# 6.896: Probability and Computation

vol. 1: Spring 2011

lecture 1

**Constantinos (Costis) Daskalakis**
costis@mit.edu

# An overview of the class

- *Card Shuffling*
- *The MCMC Paradigm*
- *Administrivia*
- *Spin Glasses*
- *Phylogenetics*

An overview of the class

- *Card Shuffling*
- *The MCMC Paradigm*
- *Administrivia*
- *Spin Glasses*
- *Phylogenetics*

# Card Shuffling

## Why do we shuffle the card deck?

We want to start the game with a uniform random permutation of the deck.

i.e. each permutation should appear with probability $1/52! \approx 1/2^{257} \approx 1/10^{77}$.

## Obtaining a random permutation:

- mathematician's approach: dice with 52! facets in 1-to-1 correspondence with the permutations of the deck

- shuffling ≈ imaginary dice

algorithm – how can we analyze it?

# Card Shuffling

Simulating the perfect dice; approaches:

- Random Transpositions

  Pick two cards *i* and *j* uniformly at random with replacement, and switch cards *i* and *j*; repeat.

- Top-in-at-Random:

  Take the top card and insert it at one of the n positions in the deck chosen uniformly at random; repeat.

- Riffle Shuffle:

# Card Shuffling

Simulating the perfect dice; approaches:

- Random Transpositions

  Pick two cards $i$ and $j$ uniformly at random with replacement, and switch cards $i$ and $j$; repeat.

- Top-in-at-Random:

  Take the top card and insert it at one of the n positions in the deck chosen uniformly at random; repeat.

- Riffle Shuffle:

  a. Split the deck into two parts according to the binomial distribution Bin($n$, 1/2).

  b. Drop cards in sequence, where the next card comes from the left hand $L$ (resp. right hand $R$) with probability $\frac{|L|}{|L|+|R|}$ (resp. $\frac{|R|}{|L|+|R|}$ ).

  c. Repeat.

# Best Shuffle?

**almost** (say within 20% from uniform)

Number of repetitions to sample a uniform permutation?

- Random Transpositions

  about 100 repetitions

- Top-in-at-Random:

  about 300 repetitions

- Riffle Shuffle:

  8 repetitions

An overview of the class

- *Card Shuffling*
- *The MCMC Paradigm*
- *Administrivia*
- *Spin Glasses*
- *Phylogenetics*

# The MCMC Paradigm

**Input:** a. very large, but finite, set $\Omega$ ;
b. a positive weight function $w : \Omega \to R^+$.

**Goal:** Sample $x \in \Omega$, with probability $\pi(x) \propto w(x)$.

in other words: $\pi(x) = \frac{w(x)}{Z}$

the "partition function"
$Z = \sum_{x \in \Omega} w(x)$

**MCMC approach:**

construct a Markov Chain (think sequence of r.v.'s) $(X_t)_t$ converging to $\pi$, i.e.

$\Pr[X_t = y \mid X_0 = x] \to \pi(y)$ as $t \to +\infty$ (independent of $x$)

**Crucial Question:** Rate of convergence to $\pi$ ("mixing time")

# e.g. Card Shuffling

State space: $\Omega$ = {all possible permutations of deck}

Weight function $w(x) = 1$ (i.e. sample a uniform permutation)

Can visualize ***shuffling method*** as a weighted directed graph on $\Omega$ whose edges are labeled by the transition probabilities from state to state.

Different shuffling methods ➔ different connectivity, transition prob.

Repeating the shuffle is performing a random walk on the graph of states, respecting these transition probabilities.

# Mixing Time of Markov Chains

**Def:** Time needed for the chain to come to within 1/2e of $\pi$ in ***total variation*** distance.

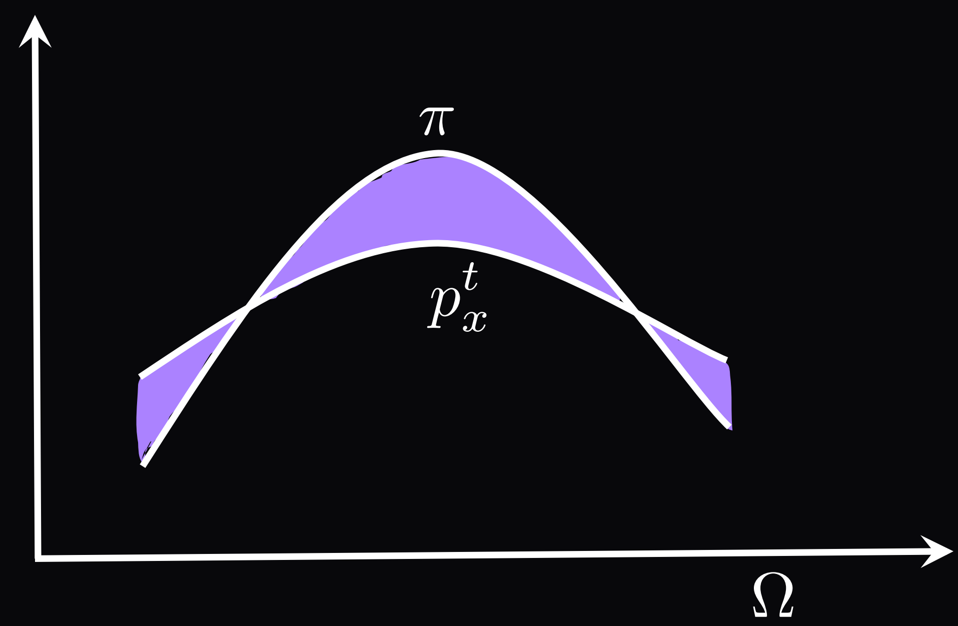
I.e.

$$\tau_{\text{mix}} := \min_t \left\{ \|p_x^t - \pi\|_{\text{TV}} \leq \frac{1}{2e},\ \forall x \right\}$$

$$\Pr[X_t = y \mid X_0 = x]$$

# [ total variation distance

$$\|\pi - p_x^t\|_{\mathrm{TV}} := \frac{1}{2} \sum_{y \in \Omega} |\pi(y) - p_x^t(y)|$$




]

# Mixing Time of Markov Chains

Time needed for the chain to come to within 1/2e of $\pi$ in ***total variation*** distance.

I.e.

$$\tau_{\text{mix}} := \min_t \left\{ \|p_x^t - \pi\|_{\text{TV}} \leq \frac{1}{2e},\ \forall x \right\}$$

$$\Pr[X_t = y \mid X_0 = x]$$

1/2e: arbitrary choice, but captures mixing

**Lemma:** $\|p_x^t - \pi\|_{\text{TV}} \leq e^{-\left\lfloor \frac{t}{\tau_{\text{mix}}} \right\rfloor}$

# Back to Card Shuffling

State space: $\Omega$ = {all possible permutations of deck}

Weight function $w(x) = 1$ (i.e. sample a uniform permutation)

- Random Transpositions

  about 100 repetitions $\tau_{\text{mix}} \sim \frac{1}{2} n \ln n$

- Top-in-at-Random:

  about 300 repetitions $\tau_{\text{mix}} \leq \frac{1}{2} n \ln n + \lceil (1 + \ln 2) n \rceil$

- Riffle Shuffle:

  8 repetitions $\tau_{\text{mix}} \sim \frac{3}{2} \log_2 n$

# Applications of MCMC

- Combinatorics
  - Examining typical members of a combinatorial set (e.g. random graphs, random SAT formulas, etc.)
  - Probabilistic Constructions (e.g. graphs w/ specified degree distributions)
- Approximate Counting (sampling to counting)
  - Counting the number of matchings of a graph/#cliques/#Sat assignments
  - E.g. counting number of people in a large crowd Ω
    - Partition Ω into two parts, e.g. those with Black hair B and its complement
    - Estimate p:=|B|/|Ω| (by taking a few samples from the population)
    - Recursively estimate number of people with Black hair $\widehat{N}_B$
    - Output estimate for size of Ω: $\widehat{N} := \widehat{N}_B \cdot 1/p$
- Volume and Integration
- Combinatorial optimization (e.g. simulated annealing)

An overview of the class

- *Card Shuffling*
- *The MCMC Paradigm*
- *Administrivia*
- *Spin Glasses*
- *Phylogenetics*

# Administrivia

*Everybody is welcome*

*If registered for credit (or pass/fail):*

- Scribe two lectures
- Collect 20 points in total from problems given in lecture
  open questions will be 10 points, decreasing
  # of points for decreasing difficulty
- Project: Survey or Research (write-up + presentation)

*If just auditing:* - Strongly encouraged to register as listeners

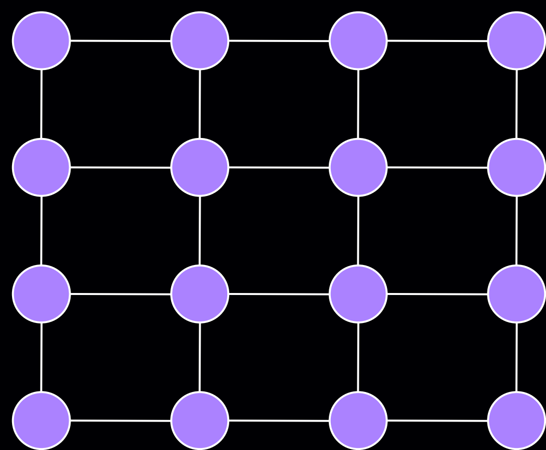
this will increase the chance we'll get a TA for the class and improve the quality of the class

An overview of the class

- *Card Shuffling*
- *The MCMC Paradigm*
- *Administrivia*
- *Spin Glasses*
- *Phylogenetics*

# Gibbs Distributions

- $\Omega$ = {configurations of a physical system comprising particles}



- Every configuration $x$, has an energy $H(x)$
- Probability of configuration $x$ is

inverse temperature

$$\pi(x) = \frac{e^{-\beta \cdot H(x)}}{Z} \qquad \text{(Gibbs distribution)}$$

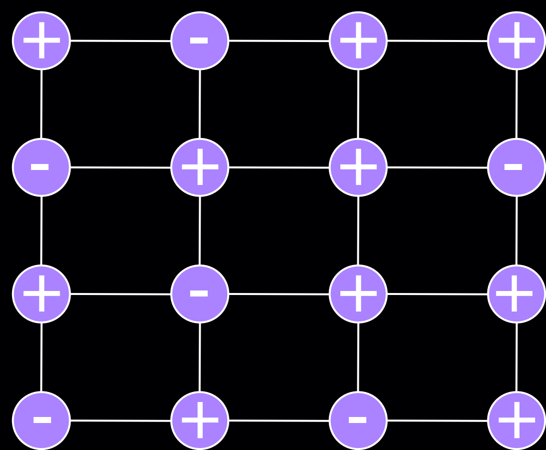
# e.g. the Ising Model

(a.k.a. Spin Glass Model, or simply a Magnet)

- $\Omega$ = {configurations of a physical system}

+: spin up

-: spin down




$$\pi(x) = \frac{e^{-\beta \cdot H(x)}}{Z}$$

(Gibbs distribution)

$$H(x) = -\sum_{i \sim j} x_i x_j$$

$\pi$ favors configurations where neighboring sites have same spin

Phenomenon is intensified as temperature decreases ($\beta$ increases)

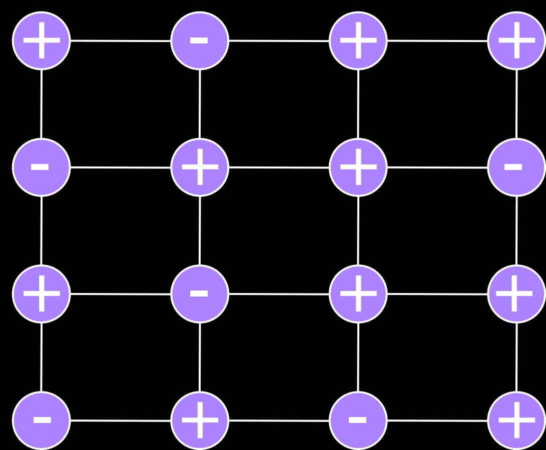
# e.g. the Ising Model

(a.k.a. Spin Glass Model, or simply a Magnet)

- $\Omega$ = {configurations of a physical system}

+: spin up

-: spin down




$$\pi(x) = \frac{e^{-\beta \cdot H(x)}}{Z}$$

$$H(x) = -\sum_{i \sim j} x_i x_j$$

known fact: Exists critical $\beta_c$ such that

$\beta < \beta_c$: system is in disordered state (random sea of + and -)

$\beta > \beta_c$: system exhibits long-range order

(system likely to exhibit a large region of + or of -)

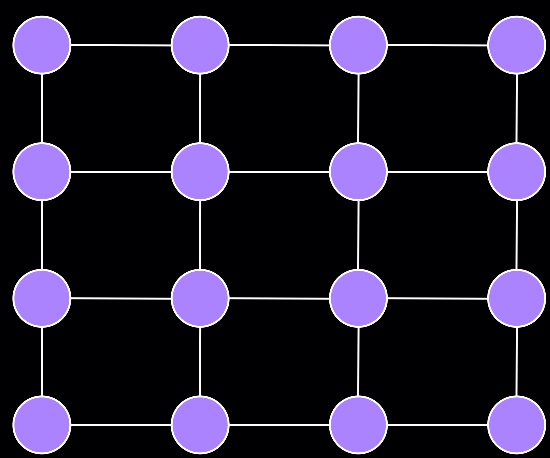
spontaneous magnetization

# Sampling the Gibbs Distribution

Uses of sampling:

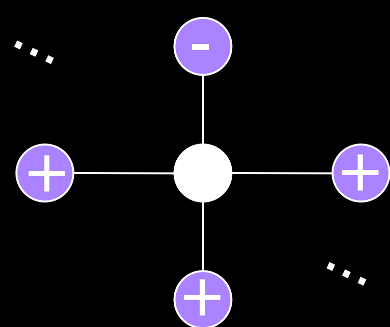
- Examine typical configurations of the system at a temperature
- Compute expectation w.r.t. to $\pi$. (e.g. for the Ising model the mean magnetization)
- Estimate the partition function $Z$ *(related to the entropy of the system)*

$$\pi(x) = \frac{e^{-\beta \cdot H(x)}}{Z}$$



# Glauber Dynamics

- Start from arbitrary configuration.
- At every time step $t$:
  - *Pick random particle*
  - *Sample the particle's spin conditioning on the spins of the neighbors*




$+$, w.pr. $\dfrac{e^{2\beta}}{e^{2\beta}+e^{-2\beta}}$

-, w.pr. $\dfrac{e^{-2\beta}}{e^{2\beta}+e^{-2\beta}}$

# Physics ➔ Computation !

**Theorem [MO '94]:** The mixing time of Glauber dynamics on the $\sqrt{n} \times \sqrt{n}$ box is

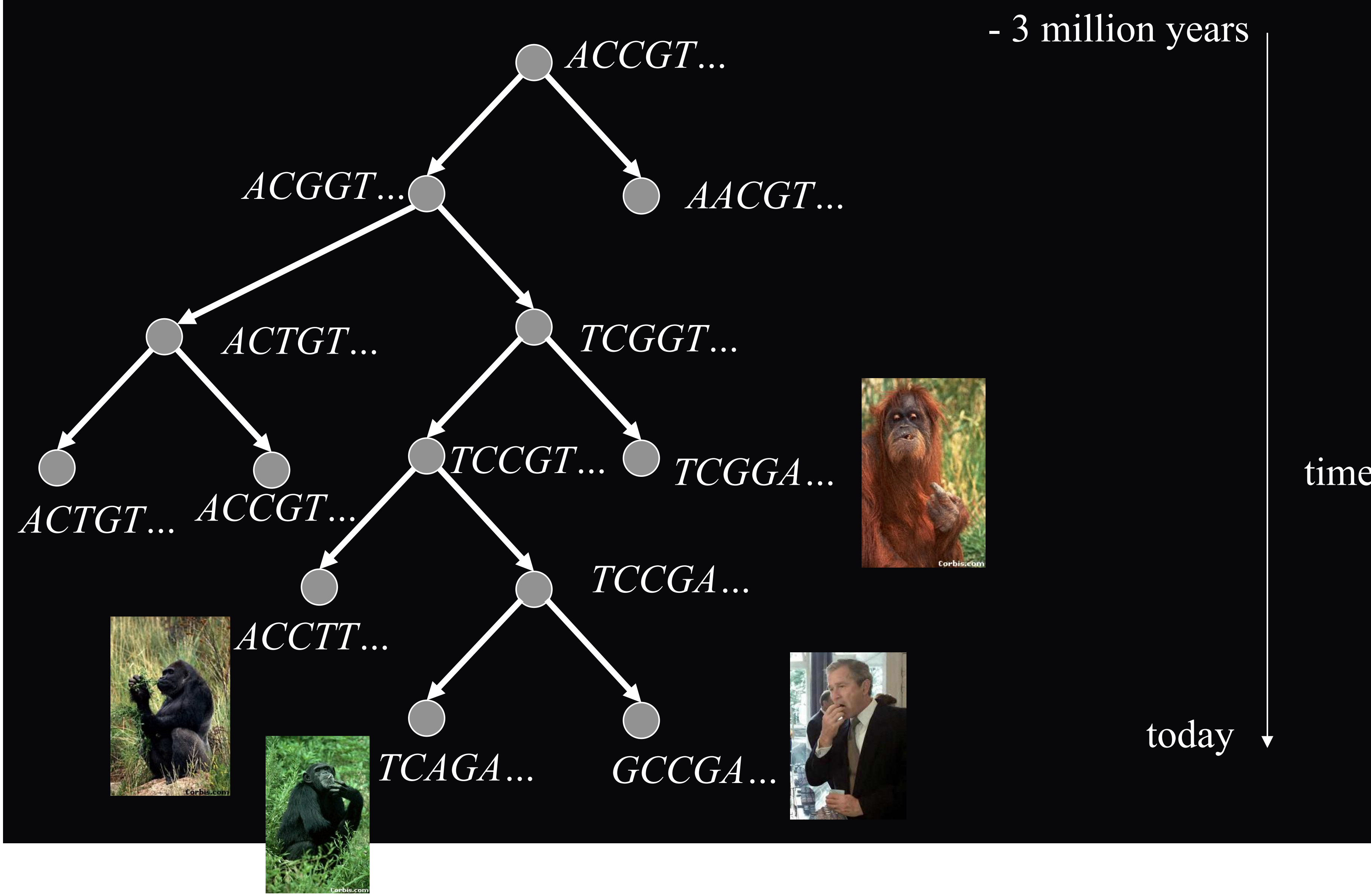
$$\begin{cases} O(n \log n), \text{ if } \beta < \beta_c & \text{(i.e. high temperature)} \\ e^{\Omega(\sqrt{n})}, \text{ if } \beta > \beta_c & \text{(i.e. low temperature)} \end{cases}$$

where $\beta_c$ is the critical (inverse) temperature.

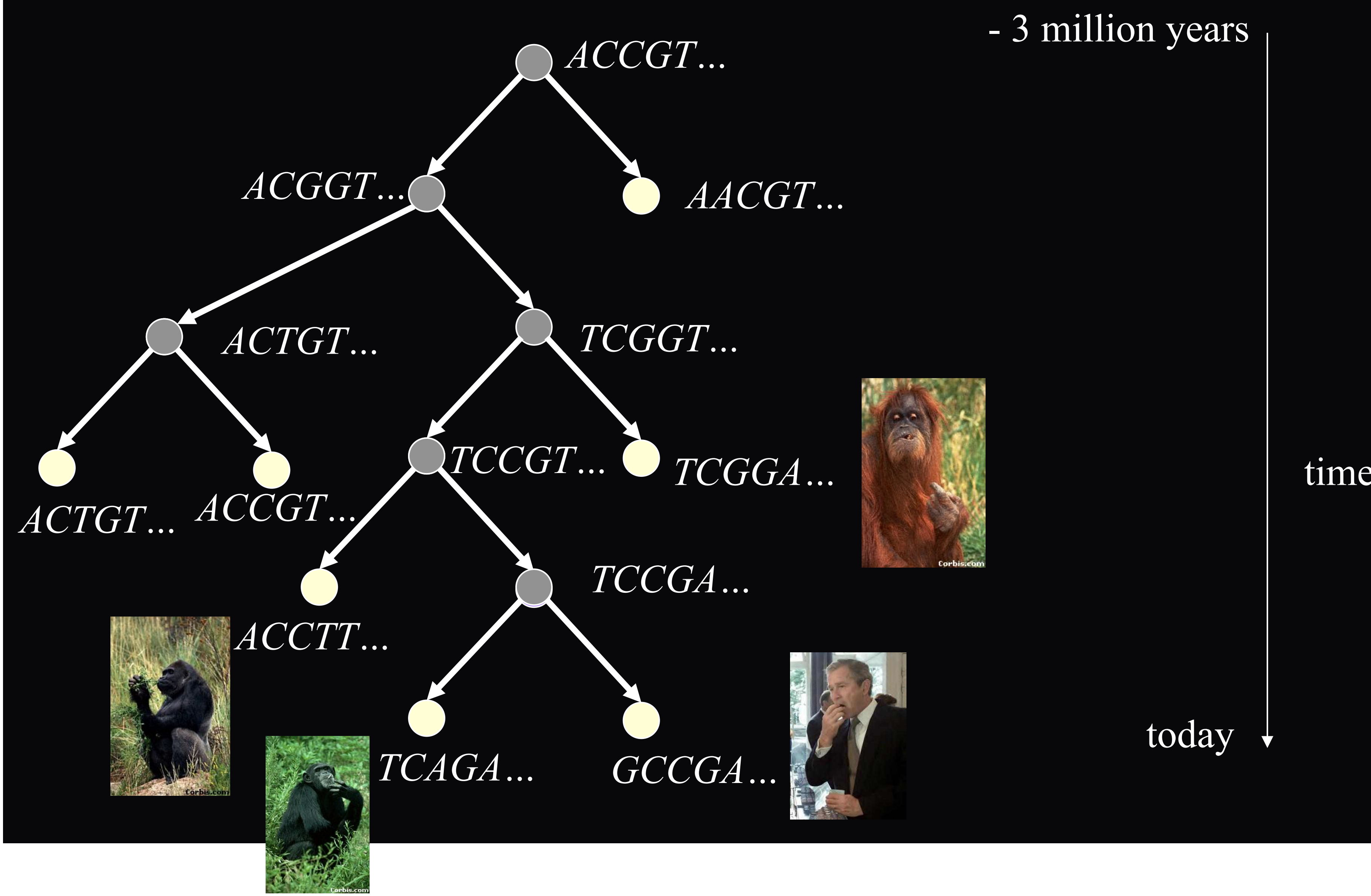
An overview of the class

- *Card Shuffling*
- *The MCMC Paradigm*
- *Administrivia*
- *Spin Glasses*
- *Phylogenetics*

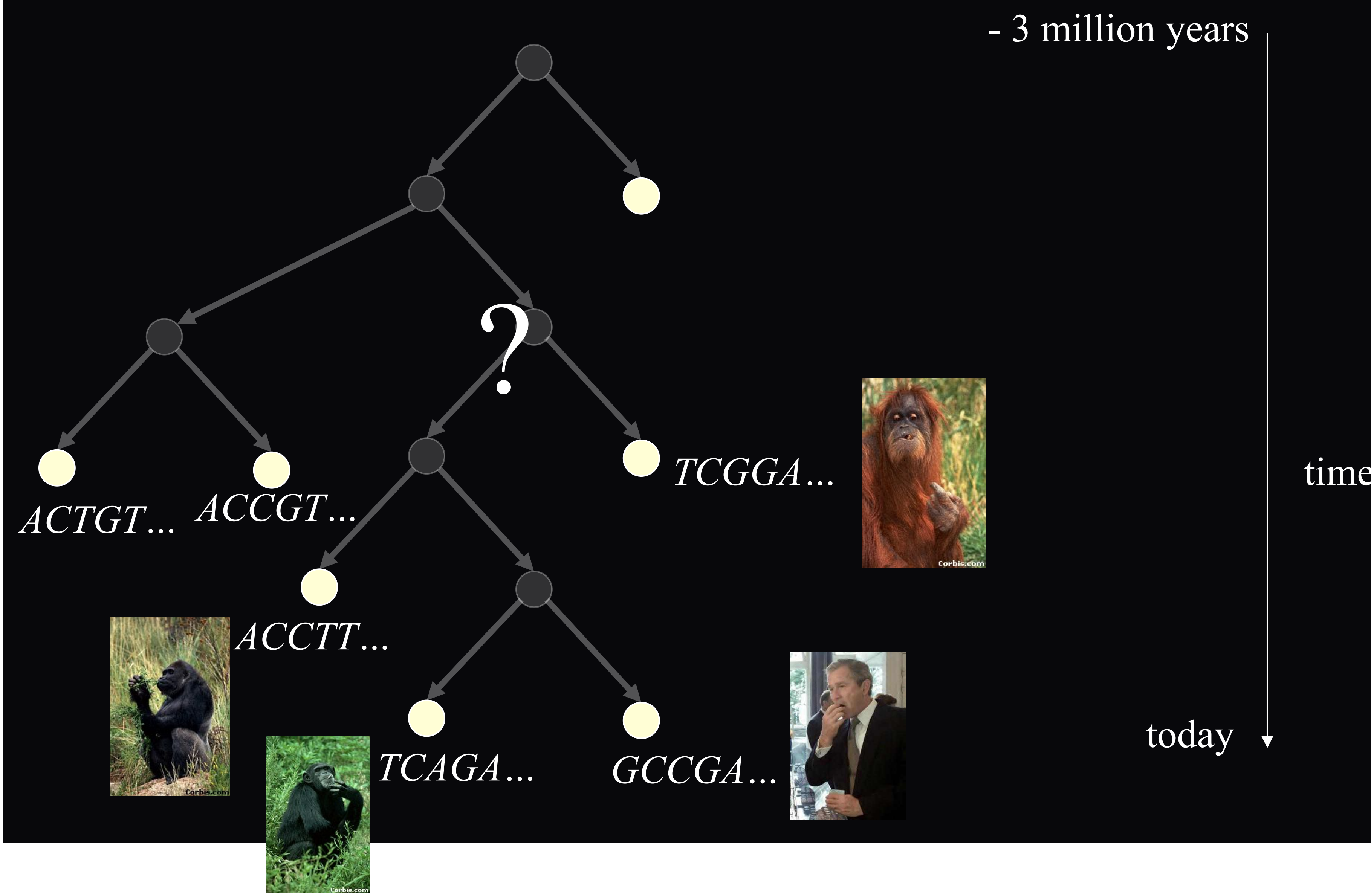
# evolution


- 3 million years
time
today
ACCGT…
ACGGT…
AACGT…
ACTGT…
TCGGT…
ACTGT…
ACCGT…
TCCGT…
TCGGA…
ACCTT…
TCCGA…
TCAGA…
GCCGA…

# the computational problem



# the computational problem

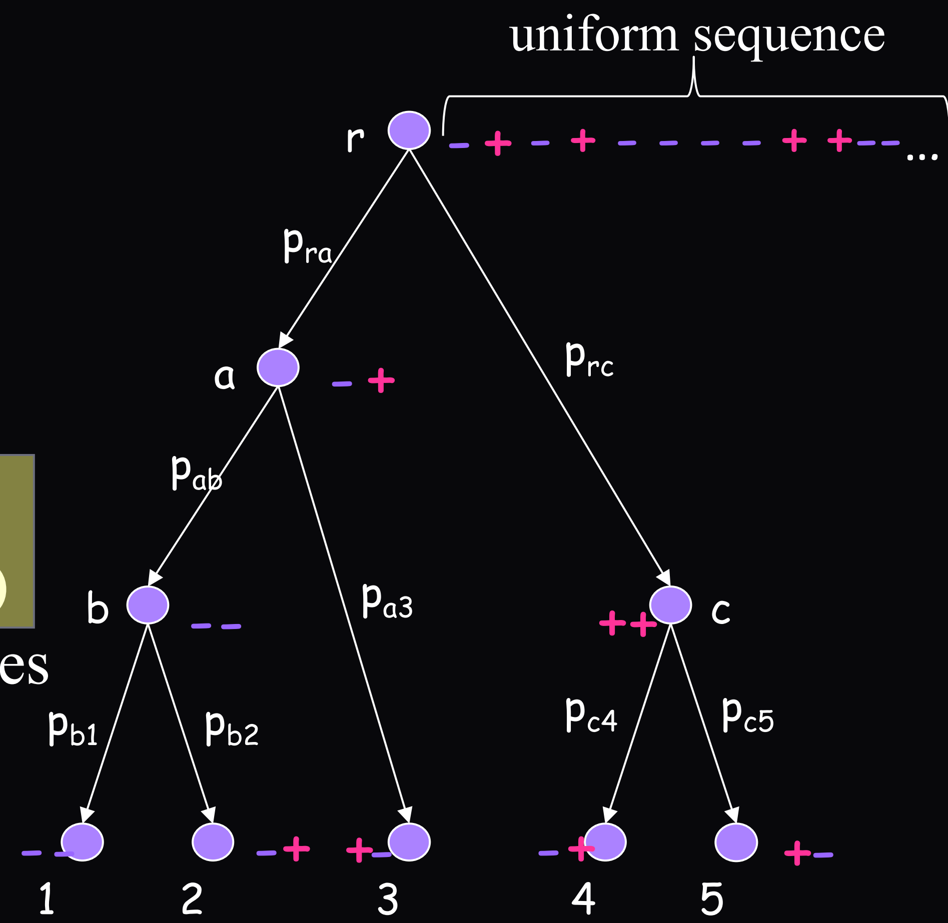


# Markov Model on a Tree

0. Tree T = (V, E) on *n* leaves
1. State Space: Σ= {-1, +1}

**-1: Purines (A,G)**
**+1: Pyrimidines (C,T)**

2. Mutation Probabilities on edges
3. Uniform state at the root



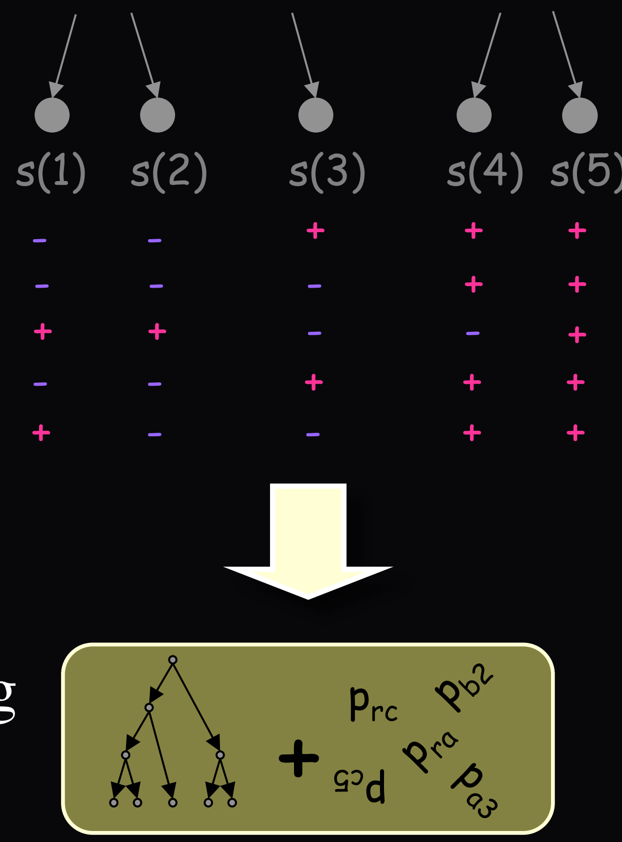

**Lemma**: Equivalent to taking independent samples from the Ising model (with appropriate temperatures related to mutation probabilities).

# The Phylogenetic Reconstruction Problem

Input: $k$ independent samples of the process at the leaves of an $n$ leaf tree – but **tree not known!**

**Task:** fully reconstruct the model,
i.e. find **tree** *and* **mutation probabilities**

**Goal:** complete the task **efficiently**
use **small sequences (i.e. small** ***k*****)**



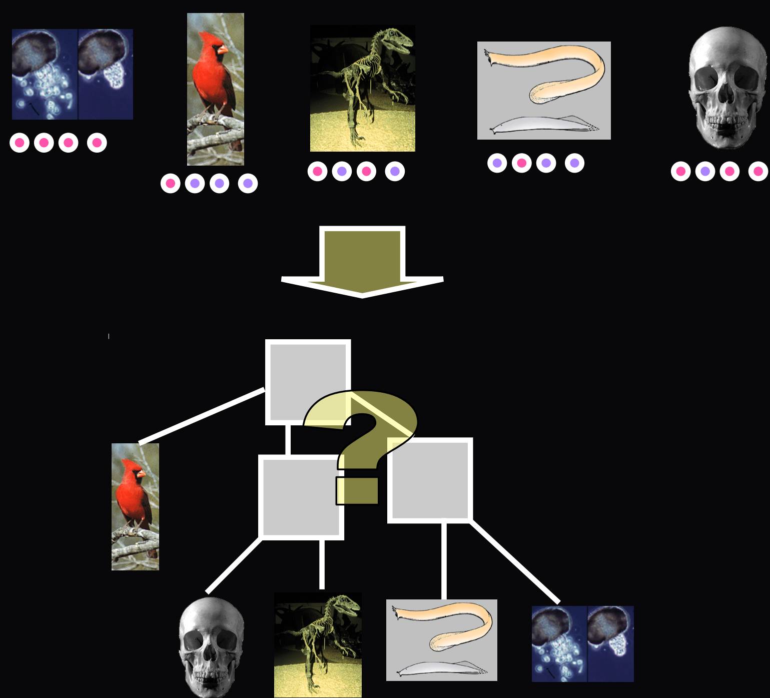

**In other words:** Given k samples from the Ising model on the tree, can we reconstruct the tree?

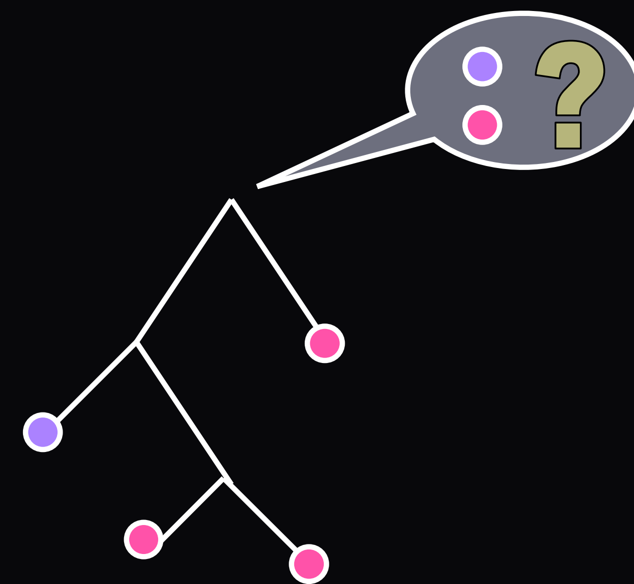
A: yes, taking $k = \text{poly}(n)$

*Can we perform reconstruction using shorter sequences?*

A: Yes, if “temperature” (equivalently the mutation probability) is sufficiently low”.

# Phylogenetics ⟷ Physics !!





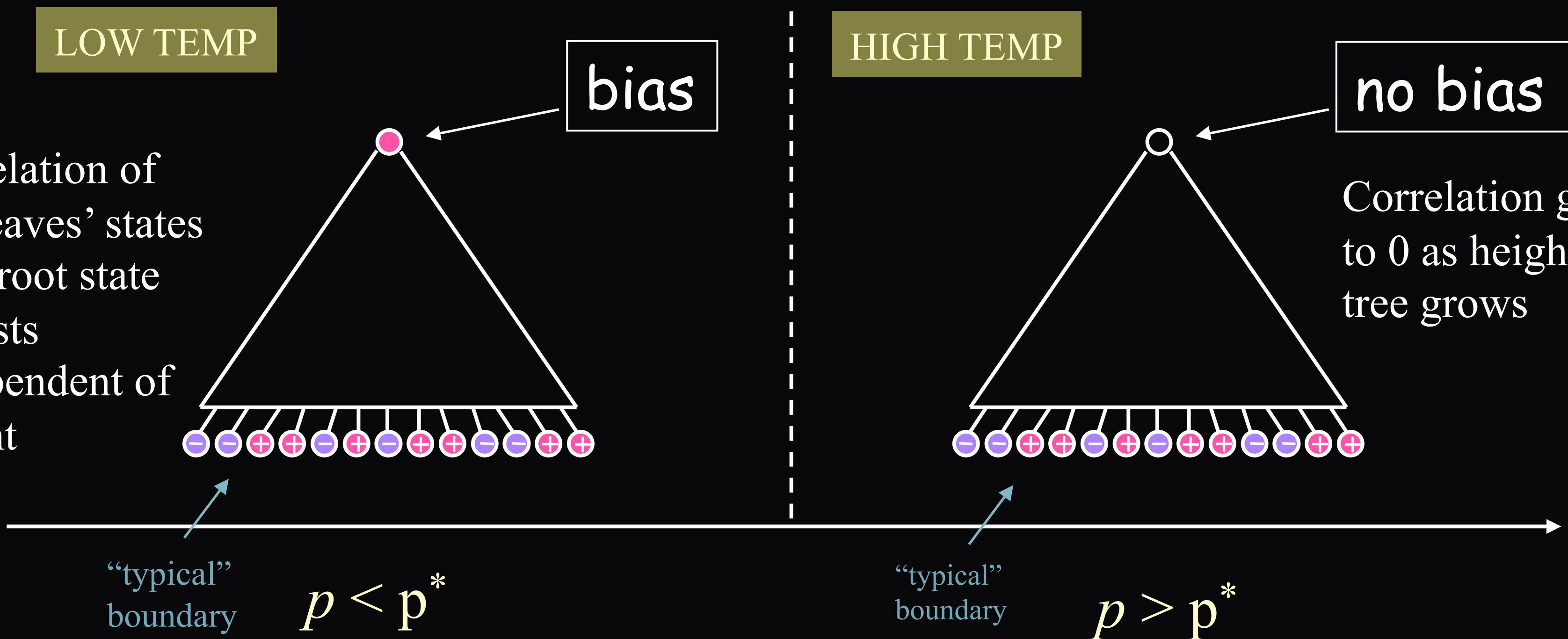

[Daskalakis-Mossel-Roch 06]

**The phylogenetic reconstruction** problem can be solved from **very short sequences** ⟷ The Ising model on the tree exhibits long-range order

**phylogeny** | **statistical physics**

# The Underlying Phase Transition: Root Reconstruction in the Ising Model

Correlation of the leaves' states with root state persists independent of height




Correlation goes to 0 as height of tree grows

The transition at p* was proved by:
[Bleher-Ruiz-Zagrebnov'95], [Ioffe'96],[Evans-Kenyon-Peres-Schulman'00], [Kenyon-Mossel-Peres'01],[Martinelli-Sinclair-Weitz'04], [Borgs-Chayes-Mossel-R'06]. Also, "spin-glass" case studied by [Chayes-Chayes-Sethna-Thouless'86]. Solvability for p* was first proved by [Higuchi'77] (and [Kesten-Stigum'66]).

$$p^* = \frac{\sqrt{2}-1}{\sqrt{8}}$$

# Physics $\Leftrightarrow$ Phylogenetics

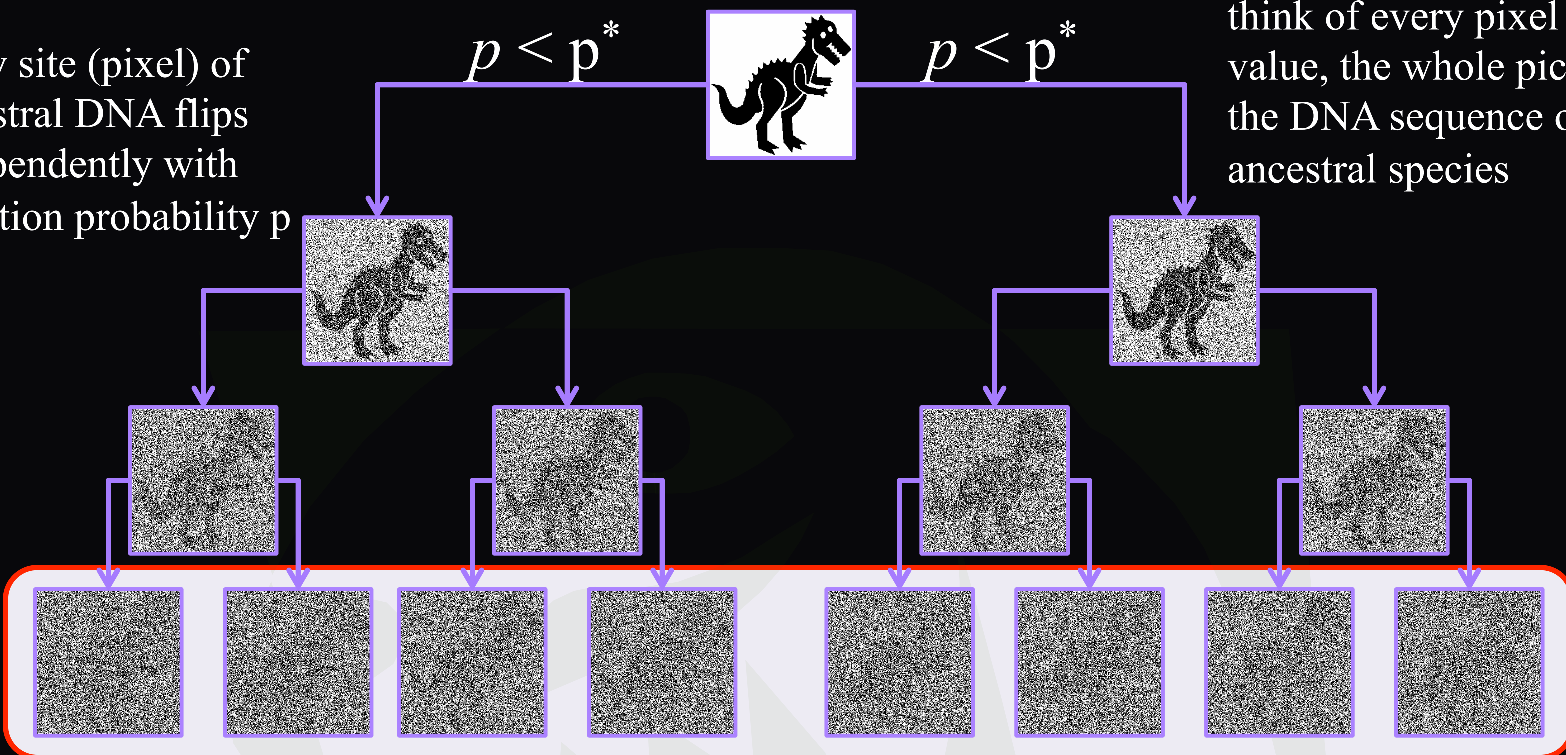
| | Statistical physics | | Phylogeny |
|---|---|---|---|
| [DMR'05] | Low Temp | $\Leftrightarrow$ | $k = \Theta(\log n)$ |
| | | $p = p^*$ | |
| [Mossel'03] | High Temp | $\Leftrightarrow$ | $k = \Theta(\mathrm{poly}(n))$ |

**Resolution of Steel's Conjecture**

*Low-temperature behavior of the Ising model…*

# The Root Reconstruction Problem (low temperature)

every site (pixel) of ancestral DNA flips independently with mutation probability p




think of every pixel as a +1/-1 value, the whole picture being the DNA sequence of ancestral species

the question is how correlated are the states at the leaves with the state at the root

let's try taking majority across leaves for every pixel

**Thm: Below $p^*$ correlation will persist, no matter how deep the tree is !**

picture will look as clean independently of depth

**Thm: Above $p^*$ correlation goes to 0 as the depth of the tree grows.**

no way to get close to root picture, using leaf pictures

# Physics $\Longleftrightarrow$ Phylogenetics

| Statistical physics | | Phylogeny |
| --- | --- | --- |
| [DMR'05] Low Temp | $\Longleftrightarrow$ | $k = \Theta(\log n)$ |
| | $p = p^*$ | |
| [Mossel'03] High Temp | $\Longleftrightarrow$ | $k = \Theta(\mathrm{poly}(n))$ |

**Resolution of Steel's Conjecture**

The point of this result is that phylogenetic reconstruction can be done with O(log *n*) sequences IFF the underlying Ising model exhibits long-range order.