Bounding Chain Techniques for Perfect Sampling

Mark Huber
Dept. of Mathematics and Institute of Statistics and Decision Sciences
Duke University
mhuber@math.duke.edu
www.math.duke.edu/~mhuber

Water, Markov chains, and NP

Q: What do they have in common?

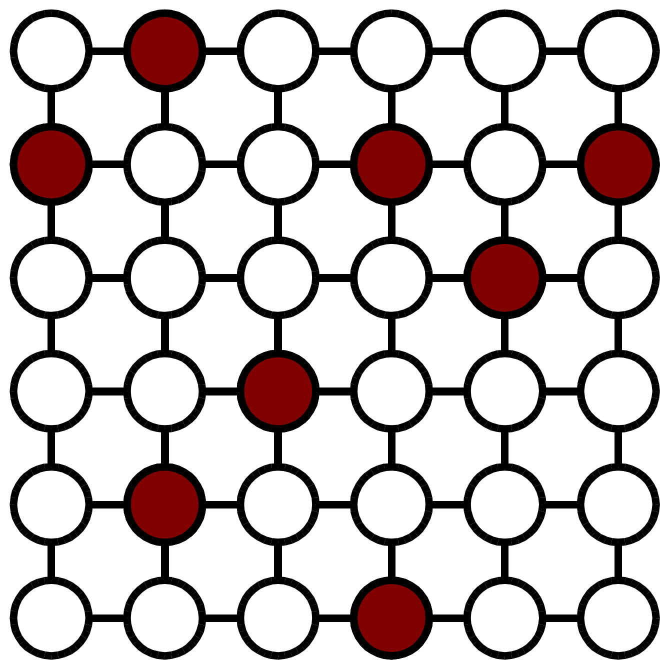
A: All three exhibit phase transitions

- Water: Liquid-gaseous form at temperature
- MC: Some parameters fast-others slow
- NP: Approximate within 20% fast
  within 19.9999999% slow

Linked through simple physical models

Example: Hard Core Gas Model

Very simple model of gases:

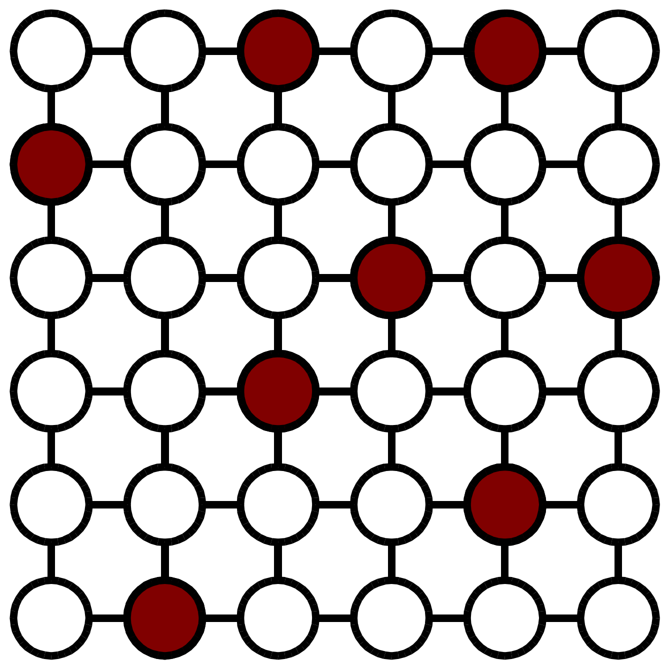


The rule: no two adjacent sites occupied

CS version: Occupied sites an independent set

Adding a temperature...

Probability distribution on independent sets



$$x(i) = \begin{cases} 1 \\ 0 \end{cases}$$

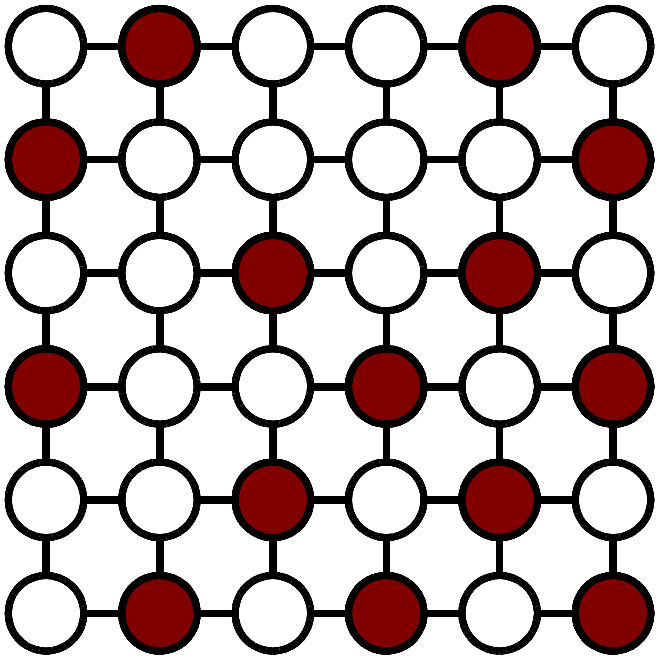
$$\pi(x) = \frac{\lambda^{\sum x}}{Z_\lambda}$$

λ known as "activity" or "fugacity"

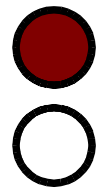
Z_λ unknown normalizing constant

High Activity

As λ changes, distribution shifts:



$$x(i) = \begin{cases} 1 \\ 0 \end{cases}$$



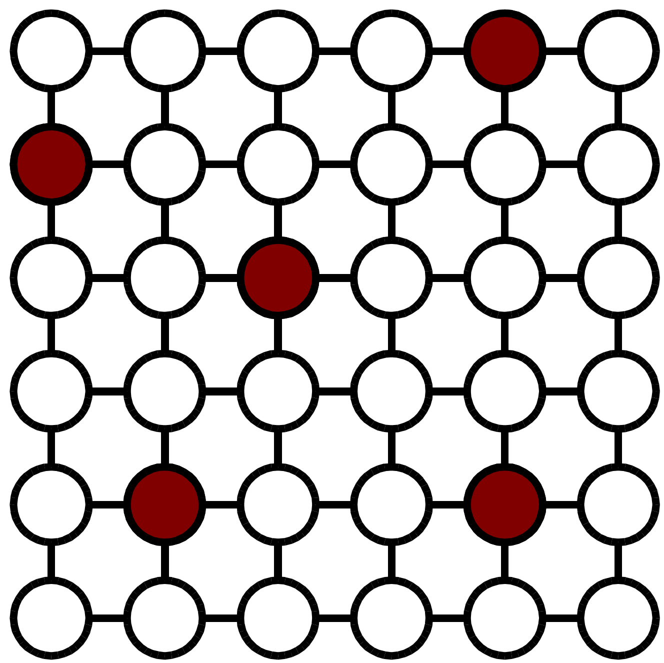
$$\pi(x) = \frac{\lambda^{\sum x}}{Z_\lambda}$$

λ large gives checkerboard pattern

Long range behavior

Low Activity

Probability distribution on independent sets



$$x(i)=\begin{cases}1 \\ 0\end{cases}$$

$$\pi(x)=\frac{\lambda^{\sum x}}{Z_\lambda}$$

λ low, no pattern

Short range behavior

The phase transition

Z_λ does not change analytically in λ

(discontinuities in derivatives)

Effect: correlations decline polynomially instead of exponentially

The complexity connection

Linked to maximum independent set

- NP complete problem
- for fixed λ and Δ (max degree of graph) sampling gives an approximation algorithm
- known that cannot approximate within a constant factor (Dyer, Frieze, Jerrum '98) when

$$\lambda \geq \frac{25}{\Delta}$$

unless

$$RP = NP$$

Markov chains

Widespread use for Monte Carlo algorithms

- Often only method available
- Simple Example: Shuffling Cards
- Idea: take lots of small random moves
- Under simple conditions (connectness, aperi-odicity) converges to stationary distribution
- Drawback: do not know *mixing time* of chain
- For (unknown) reasons, mixing time can be slow at phase transitions

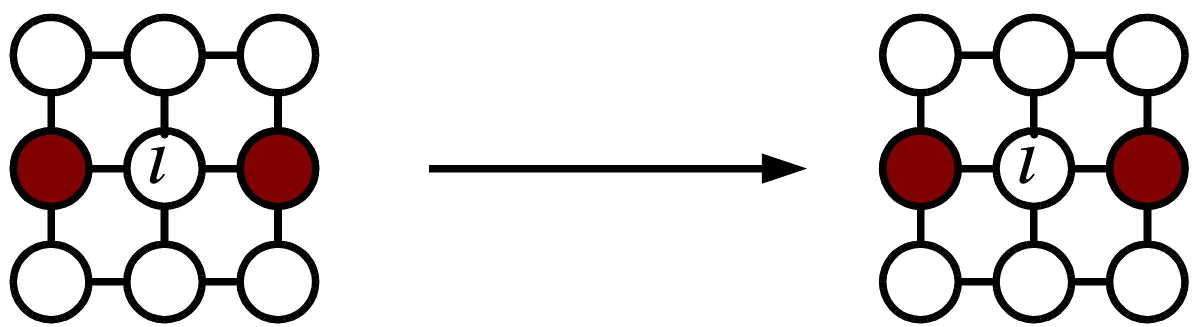
Standard Gibbs sampler

At each step, make small random change:

Step 1: Choose i uniformly from nodes V

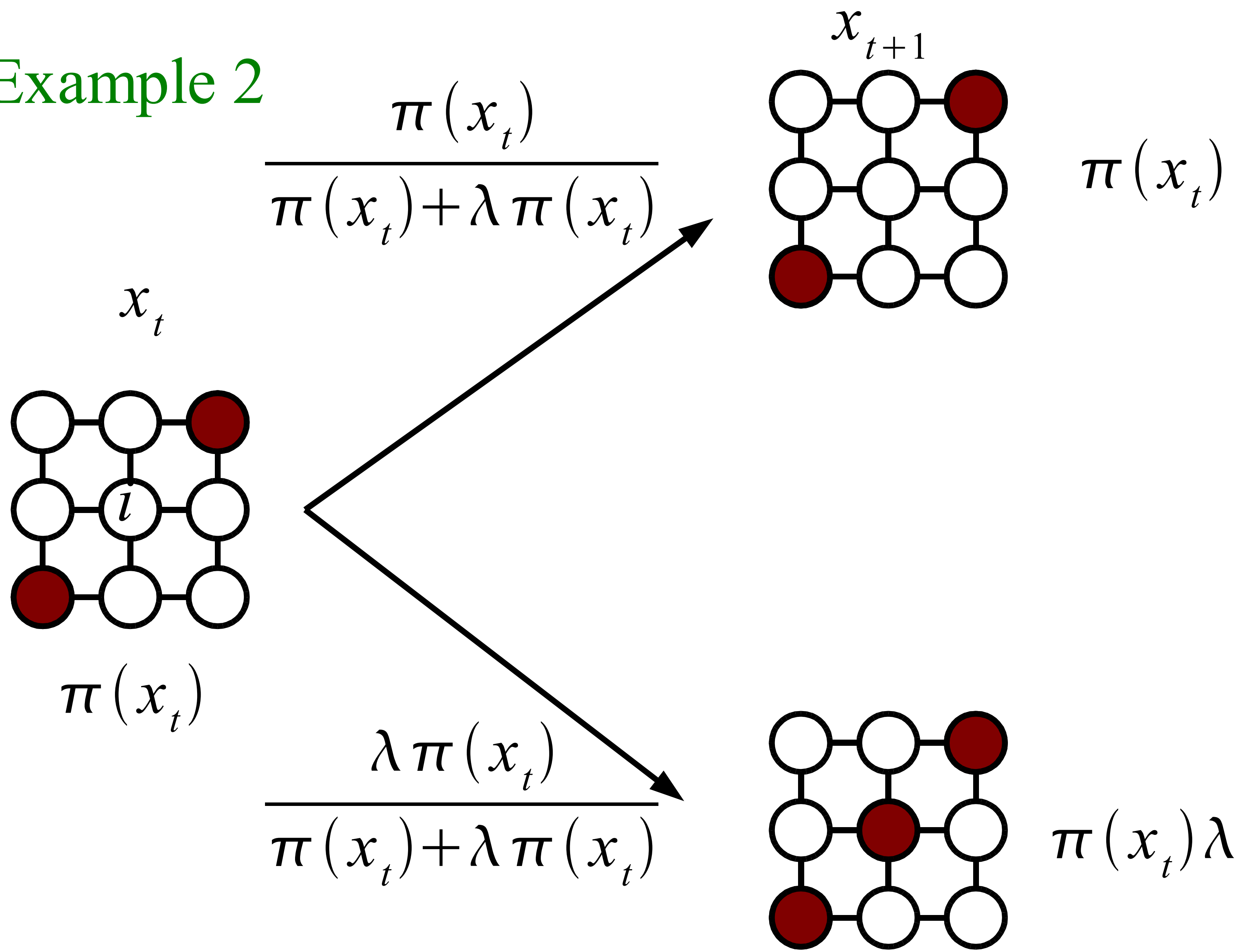
Step 2: Choose $x(i)$ from distribution conditioned on $x(V \setminus \{i\})$

Example 1



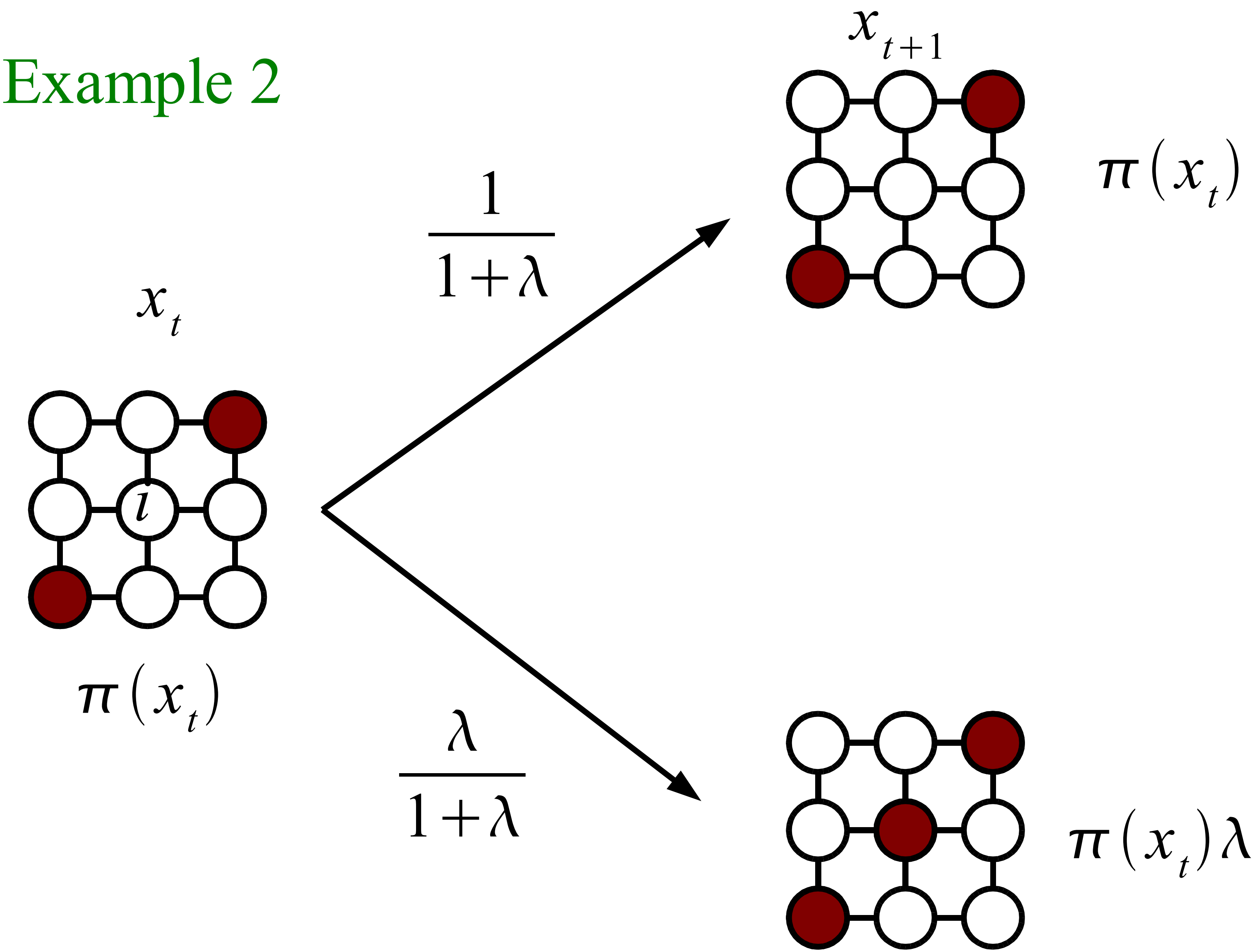
A more interesting example...

Example 2



A more interesting example...

Example 2



The Markov chain approach

1) Start with a configuration x_0
2) Take “lots” of steps in the Markov chain
3) Output final state x_t as random sample

The Good News:

Can design a chain for nearly all distributions

The Bad News:

Don't know what “lots” is

Mixing Time

Markov chain approach often rediscovered
(right along with its problems)

Mixing Time **Warm-up Time**

Initialization Time **Burn-in Time**

Dememorization Time

While mixing time unknown, MCMC
heurestic rather than algorithm

Bounding chain idea

Huber '98, Haggstrom & Nelander '99

Idea: Don't need to know state of chain in order to take steps

Requires: Implementation of chain on computer (formally: complete coupling)

Perfect Sampling: (CFTP Propp Wilson '95)
- Start chain in unknown stationary state,
- run chain forward hoping state becomes known

Computer implementation

Choose i uniformly from V

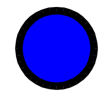
Choose U uniformly from $[0,1]$

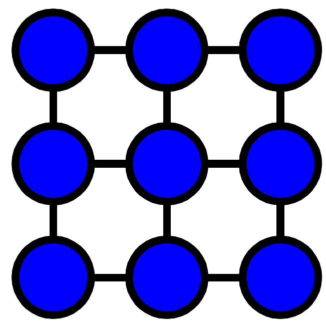
If $U \leq \frac{1}{1+\lambda}$ or j neighbors i and $x(i)=1$

Let $x(i) \leftarrow 1$

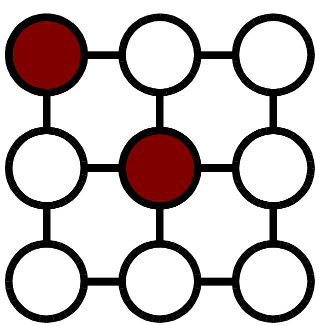
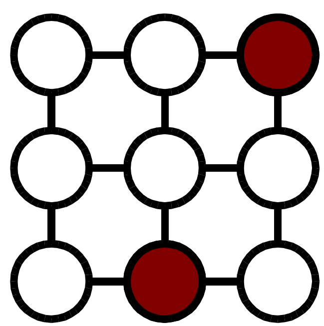
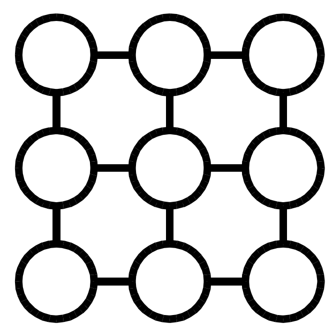
Bounding chain

Introduce new value for $x(i)$ means unknown

$x(i) = ?$ 

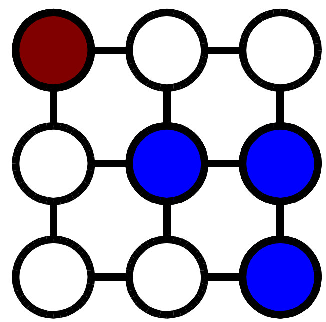


Consistent with

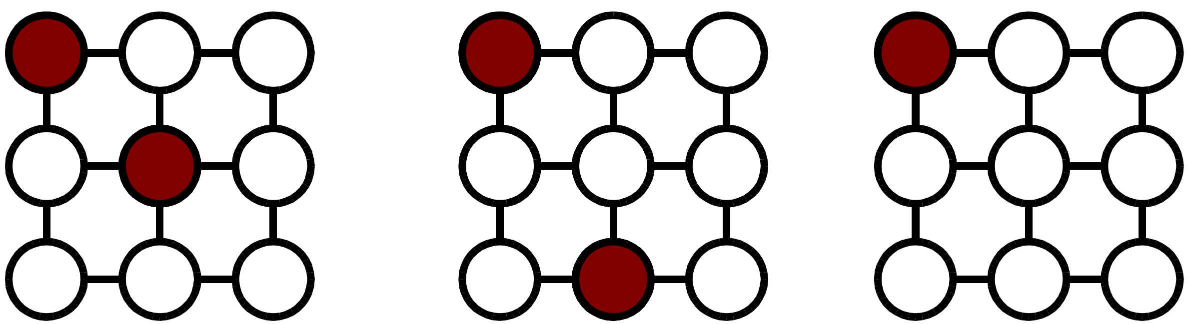
   ...

Bounding chain

Can have mix of known and unknown

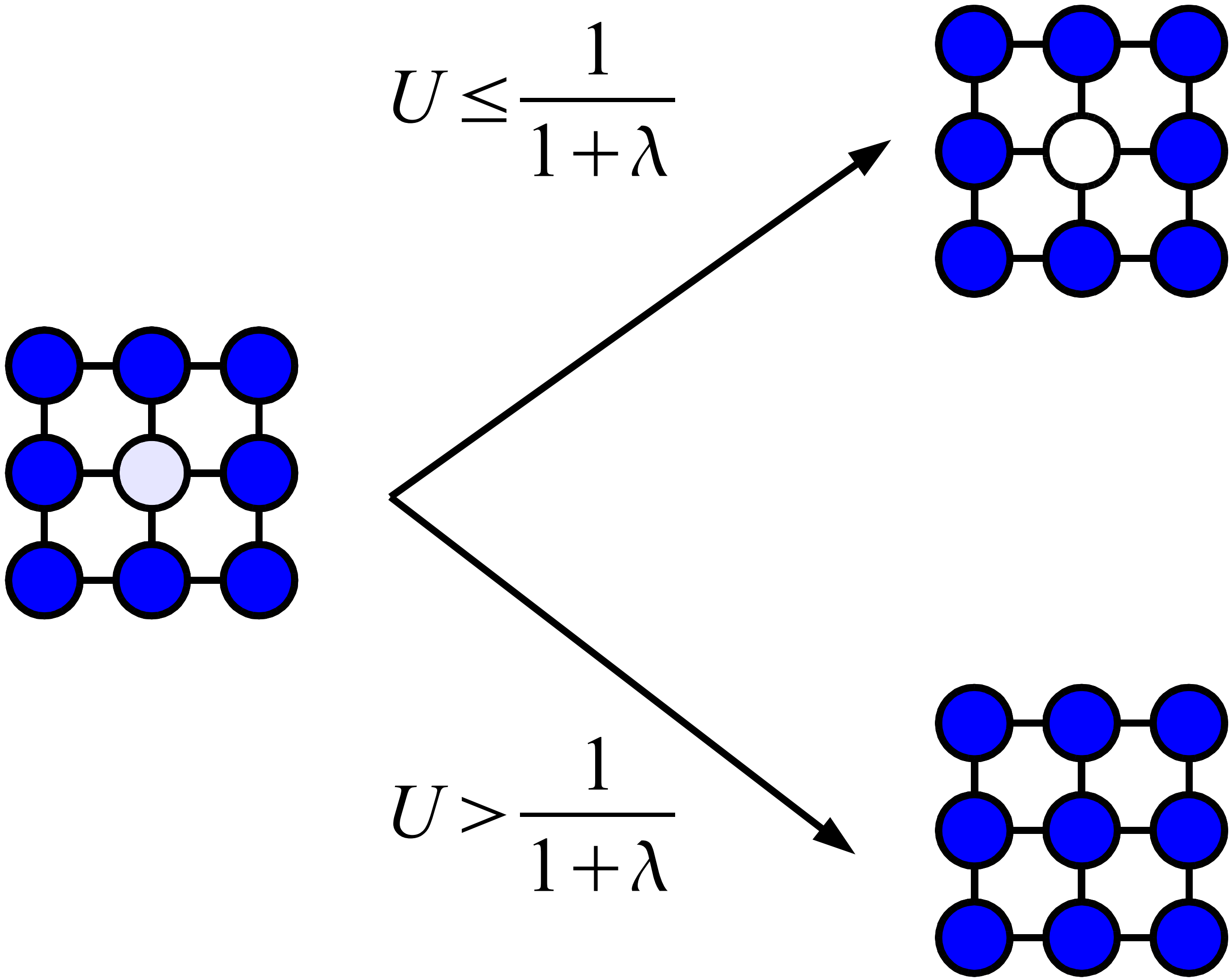


Consistent with

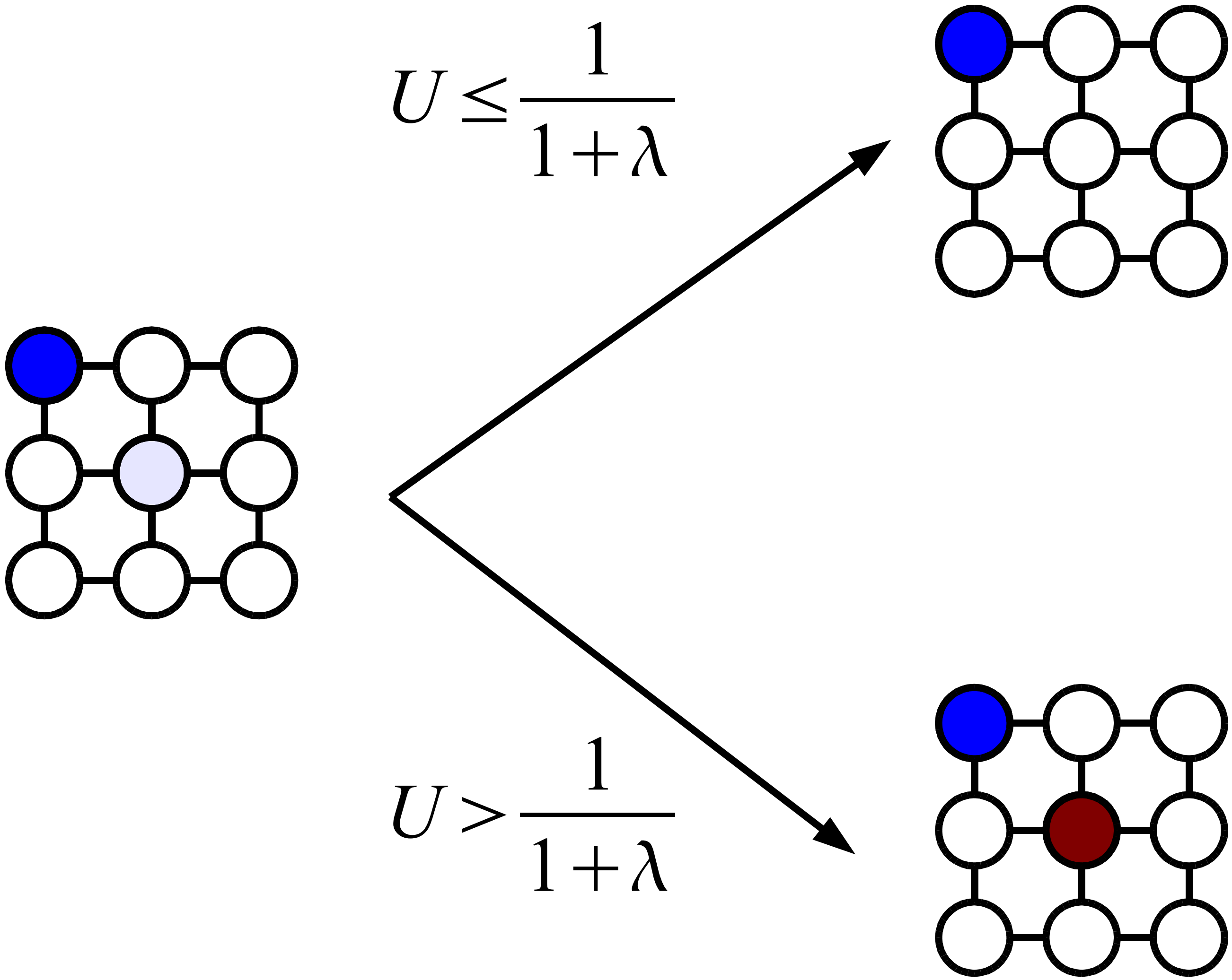


...

Bounding chain



Bounding chain



Formal definition

In General

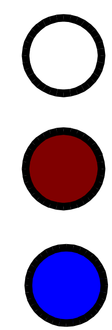
$$x_t \in C^V$$

$$y_t \in (2^C)^V$$

Independent Sets

$$x_t \in \{0,1\}^V$$

$$y_t \in \begin{cases} \{0\} \\ \{1\} \\ \{0,1\} \end{cases}$$



$\{y_t\}$ is a bounding chain if Markovian and

$$x_t(i) \in y_t(i) \ \ \forall i \rightarrow x_{t+1}(i) \in y_{t+1}(i) \ \ \forall i$$

Overall

Start
Unknown
Sample

Take Fixed
# of Steps

Hopefully
State Ends
Known

A nagging question…

Q: What do we do if the sample is still unknown?

A: Use Coupling from the past (Propp/Wilson '96)

CFTP(T)

Set y_0 to unknown state $(y_0(i) = C \quad \forall i)$

Take T steps

If state known, output single $x_T \in y_T$

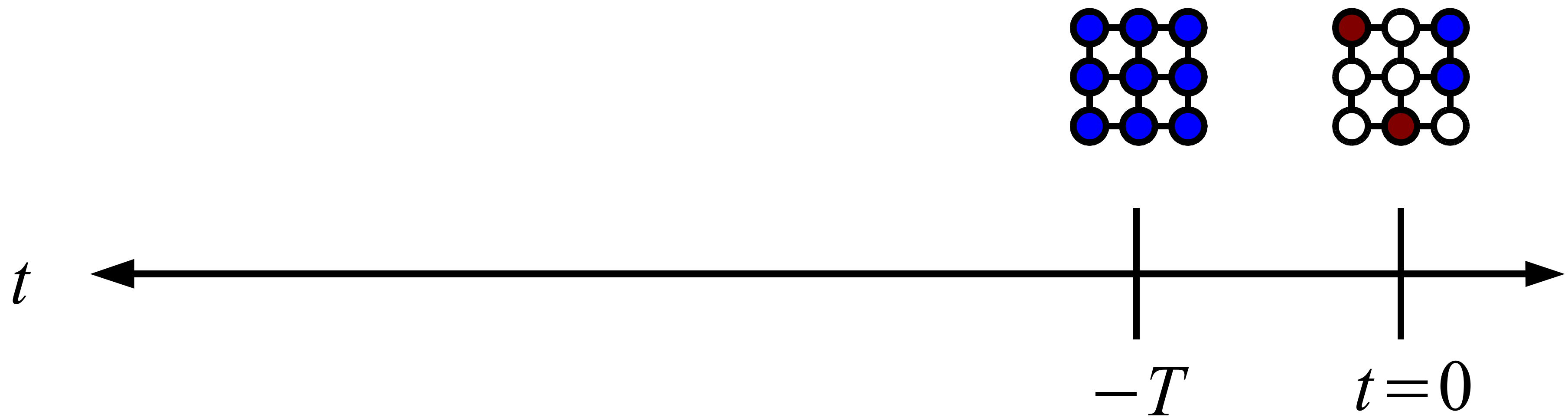
Else

Set $x_0 = CFTP(T)$

Output x_T

"from the past" point of view

Alternate way of looking at CFTP:



“from the past” point of view

Alternate way of looking at CFTP:

Keep going back in time until succeed

Results

Theorem: Expected running time for this procedure is $O(n \ln n)$ when

$$\lambda \leq \frac{1}{\Delta - 3}$$

Proof Idea: Number of ● decreases on average at each step when λ small (note ● only created next to ●)

More complex chain

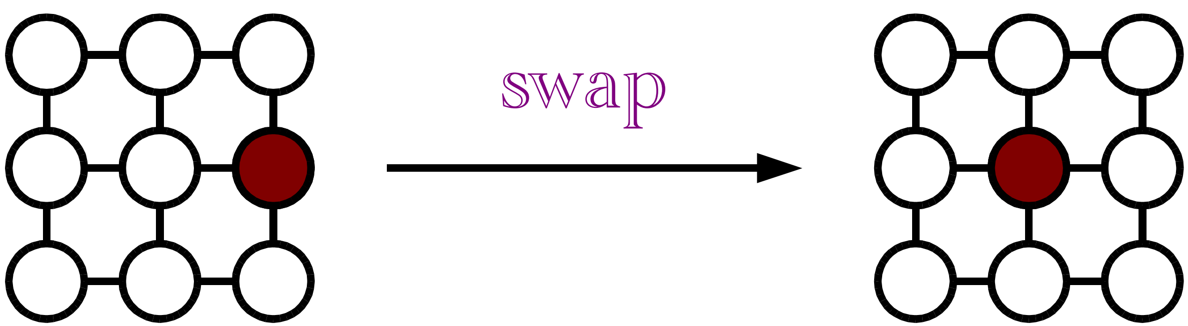
Dyer and Greenhill '00 created improved chain:

Swap Move

If exactly one neighbor occupied

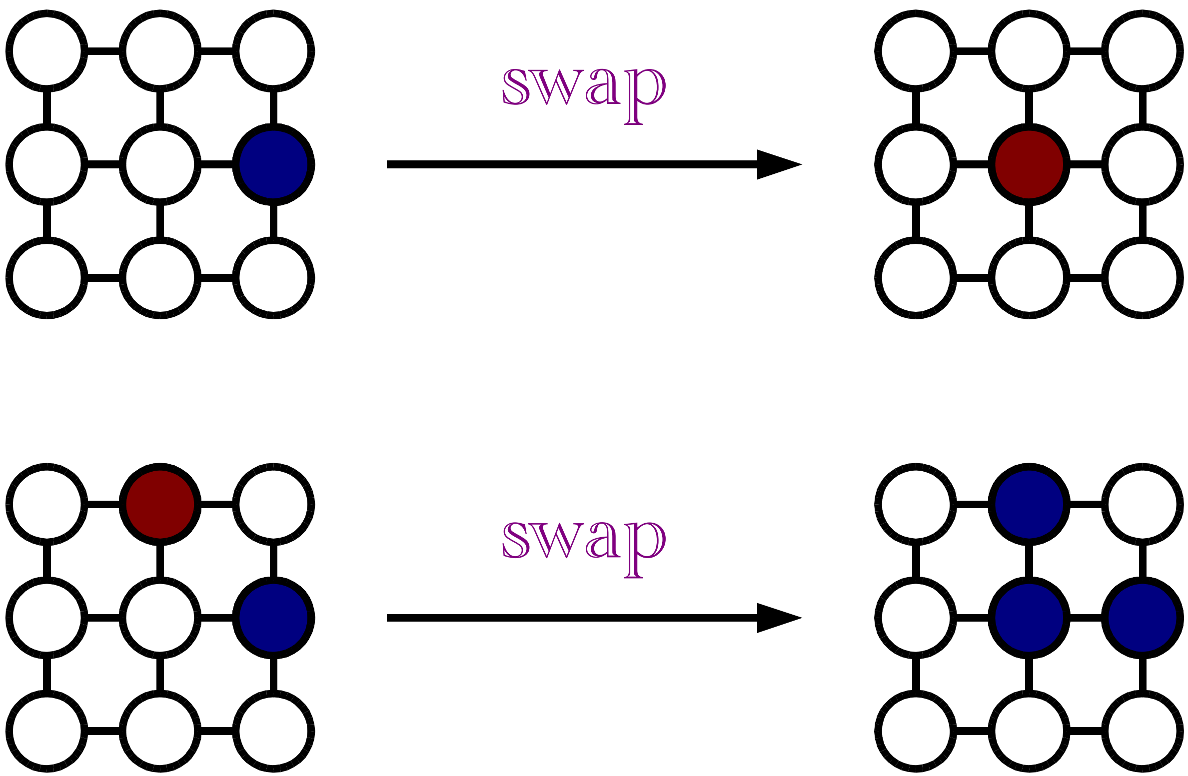
(and roll to add node to ind set)

Swap with probability p



Works well for bounding chain

(Huber '99)



Better Results

Theorem: Expected running time for this procedure is $O(n \ln n)$ when

$$\lambda \leq \frac{2}{\Delta - 2}$$

(about 3 or 4 times as fast in practice as previous method)

Other approaches

Other perfect sampling methods

- CFTP like algorithms
  (Fill, Machida, Murdoch, Rosenthal '99)
  (Read once CFTP Wilson '99)
- The Randomness Recycler
  (Fill, Huber '99)
- All provably polynomial when

$$\lambda = O\left(\frac{1}{\Delta}\right)$$

Conclusions

The problem:

Phase transitions appear in unlikely situations
Of practical importance to try to eliminate

What bounding chains give:

Avoids the mixing time question for Markov chains
Allows use of perfect sampling protocols like CFTP
Delivers samples exactly from desired distribution
Cannot beat the underlying Markov chain
Still a gap between where we can efficiently sample and where we can't unless $P = NP$