Maybe some local algorithms aren't that bad...?

Timothy M. Garoni

t.garoni@ms.unimelb.edu.au

MASCOS

Department of Mathematics and Statistics

The University of Melbourne

Australia

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Overview

How do we efficiently simulate models near criticality?

- Problem: critical slowing-down
- The current state-of-the-art: cluster algorithms
  - Use global moves in clever way
- We will discuss two local algorithms:
  - Sweeny algorithm
    - Simulates the random-cluster model
  - Worm algorithm for the Ising model
    - Simulates the high-temperature graphs
- Both display critical speeding-up and multiple time-scales

References/Collaborators

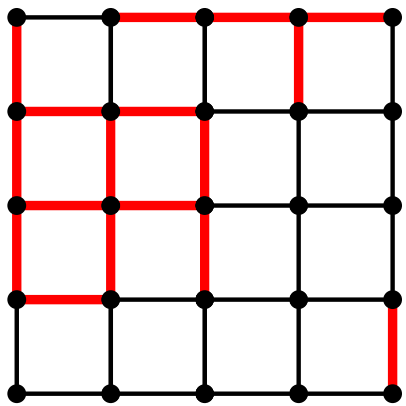
Youjin Deng, Timothy M. Garoni, and Alan D. Sokal, *Critical Speeding-Up in the Local Dynamics of the Random-Cluster Model*, Phys. Rev. Lett. 98, 230602 (2007).

Youjin Deng, Timothy M. Garoni, and Alan D. Sokal, *Dynamic Critical Behavior of the Worm Algorithm for the Ising Model*, Phys. Rev. Lett. 99, 110601 (2007).

Sweeny's algorithm

Random-cluster model

- Fortuin-Kasteleyn 1969
- Fix a finite graph $G = (V, E)$ and real numbers $q, v > 0$



- Pick a random bond configuration $A \subseteq E$ with probability

$$\mathbb{P}(A) \propto q^{k(A)}\, v^{|A|}$$

- $k(A) =$ number of components of (V, A)

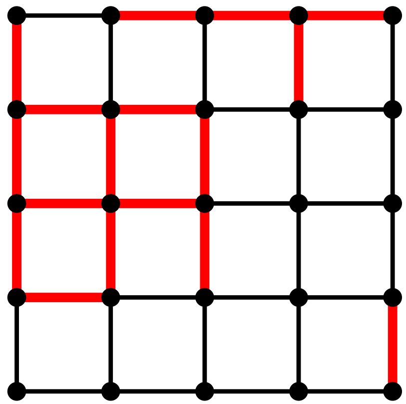
- Integer $q \geq 2$ equivalent to q-state Potts model ($q = 2$ Ising)
- $q = 1$ reduces to bond percolation
- $q \to 0$ gives connected spanning subgraphs, spanning forests, spanning trees

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

ARC

Random-cluster model

- Focus on two observables:



$$\mathcal{N}(A) = |A|$$

$$\mathcal{S}_2(A) = \sum_{\text{clusters } \mathcal{C} \text{ in } (V,A)} |\mathcal{C}|^2$$

- $\mathcal{N}$ is an "energy"
- $\chi = \langle \mathcal{S}_2 \rangle / V$ is the mean cluster size, or "susceptibility"
- Chayes-Machta 1997 devised a cluster algorithm valid for all real $q \geq 1$
  - Simulates a coupled measure of bond and vertex variables
  - Equivalent to Swendsen-Wang when q is an integer

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random
  - Determine if $x \leftrightarrow y$ via a path not including xy

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random
  - Determine if $x \leftrightarrow y$ via a path not including xy
    - If so, occupy xy with probability $v/(1+v)$

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random
  - Determine if $x \leftrightarrow y$ via a path not including xy
    - If so, occupy xy with probability $v/(1+v)$
    - If not, occupy xy with probability $v/(q+v)$

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random
  - Determine if $x \leftrightarrow y$ via a path not including xy
    - If so, occupy xy with probability $v/(1+v)$
    - If not, occupy xy with probability $v/(q+v)$
- Valid for all real $q \geq 0$

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random
  - Determine if $x \leftrightarrow y$ via a path not including xy
    - If so, occupy xy with probability $v/(1+v)$
    - If not, occupy xy with probability $v/(q+v)$
- Valid for all real $q \geq 0$
- Need an efficient way to check connectivity. . .

Sweeny's algorithm

- The heat-bath version proceeds as follows:
  - Start with some configuration $A \subseteq E$
  - Choose an edge $xy \in E$ uniformly at random
  - Determine if $x \leftrightarrow y$ via a path not including xy
    - If so, occupy xy with probability $v/(1+v)$
    - If not, occupy xy with probability $v/(q+v)$
- Valid for all real $q \geq 0$
- Need an efficient way to check connectivity...
- How do we measure the efficiency of an MCMC algorithm?
- Compare Sweeny with Chayes-Machta

General setting for MCMC

- Irreducible, aperiodic, reversible Markov chain
  - State space S, with $|S| < \infty$
  - Transition matrix P
  - Stationary distribution π
- Observable (random variable) X
  - E.g. $X = \mathcal{N}$ or $\mathcal{S}_2 \ldots$
- Simulate Markov chain $\implies$ time series $X_0, X_1, \ldots$
- Define the autocorrelation function

$$\rho_X(t) := \frac{\langle X_s X_{s+t} \rangle_\pi - \langle X \rangle_\pi^2}{\text{var}_\pi(X)}$$

- Stationary process – start "in equilibrium" (or wait "long enough")

Autocorrelation times

- We must consider two distinct autocorrelation times
- The integrated autocorrelation time

$$\tau_{\text{int},X} := \frac{1}{2} \sum_{t=-\infty}^{\infty} \rho_X(t)$$

- If $\widehat{X}$ is the sample mean of $\{X_t\}_{t=1}^{T}$ then we have

$$\text{var}(\widehat{X}) \sim 2\,\tau_{\text{int},X} \frac{\text{var}(X)}{T}, \qquad T \to \infty$$

- We get one "effectively independent" observation every $2\,\tau_{\text{int},X}$ time steps

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Autocorrelation times

- $\rho_X(t)$ typically decays exponentially as $t \to \infty$
- The exponential autocorrelation time

$$\tau_{\exp,X} := \limsup_{t\to\infty} \frac{t}{-\log|\rho_X(t)|} \quad \text{and} \quad \tau_{\exp} := \sup_X \tau_{\exp,X}$$

- Typical observables have $\tau_{\exp,X} = \tau_{\exp}$
- *Nice* chains with $|S| < \infty$ have $\tau_{\exp} < \infty$
- $\tau_{\text{int},X} \leq \tau_{\exp}$ for all X (need NOT be equal)
- Start the chain with arbitrary distribution α
  - Distribution at time t is αP^t
  - αP^t tends to π with rate bounded by $e^{-t/\tau_{\exp}}$

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics
and Statistics of Complex Systems

ARC

Critical slowing-down

- Near a critical point the autocorrelation times typically diverge like

$$\tau \sim \xi^z$$

- More precisely, we have a family of exponents:

  z_{exp}, and $z_{\text{int},X}$ for each observable X.

- Different algorithms for the same model can have very different z

- E.g. $d = 2$ Ising model
  - Glauber (Metropolis) algorithm $z \approx 2$
  - Swendsen-Wang algorithm $z \approx 0.2$

Back to Sweeny's algorithm

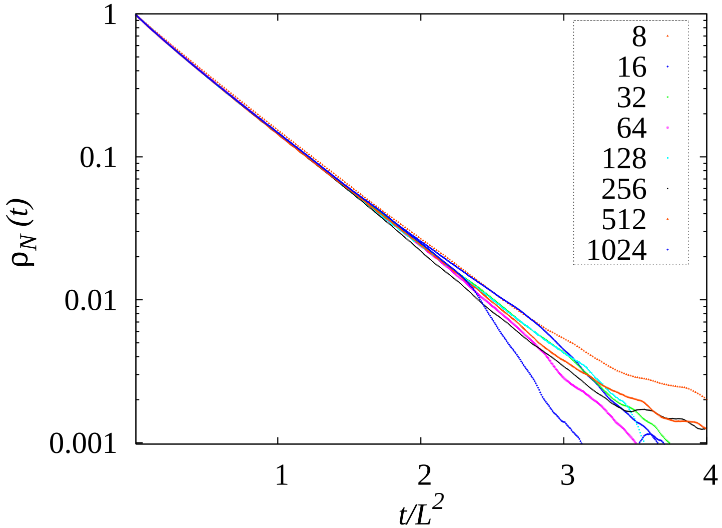
- Simulated the $d = 2$ critical random-cluster model
  - On $L \times L$ square lattice
  - Simulated a number of values of $0 \leq q \leq 4$
  - Measured:
    - $\mathcal{N}(A) = |A|$
    - $\mathcal{S}_2(A) = \sum_{\text{clusters } \mathcal{C} \text{ in } (V,A)} |\mathcal{C}|^2$
  - Measured observables after every hit

    – i.e. every bond update
  - Natural unit of time is one sweep

    – i.e. L^d hits
  - Cluster algorithms perform one sweep every iteration

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Dynamics of $\mathcal{N}$

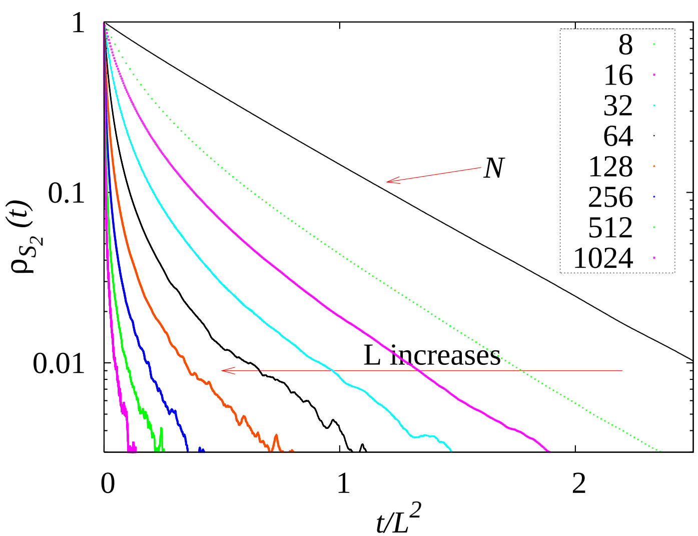


Plot shows $q = 0.2$ and $8 \leq L \leq 1024$

Suggests $\tau_{\text{exp}} \sim L^2$ hits

- Good data collapse
- Empirically $z_{\text{exp}} = 0$ for $q \lessapprox 2$
- $\rho_{\mathcal{N}}(t)$ is almost a perfect exponential
- Li-Sokal bound: $z_{\text{exp}}, z_{\text{int},\mathcal{N}} \geq \alpha/\nu$
  - Applies to Sweeny and Chayes-Machta
  - Empirically $z^{\text{Sweeny}} \approx z^{\text{Chayes-Machta}}$

Dynamics of $\mathcal{S}_2$



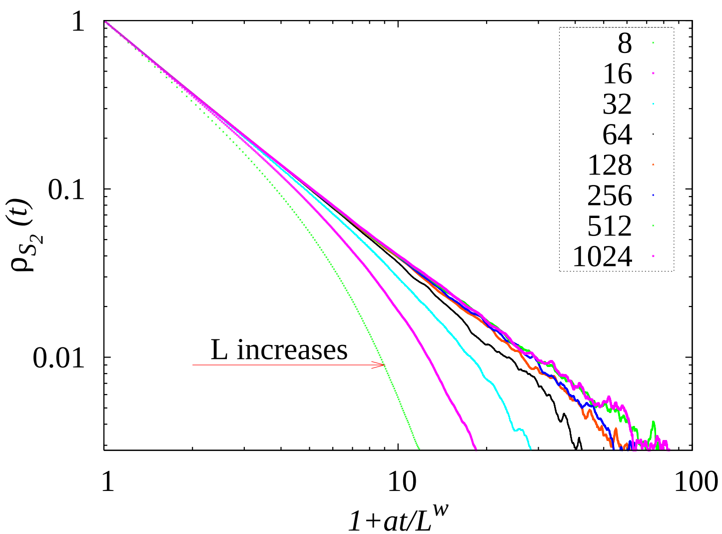
- Plot shows $q = 0.2$ and $8 \leq L \leq 1024$
- $\rho_{\mathcal{N}}(t)$ is shown for comparison

- $\rho_{\mathcal{S}_2}(t)$ decays significantly in a time much less than one sweep
- Critical speeding-up

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

ARC

Critical speeding-up



- Plot shows $q = 0.2$ and $8 \leq L \leq 1024$
- $\rho_{\mathcal{S}_2}(t)$ vs $1 + at/L^w$ with $w = 0.99$

- Good data collapse
- $\mathcal{S}_2$ exhibits strong decorrelation on a time scale $O(L^w)$ hits
- Initial decay $\rho_{\mathcal{S}_2}(t) = f(t/L^w)$ with $f(x) \sim x^{-r}$
- Empirically $w < d$ for $q \lessapprox 2$

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics
and Statistics of Complex Systems

Some hand-waving...

- Critical FK clusters are fractal
  - $O(1)$ bond deletions can split a large cluster into two large clusters
  - $O(1)$ bond additions can join two large clusters
- There are $O(L^{d_{\text{red}}})$ edges whose removal would split a big cluster
- There are $O(L^{d_{\text{red}}})$ edges whose addition would connect two big clusters
  - d_{red} is the red bond exponent
  - Coniglio 1989 gives d_{red} for all $0 \leq q \leq 4$ in $d = 2$

A conjecture...

- “The decorrelation of $\mathcal{S}_2$ is due to hitting $O(1)$ red bonds”
  - This takes time $O(L^{d-d_{\mathrm{red}}})$
  - So $w = d - d_{\mathrm{red}}$

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

A conjecture...

- "The decorrelation of $\mathcal{S}_2$ is due to hitting $O(1)$ red bonds"
  - This takes time $O(L^{d-d_{\mathrm{red}}})$
  - So $w = d - d_{\mathrm{red}}$

q	z_{exp}	α/ν	w	d_{red}
0.0005	0	-1.9576	0.77	1.2376
0.005	0	-1.8679	0.79	1.2111
0.05	0	-1.6005	0.88	1.1299
0.2	0	-1.2467	0.99	1.0168
0.5	0	-0.8778	1.11	0.8904
1.0	0	-0.5000	1.26	0.7500
1.5	0	-0.2266	1.36	0.6398
2.0	0 (log)	0 (log)	1.49	0.5417
2.5	0.26(1)	0.2036	1.64	0.4474
3.0	0.45(1)	0.4000	1.84	0.3500
3.5	0.636(2)	0.6101	2.04	0.2375

Summary of Sweeny results

- Critical slowing down is absent for small q
- $z_{\exp}, z_{\text{int},\mathcal{N}}$ comparable to their Chayes-Machta values
- $\mathcal{S}_2$ exhibits critical speeding-up for a wide range of q
  - This can lead to $z_{\text{int},\mathcal{S}_2} < 0$
  - Estimating $z_{\text{int},\mathcal{S}_2}$ is tricky ...
- Critical speeding-up and slowing-down can coexist
- All this holds in $d = 3$ too
- It is conceivable that *most* dynamics have a multiple time-scale behavior...

Worm algorithms

How can we simulate the Ising model?

- Glauber dynamics
  - Flip one Ising spin at a time
  - Severe critical slowing-down
- Sweeny dynamics
  - Transform Ising model to $q = 2$ random-cluster model
  - Flip one FK bond at a time
  - Weak critical slowing-down
- Swendsen-Wang (Chayes-Machta) dynamics
  - Transform Ising model to $q = 2$ random-cluster model
  - Simulate joint model of Ising spins and FK bonds
  - Weak critical slowing-down

Worm algorithms

- Worm dynamics
  - Prokof'ev & Svistunov PRL 2001
  - Transform Ising model to high-temperature graphs
  - Simulate high-temperature graphs via *local* moves
  - *worm diffusion*
- Consider simplest case
  - ferromagnetic, zero field, nearest-neighbor, on L^d

State space for worm dynamics

- Fix a finite graph $G = (V, E)$
- For $A \subseteq E$ let ∂A be the set of all vertices with odd degree in (V, A)
- For distinct $x, y \in V$ define

$$\mathcal{S}_{x,y} = \{A \subseteq E | \partial A = \{x, y\}\}$$

and let

$$\mathcal{S}_{x,x} = \{A \subseteq E | \partial A = \emptyset\}$$

- $\mathcal{S}_{x,x}$ is just the cycle space $\mathcal{C}(G)$
- Configuration space of our worm algorithm is

$$\mathcal{S} = \{(A, x, y) | x, y \in V \text{ and } A \in \mathcal{S}_{x,y}\}$$

High temperature expansions

- The standard Ising high-temperature expansions are:

$$Z = \sum_{A \in \mathcal{S}_{x,x}} w^{|A|} \qquad \text{Partition function}$$

$$Z\langle \sigma_x \sigma_y \rangle = \sum_{A \in \mathcal{S}_{x,y}} w^{|A|} \qquad \text{Two-point function}$$

$$Z\langle \mathcal{M}^2 \rangle = \sum_{A \in \mathcal{S}} w^{|A|} \qquad \text{Magnetization}$$

- $\mathcal{M}(\sigma) = \sum_{x \in V} \sigma_x$ is the Ising magnetization
- $w = \tanh(\beta)$
- $0 \leq w \leq 1$

Worm algorithm

- The elementary move of the worm algorithm is as follows:

Worm algorithm

- The elementary move of the worm algorithm is as follows:
  - Start in configuration (A, x, y)

Worm algorithm

- The elementary move of the worm algorithm is as follows:
  - Start in configuration (A, x, y)
  - Pick uniformly at random either x or y (say, x)

Worm algorithm

- The elementary move of the worm algorithm is as follows:
  - Start in configuration (A, x, y)
  - Pick uniformly at random either x or y (say, x)
  - Pick uniformly at random some $x' \sim x$ (in G)

Worm algorithm

- The elementary move of the worm algorithm is as follows:
  - Start in configuration (A, x, y)
  - Pick uniformly at random either x or y (say, x)
  - Pick uniformly at random some $x' \sim x$ (in G)
  - Propose moving to $(A \triangle xx', x', y)$

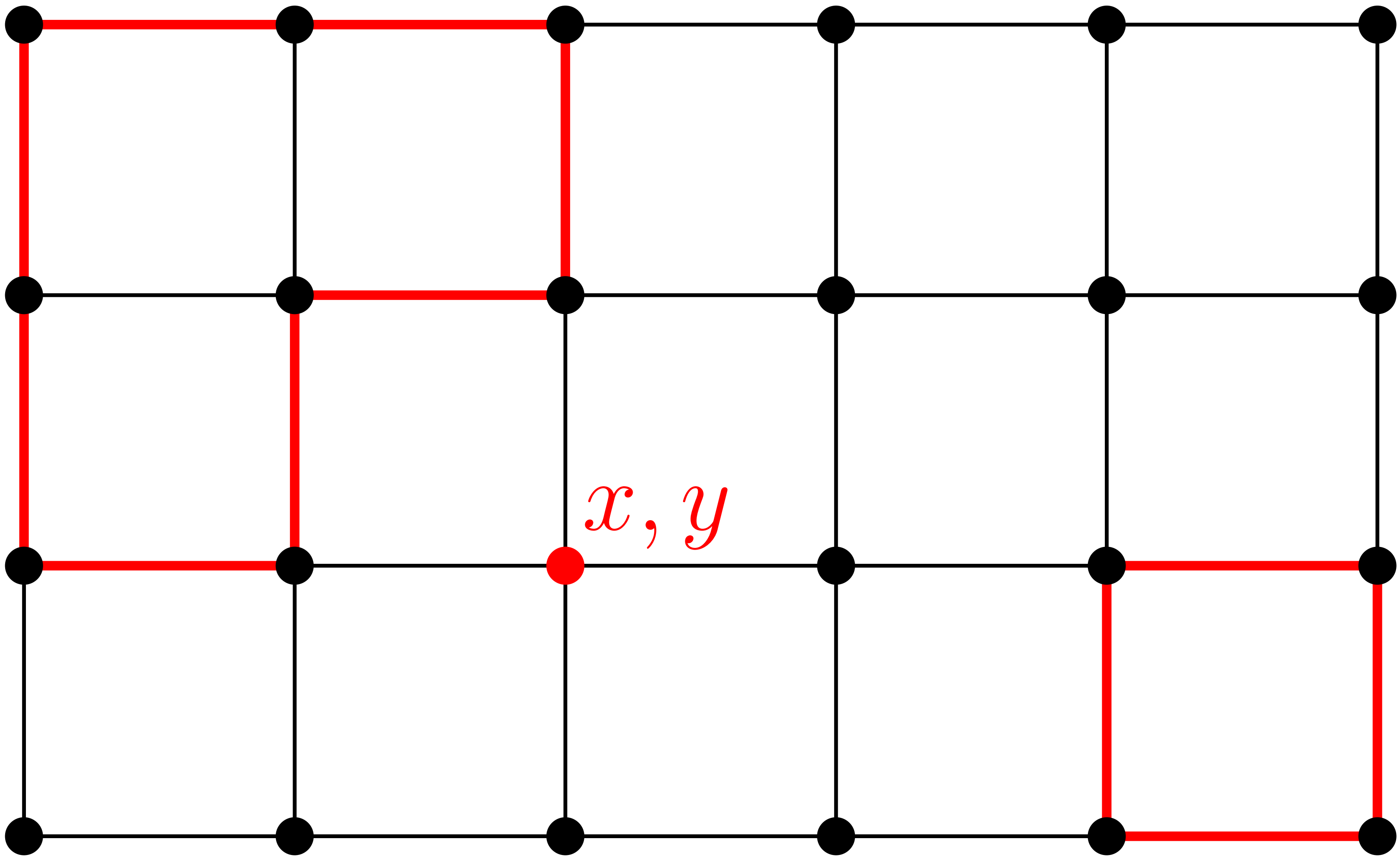
Worm algorithm

- The elementary move of the worm algorithm is as follows:
  - Start in configuration (A, x, y)
  - Pick uniformly at random either x or y (say, x)
  - Pick uniformly at random some $x' \sim x$ (in G)
  - Propose moving to $(A \triangle xx', x', y)$
    - If proposed transition would add an edge accept with probability w

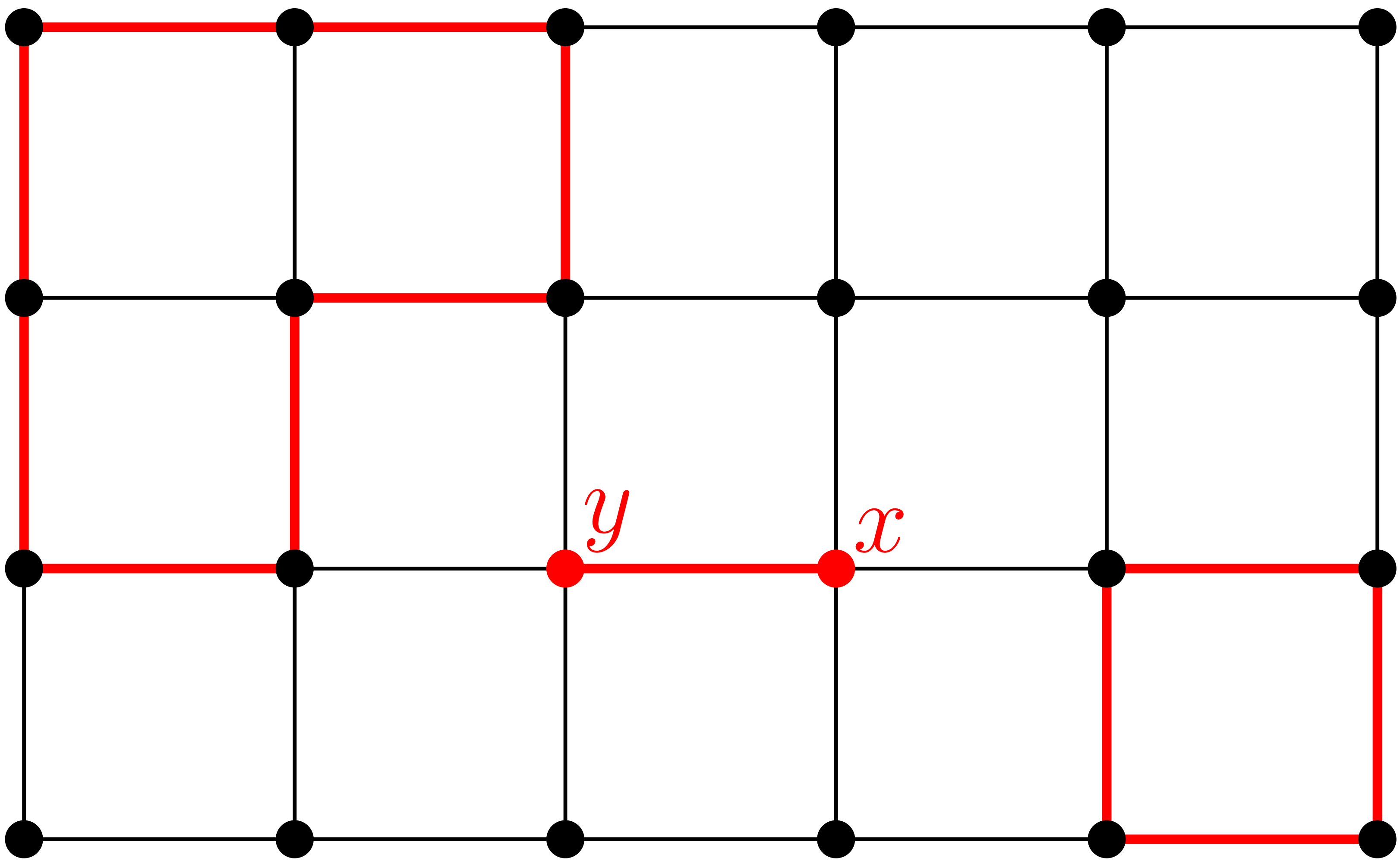
Worm algorithm

- The elementary move of the worm algorithm is as follows:
  - Start in configuration (A, x, y)
  - Pick uniformly at random either x or y (say, x)
  - Pick uniformly at random some $x' \sim x$ (in G)
  - Propose moving to $(A \triangle xx', x', y)$
    - If proposed transition would add an edge accept with probability w
    - If proposed transition would remove an edge accept with probability 1

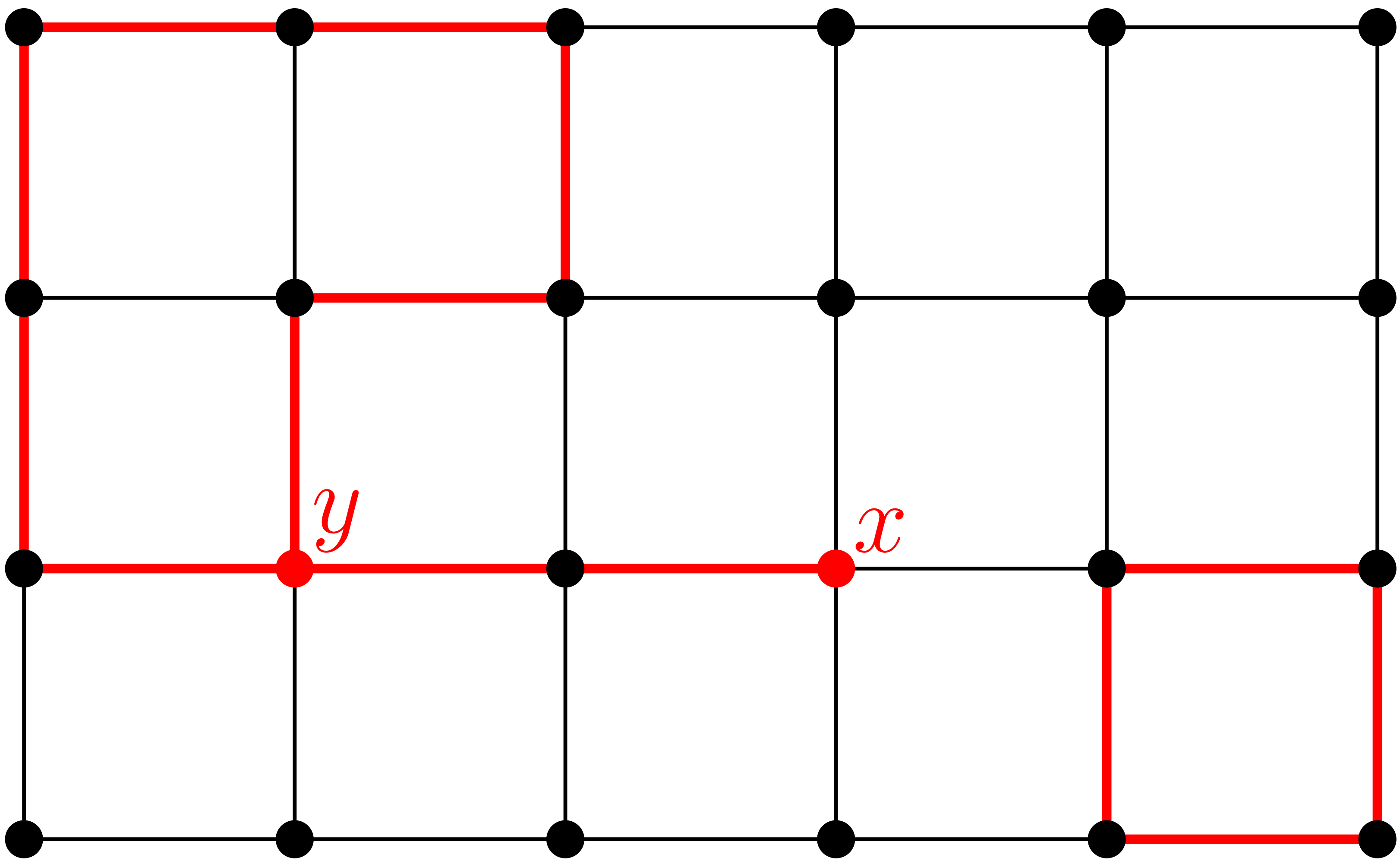
Worm dynamics $t = 0$



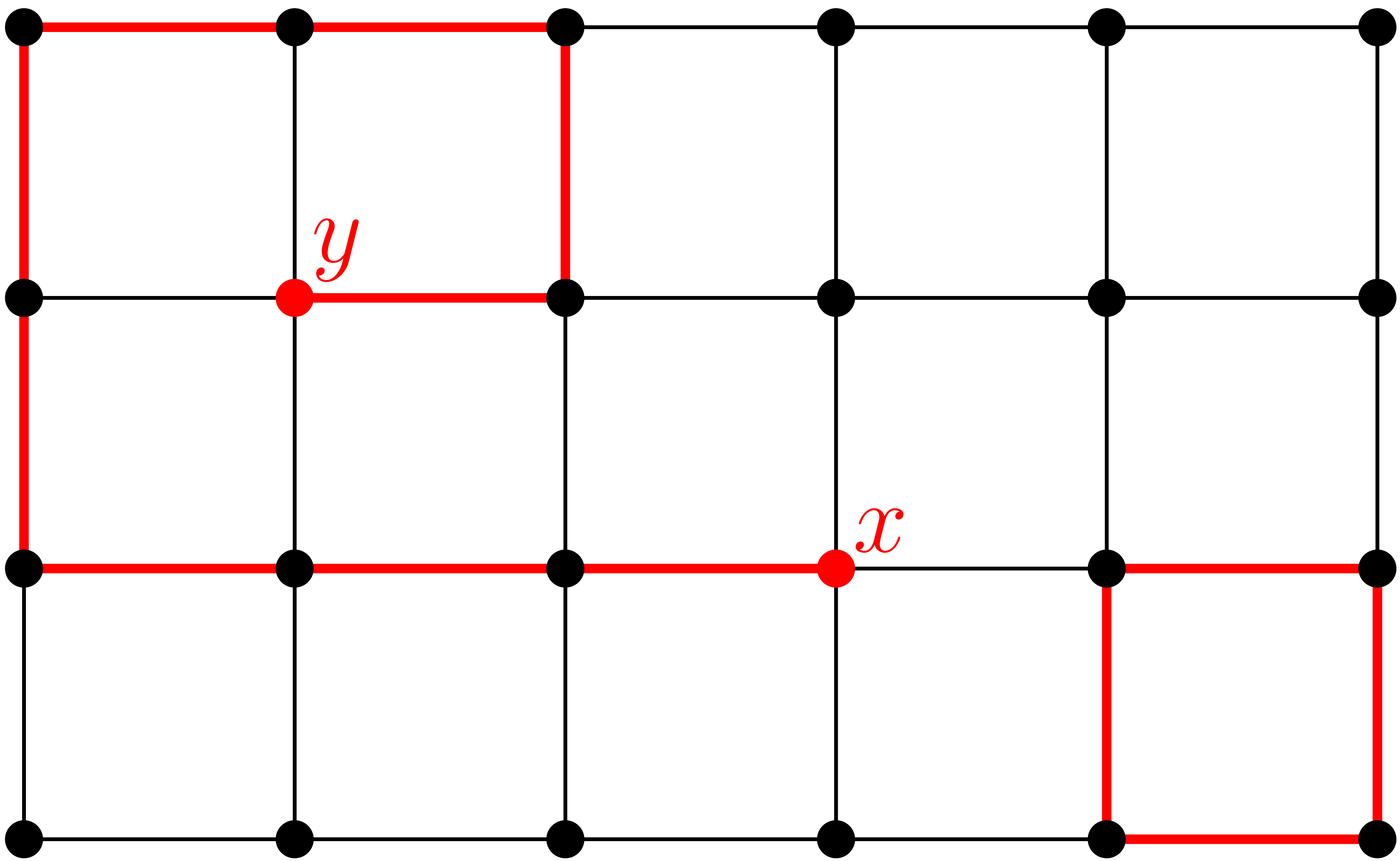
Worm dynamics $t = 1$



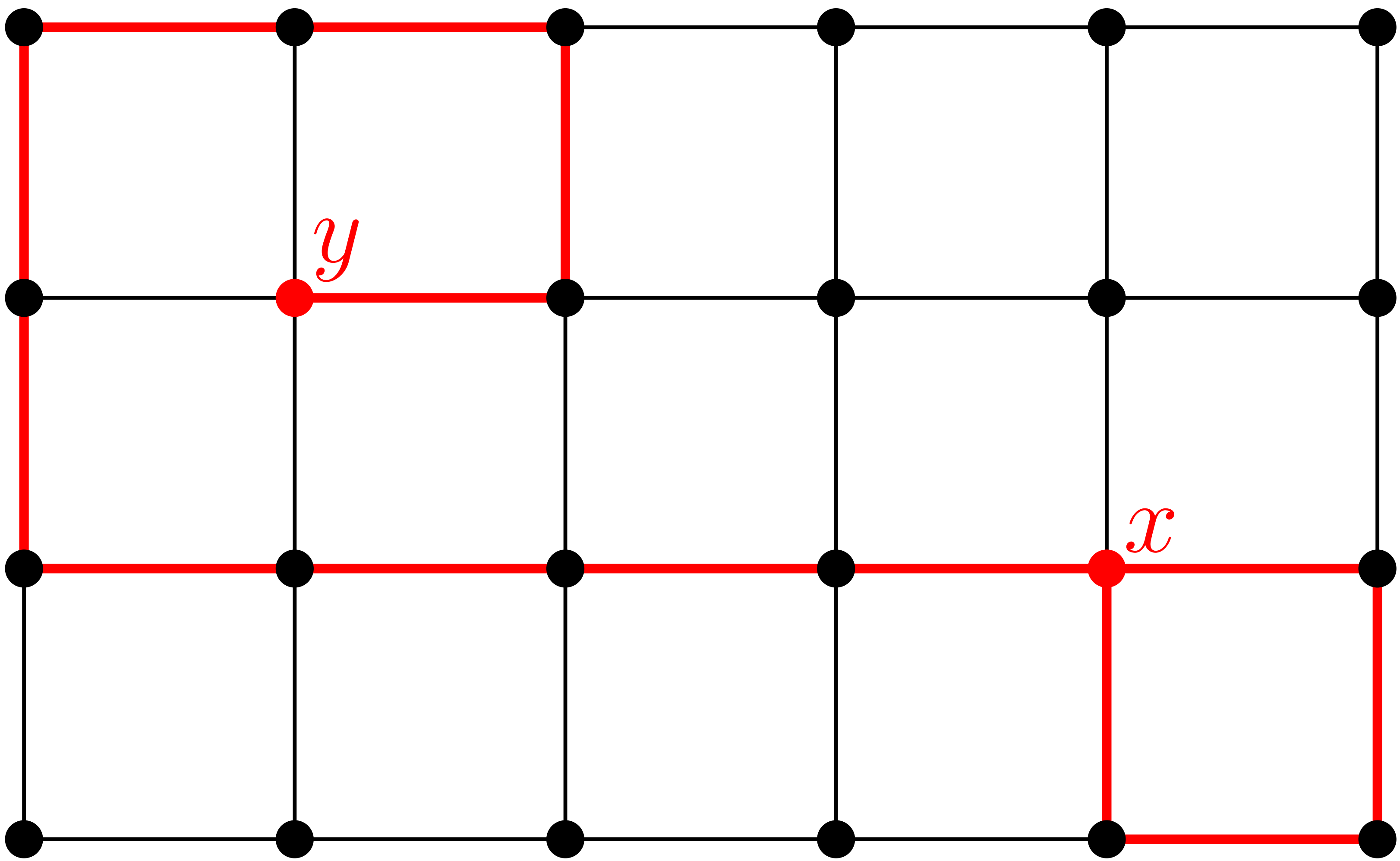
Worm dynamics $t = 2$



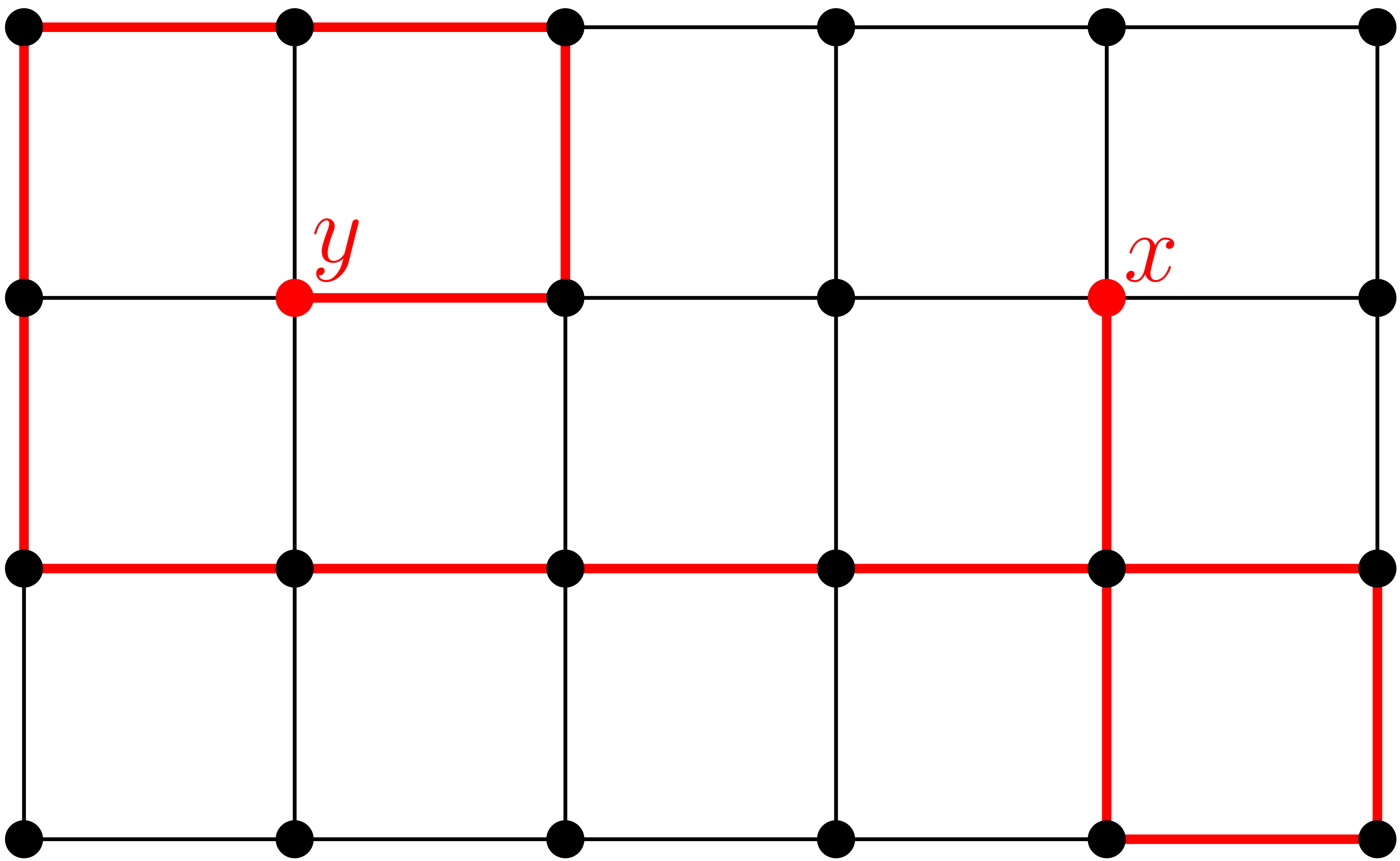
Worm dynamics $t = 3$



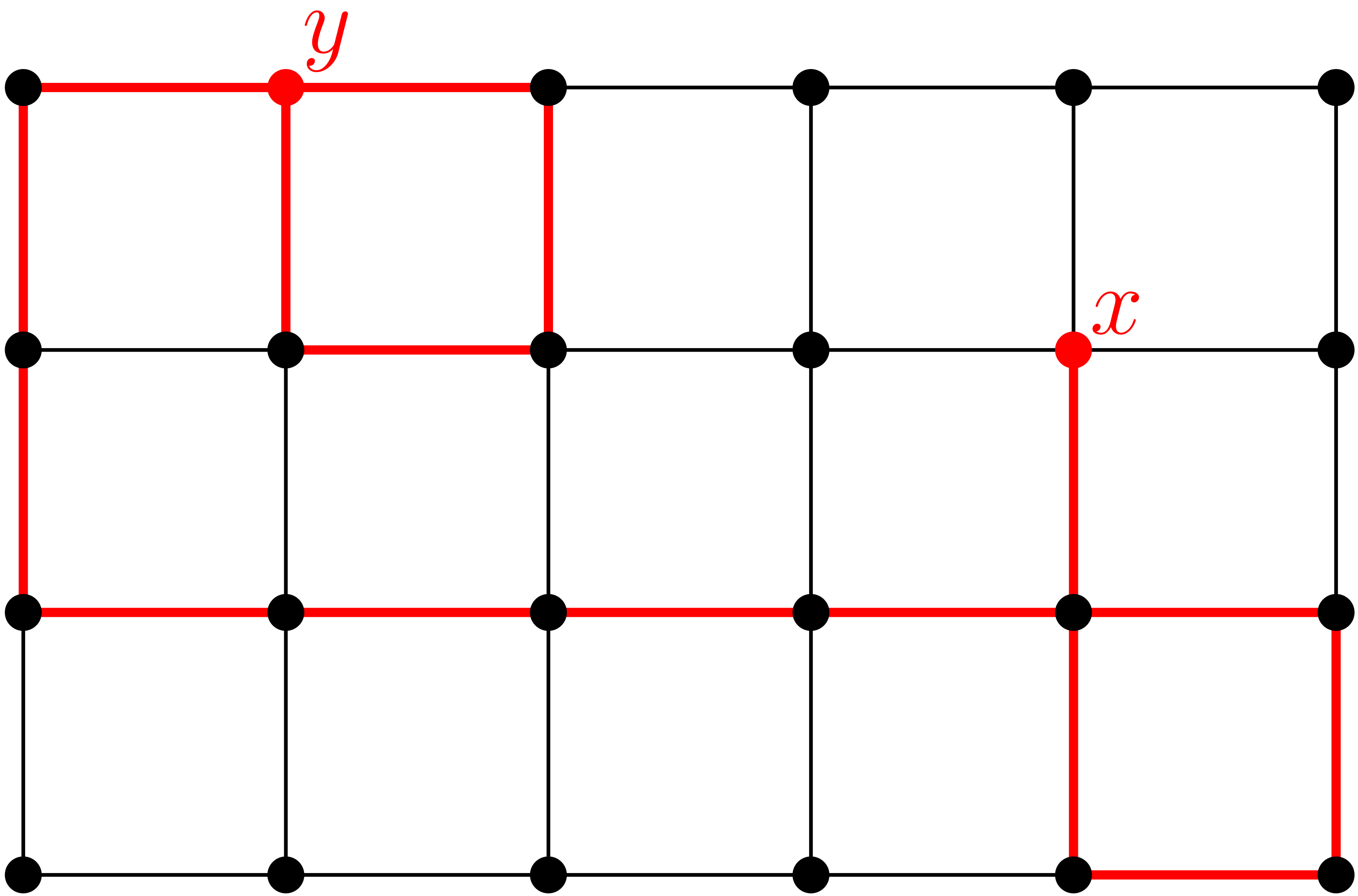
Worm dynamics $t = 4$



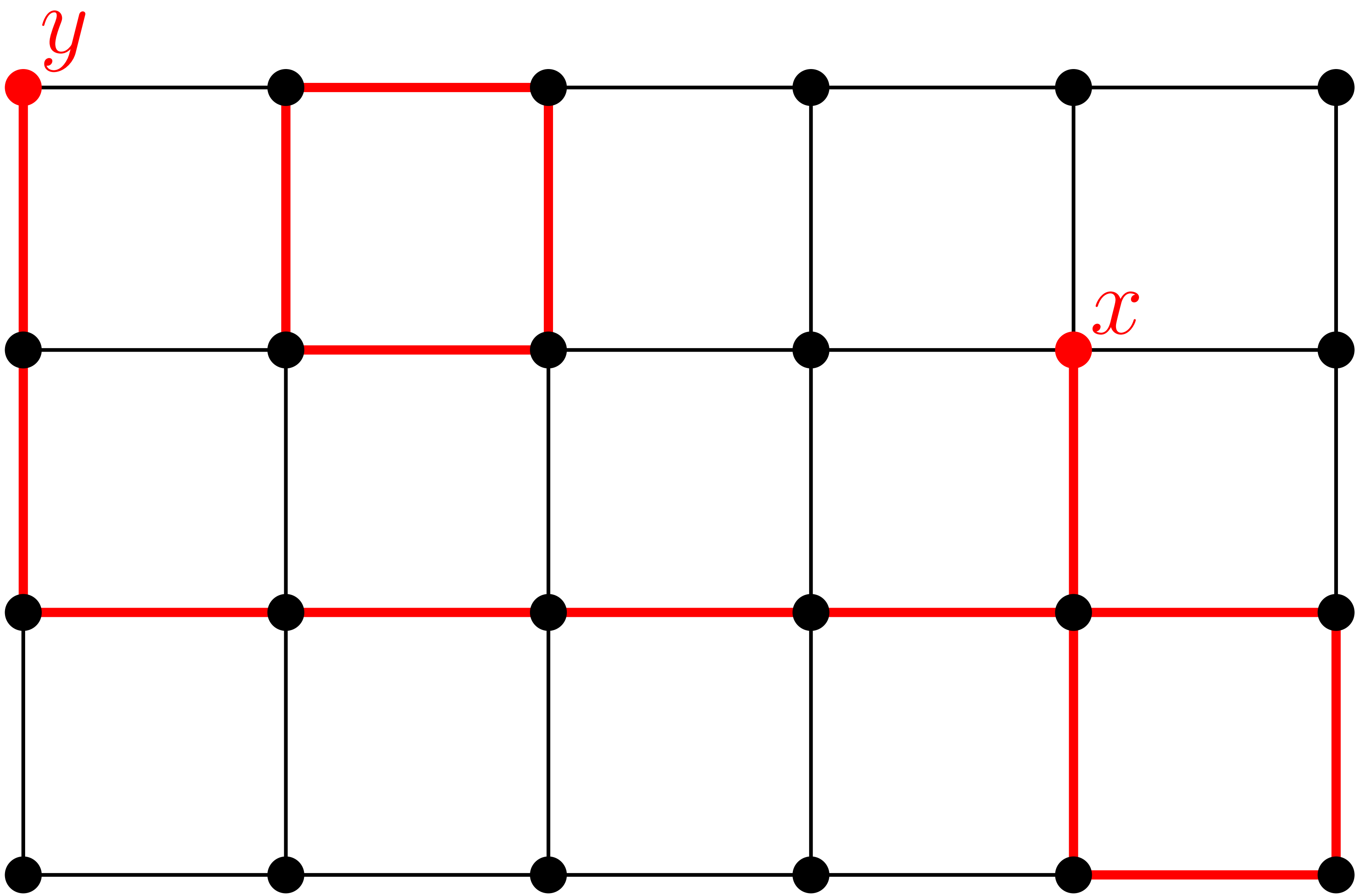
Worm dynamics $t = 5$



Worm dynamics $t = 6$



Worm dynamics $t = 7$



Transition matrix

- Let G be a regular lattice of coordination number z
- Transition matrix P on $\mathcal{S}$ is

$$P[(A,x,y) \to (A\triangle xx', x', y)] = \frac{1}{2}\,\frac{1}{z}\begin{cases} 1, & xx' \in A, \\ w, & xx' \notin A, \end{cases}$$

- And similarly for y moves...
- All other non-diagonal elements of P are zero
- P is in detailed balance with $\pi(A,x,y) = w^{|A|}/Z\langle\mathcal{M}^2\rangle$
- For translation invariant systems $\langle\mathcal{M}^2\rangle = V\,\chi$

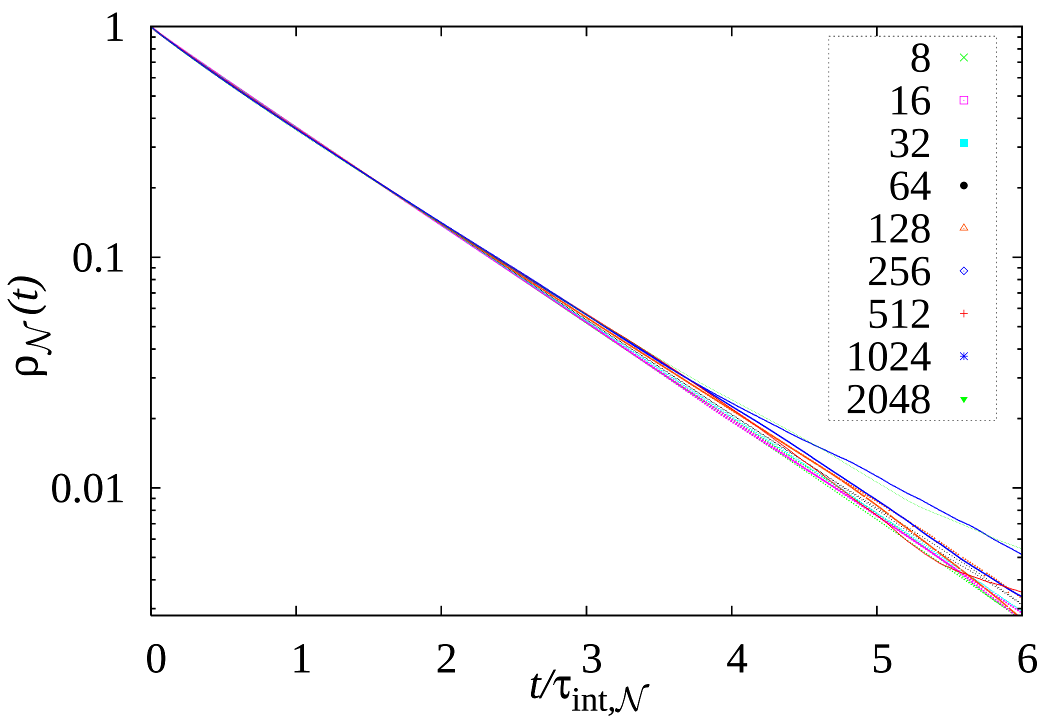
Observables

- Focus on two observables:
  - $\mathcal{N}(A, x, y) = |A|$
  - $\mathcal{D}_0(A, x, y) = \delta_{x,y}$
- $\langle \mathcal{D}_0 \rangle_\pi$ is simply related to χ

$$
\begin{aligned}
\langle \mathcal{D}_0 \rangle_\pi &= \frac{1}{Z\,V\,\chi} \sum_{(A,x,y)\in\mathcal{S}} w^{|A|} \delta_{x,y} \\
&= 1/\chi
\end{aligned}
$$

- Measured observables after every hit (worm update)
- Natural unit of time is one sweep (L^d hits)

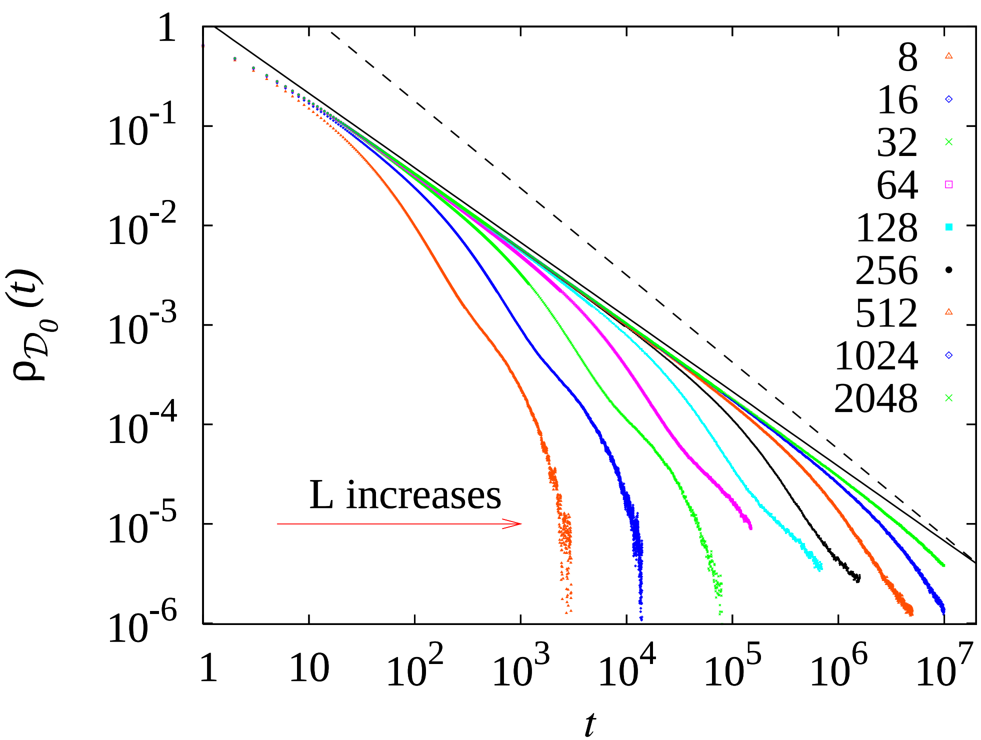
Dynamics of $\mathcal{N}$



- Critical Ising model $d = 2$
- Good data collapse
- $z_{\text{exp}} \approx z_{\text{int},\mathcal{N}} \approx 0.379$

- $\rho_{\mathcal{N}}(t)$ is almost a perfect exponential
- Li-Sokal bound $z_{\text{exp}}, z_{\text{int},\mathcal{N}} \geq \alpha/\nu$ applies to worm too

Dynamics of $\mathcal{D}_0$

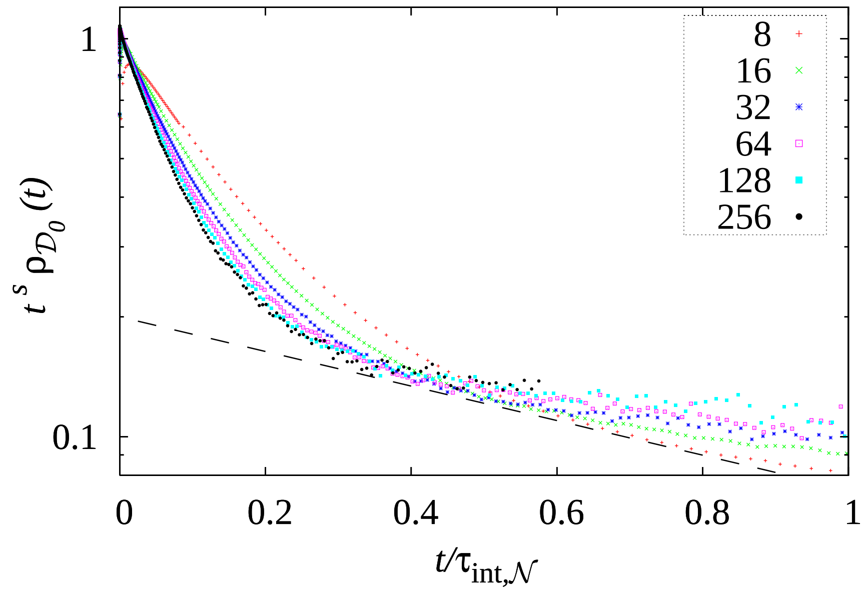


Critical Ising model $d = 2$

- $\rho_{\mathcal{D}_0}(t)$ decays significantly in $O(1)$ hits!
- $\rho_{\mathcal{D}_0}(t) \sim t^{-s}$ with $s \approx 0.75$

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Crossover



Plot $t^s\,\rho_{\mathcal{D}_0}(t)$ versus $t/\tau_{\text{int},\mathcal{N}}$

Reasonable data collapse

- Postulate $\rho_{\mathcal{D}_0}(t) = g(t)h(t/L^{d+z_{\exp}})$ with $g(t) \sim t^{-s}$ and $s < 1$

$$\implies z_{\text{int},\mathcal{D}_0} = -sd + (1-s)z_{\exp}$$

- Gives $z_{\text{int},\mathcal{D}_0} \approx -1.42$

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Three dimensions

Qualitatively similar behavior when $d = 3$:

- $\rho_{\mathcal{D}_0}(t) \sim t^{-s}$
- $s \approx 0.66$
- Implies $z_{\text{int},\mathcal{D}_0} \approx -1.92$
- $\rho_{\mathcal{N}}(t)$ roughly exponential
- $z_{\text{exp}} \approx z_{\text{int},\mathcal{N}} \approx \alpha/\nu \approx 0.174$
- Li-Sokal bound may be sharp for $d = 3$ worm algorithm
- Compare Swendsen-Wang $z_{SW} \approx 0.46$

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Practical efficiency

- Swendsen-Wang seems to outperform worm when $d = 2$
- Efficiency depends on observable, X
- A simple way to compare worm and SW is to compute $\kappa = \sigma^2_{\widehat{X}} T_{CPU}$ for both algorithms
- When $d = 3$ and $X = \chi$ we find $\kappa_{worm}/\kappa_{SW} \approx L^{-0.33}$
  - With the crossover $\kappa_{worm}/\kappa_{SW} \approx 1$ at around $L \approx 20$
- There is also a natural worm estimator for ξ
- Again SW outperforms worm when $d = 2$
  - For $d = 3$ we find $\kappa_{worm}/\kappa_{SW} \approx L^{-0.32}$
  - With the crossover $\kappa_{worm}/\kappa_{SW} \approx 1$ at around $L \approx 45$

Conclusions

- Locality is not a sufficient condition for “badness”
- Sweeny’s algorithm has comparable efficiency to Chayes-Machta
- For $q \lessapprox 2$ Sweeny’s algorithm exhibits critical speeding-up i.e. significant decorrelation in $O(L^w)$ hits with $w < d$
- We can predict w if $\alpha/\nu < 0$ (no critical slowing down)
- The worm algorithm also exhibits decorrelation on multiple time scales
- The worm algorithm outperforms Swendsen-Wang for $d = 3$ Ising model for measuring χ and ξ