Phase Transitions and Emergent Phenomena in Algorithms and Applications

Dana Randall
Georgia Institute of Technology

Georgia Tech Institute for Data Engineering and Science

- IDEaS is a new "Interdisciplinary Research Institute"
- Launched July 1, 2016
- Executive Directors: Srinivas Aluru and Dana Randall

Foundations:

- Machine Learning
- Cyber-infrastructure
- Algorithms & Optimization
- Signal Processing
- Policy
- Security

Domains:

- Medicine & Health
- Energy
- Materials
- Smart Cities
- Business Analytics
- Social Computing

- http://ideas.gatech.edu

Georgia Tech Institute for Data Engineering and Science

SOUTHBDHUB

Big Data Sharing and Infrastructure

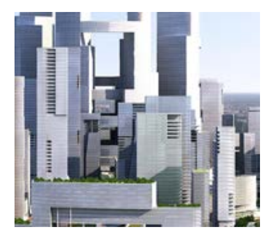
Healthcare

Coastal Hazards

Economics, Privacy and Policy Issues

Industrial Big Data

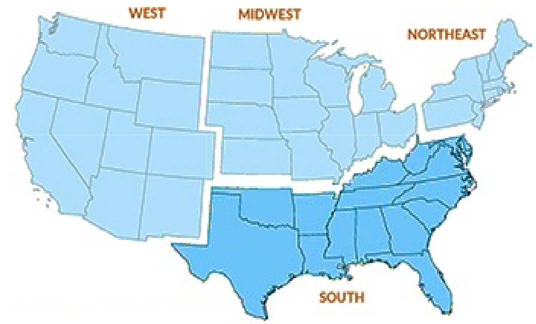
Materials and Manufacturing



Habitat Planning

Georgia Tech Institute for Data Engineering and Science

SOUTHBDHUB



Introducing *CODA*

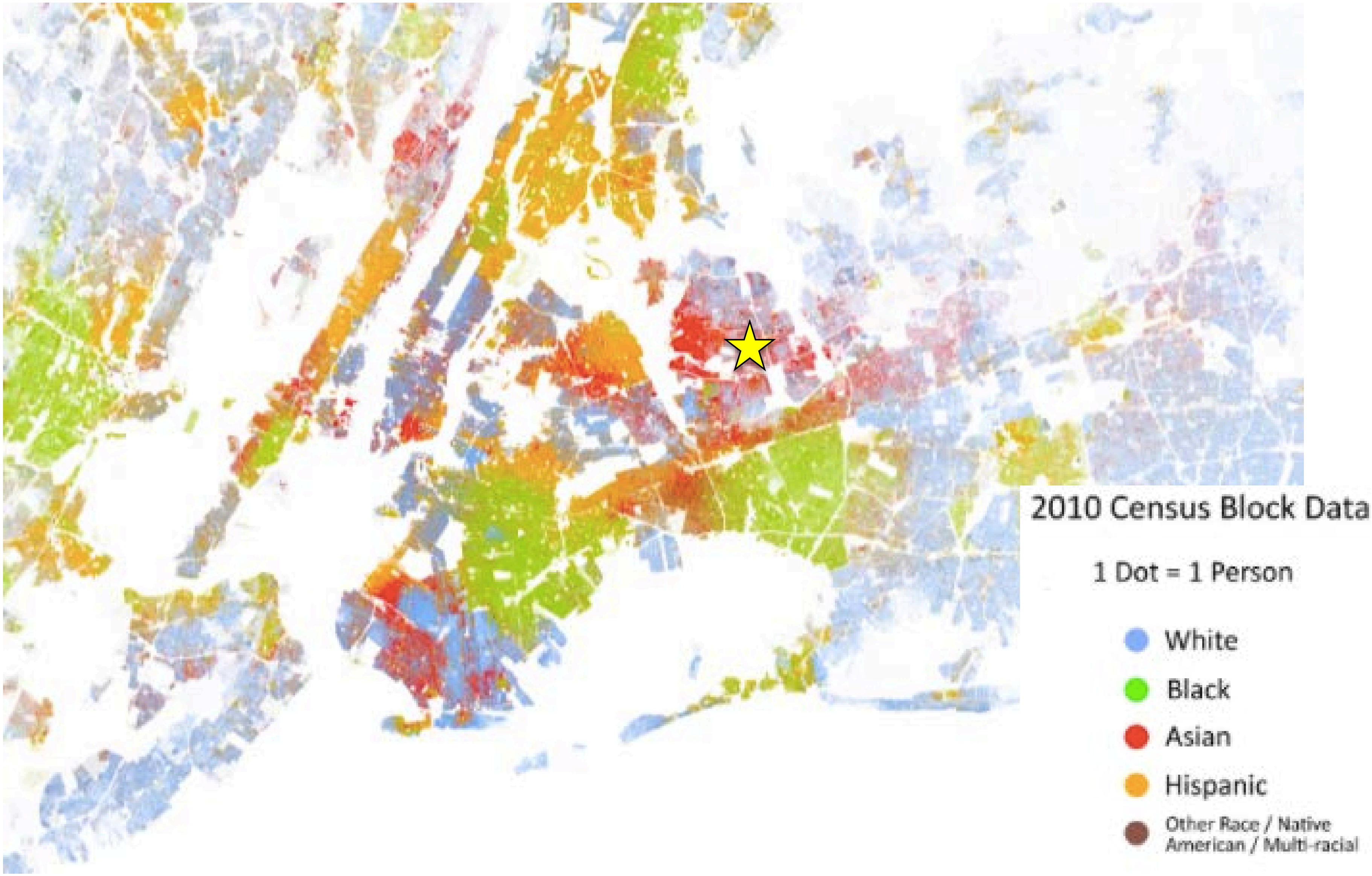
- Midtown Atlanta
- 750K sq ft mixed use
- 80K sq ft data center
- $375 million investment
- GT will occupy half of the office space; rest industry

Phase Transitions and Emergent Phenomena in Algorithms and Applications

Dana Randall
Georgia Institute of Technology

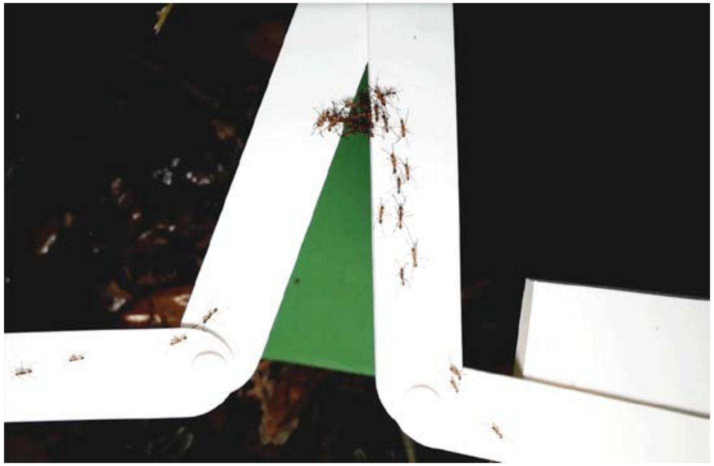
Demographic Data

- http://demographics.coopercenter.org/DotMap/



Ants and Swarm Robotics

Ants build bridges to shorten distance other ants must travel:



Bridge length is a function of angle: optimize tradeoff between shorter path and more ants to traverse path

Reid, Lutz, Powell, Kao, Couzin, and Garnier. Army ants dynamically adjust living bridges in response to a cost-benefit trade-off. *Proceedings of the National Academy of Sciences*, 112(49):15113-15118, 2015.

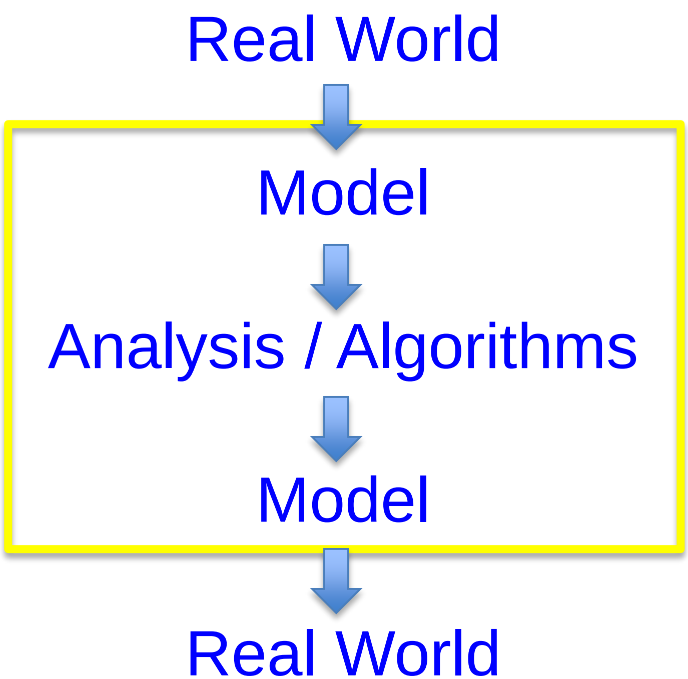
Many systems proposed and realized recently:

- Modular robotics
- Swarm robotics
- DNA computing
- Smart materials

kilobots

Understanding Data

<u>Goals</u>: Identify emergent behaviors occurring in data
Provide mechanisms to explain emergent behavior.



Algorithms

+

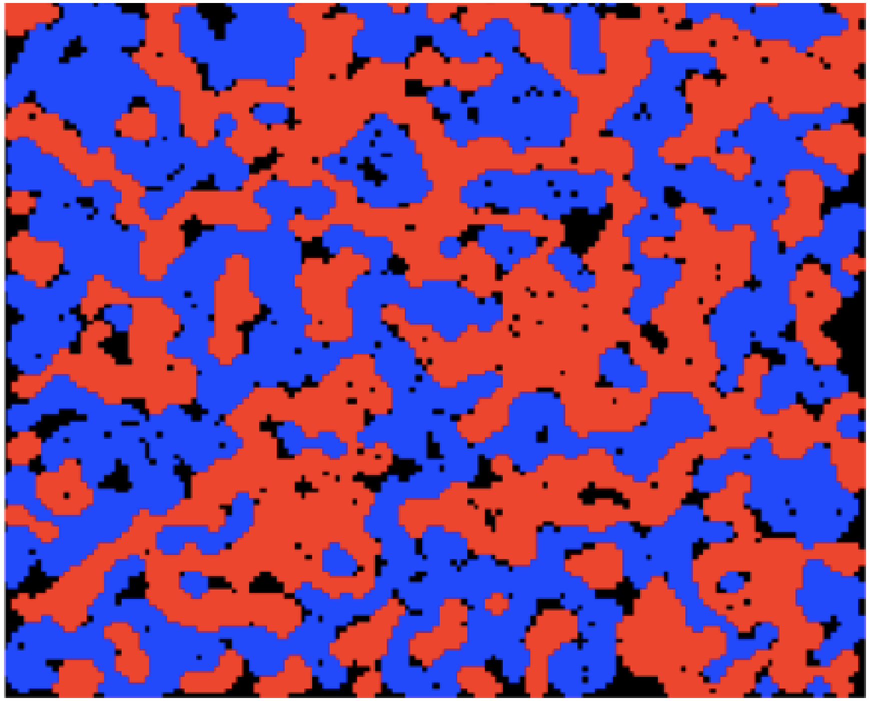
Statistical Physics

Randomized Algorithms & Large Data

I. Demographics: The Schelling Segregation Model '71

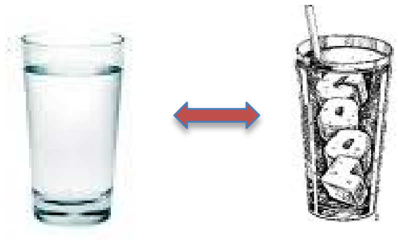
"Micro-Motives determine Macro-Behavior"

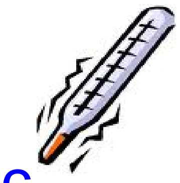
- Houses are colored red or blue
- People move if they have too many neighbors of the opposite color



Randomized Algorithms & Large Data

II. Physics: Phase transitions

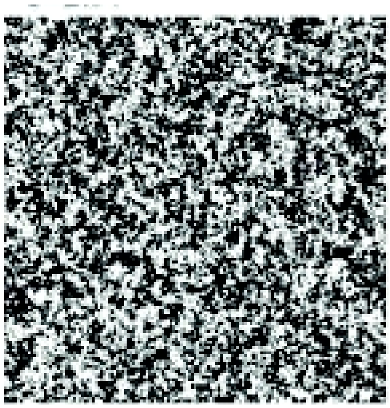




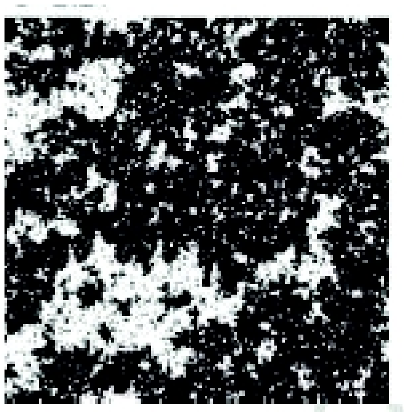
Macroscopic changes to the system due to a microscopic change to some parameter.

e.g.: gas/liquid/solid, spontaneous magnetization

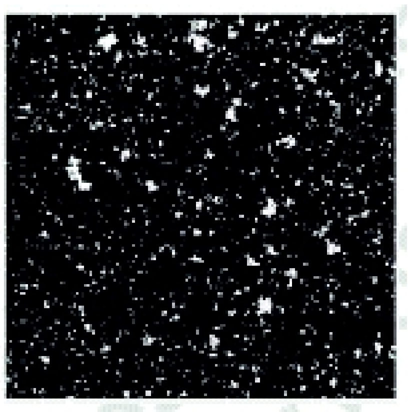
Simulations of the Ising model



High temperature



Criticality

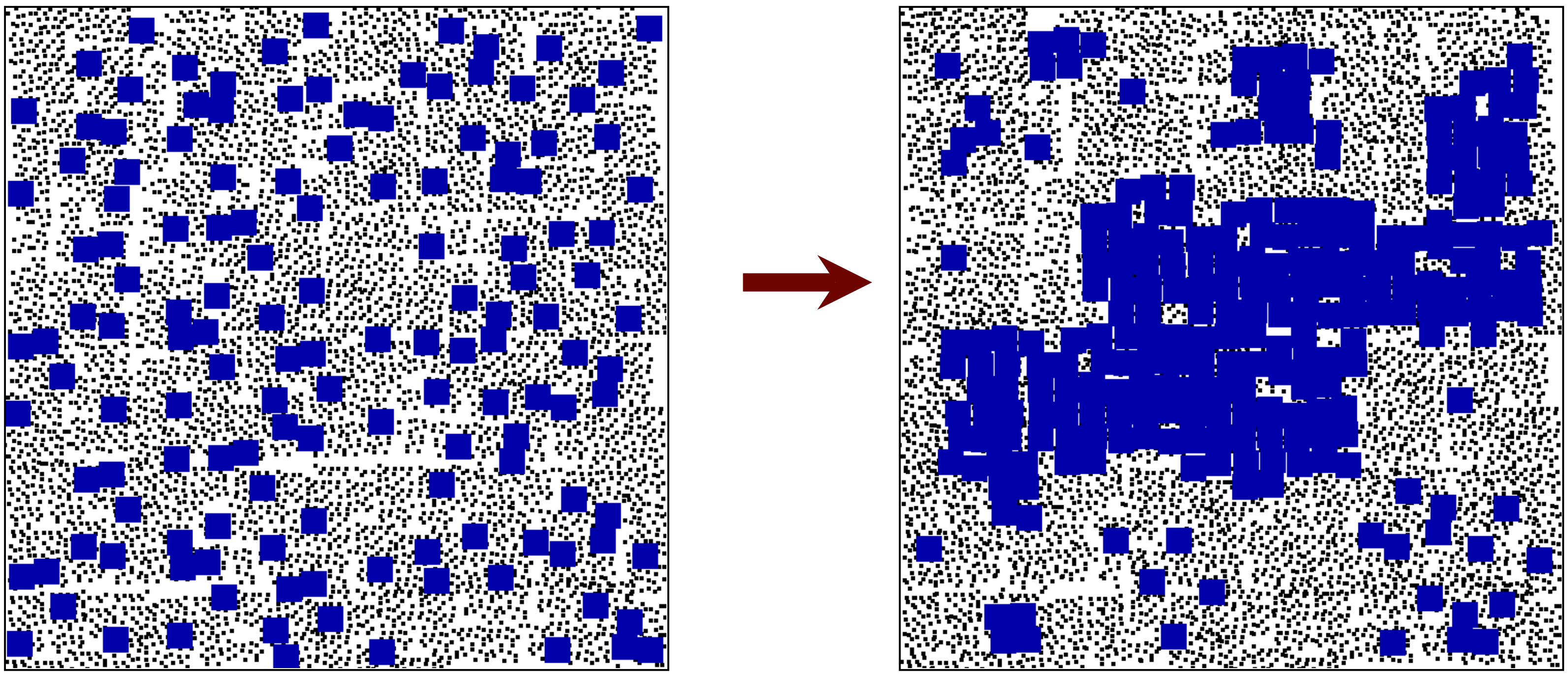


Low temperature

Randomized Algorithms & Large Data

III. Colloids: mixtures of two types of molecules.

Binary mixtures of molecules; Must not overlap.



Low density High density

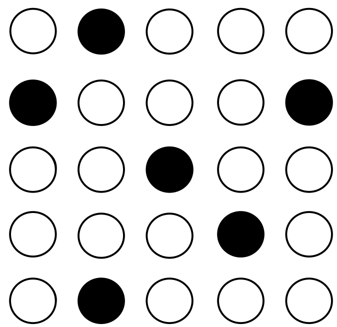
Above some density increases, large particles cluster together.

* purely entropic *

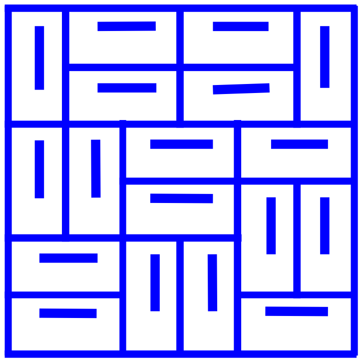
Randomized Algorithms & Large Data

IV. Sampling Algorithms: learn by sampling

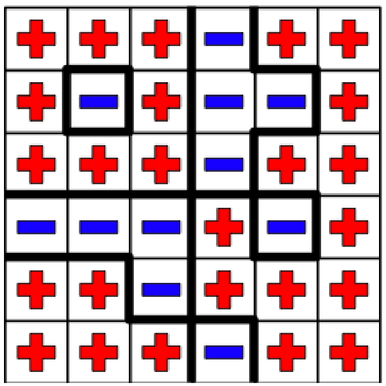
Independent Sets



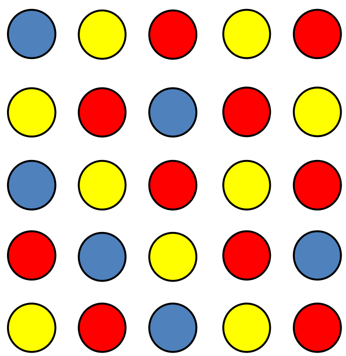
Matchings



The Ising Model



Colorings (Potts Model)



How Do We Sample?

- "Push" the squares out of the way to increase density.

Fast, but wrong distribution.

- Scramble the squares by moving one at a time to available places.

Right distribution, but slow.

Goal: *Fast* and *Correct*

Main Questions

- Is the problem efficiently computable (in polynomial time)?

  Which problems are "intractable"?

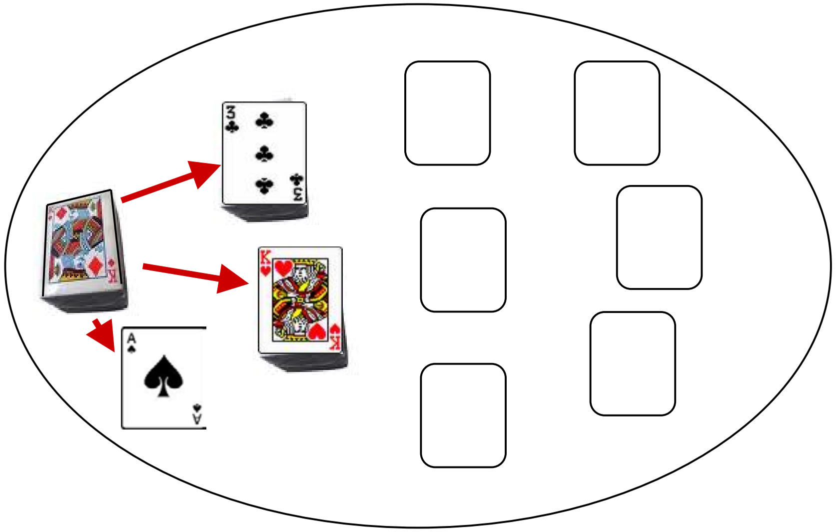
- Does the "natural" sampling method work?

Outline

- Basics of Sampling
  - Independent Sets on $\mathbf{Z}^2$
- Applications
  - Physics
  - Colloids
- Harnessing Phase Transitions
  - Self-Organizing Particle Systems

Markov chains

Perform a random walk among valid configurations



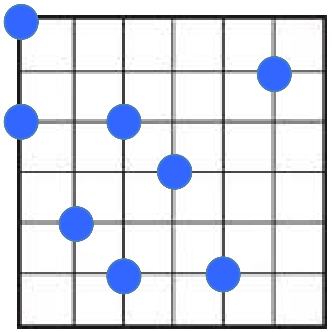
To design a useful Markov chain:

✓ ▪ Connect the state space;

✓ ▪ Define transition probabilities so that the chain will converge to π (e.g., the Metropolis Algorithm)

? ▪ Show the chain is "rapidly mixing" – i.e., distribution will be close to π in polynomial time.

Eg: Independent Sets



Given λ, let $\pi(I) = \lambda^{|I|}/Z$,

where $Z = \sum_J \lambda^{|J|}$.

MC_{IND} ("Glauber Dynamics")

Starting at I_0, Repeat:

- Pick $v \in V$ and $b \in \{0,1\}$;
- If $v \in I$, $b=0$, remove v w.p. $\min(1,\lambda^{-1})$
- If $v \notin I$, $b=1$, add v w.p. $\min(1,\lambda)$ if possible;
- O.w. do nothing.

This chain connects the state space and converges to π.

How long?

To Upper Bound the Mixing Time

- Spectral gap (1 – λ_1 > 1/poly, $\lambda_1,\ldots,\lambda_{|\Omega|-1}$ eigenvalues)
- Coupling
- Conductance (Cheeger's Inequality)
- Isoperimetric Inequalities (Dirichlet form, log Sobolev,…)

Mostly from physics

Ex: For Independent sets, Coupling gives fast mixing when $\lambda \leq ½$.

Many other improvements…

Sampling Independent Sets

Independent sets on Z^2:

$\pi(I) = \lambda^{|I|}/Z$, where $Z = \sum_J \lambda^{|J|}$.

MC$_{IND}$ is fast on Z^2 when:

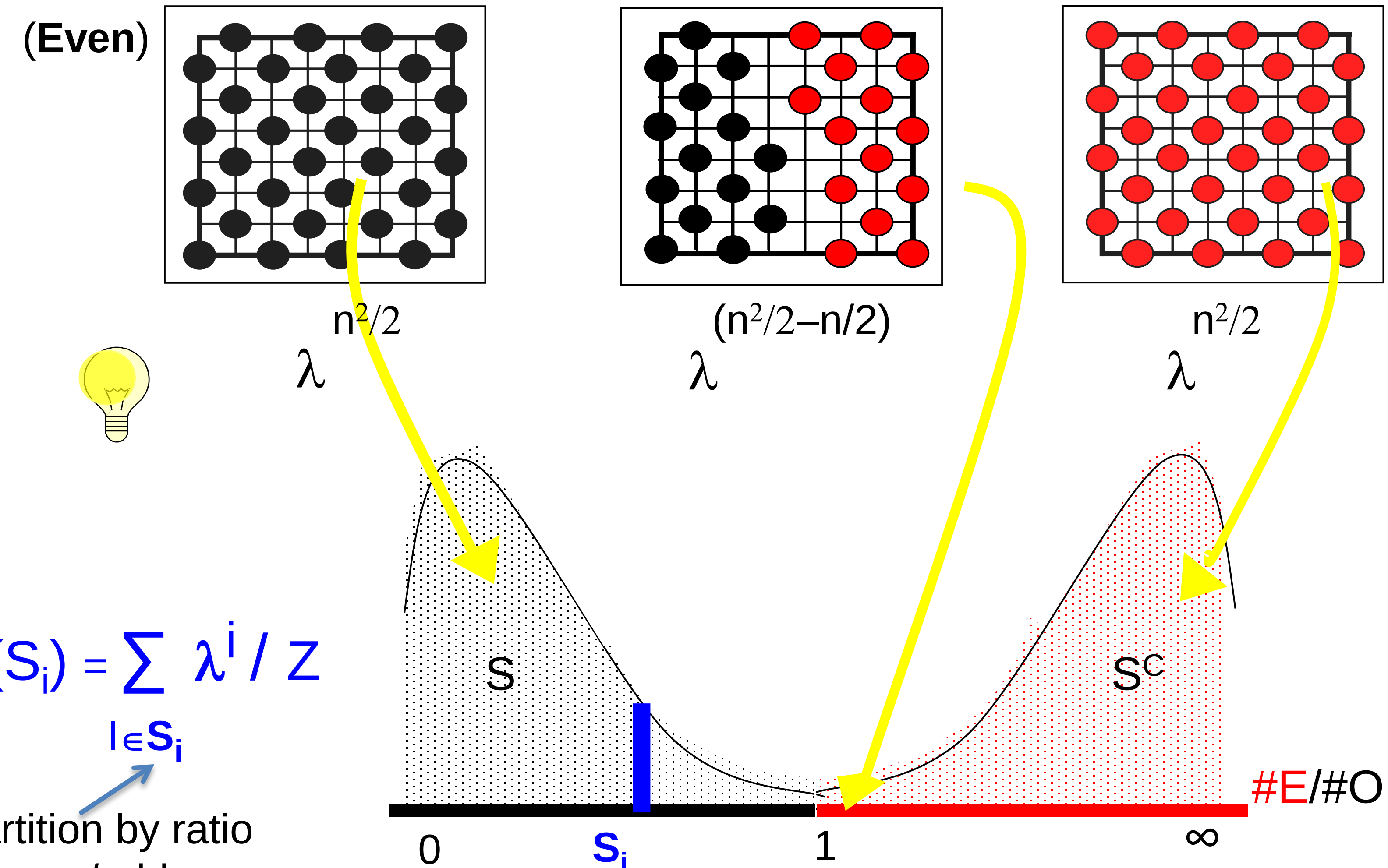
- $\lambda \leq 1$ [Luby, Vigoda]
- $\lambda \leq 1.68$ [Weitz]
- $\lambda \leq 2.48$ [VVY '13]

MC$_{IND}$ is slow on Z^2 when:

- $\lambda > 80$ [BCFKTVV]
- $\lambda > 50.6$ [R.]
- $\lambda > 5.396$ [Blanca, Galvin R., Tetali '13]

Conjecture: Fast for $\lambda < \lambda_c$ and slow for $\lambda > \lambda_c$ for λ_c H 3.79.

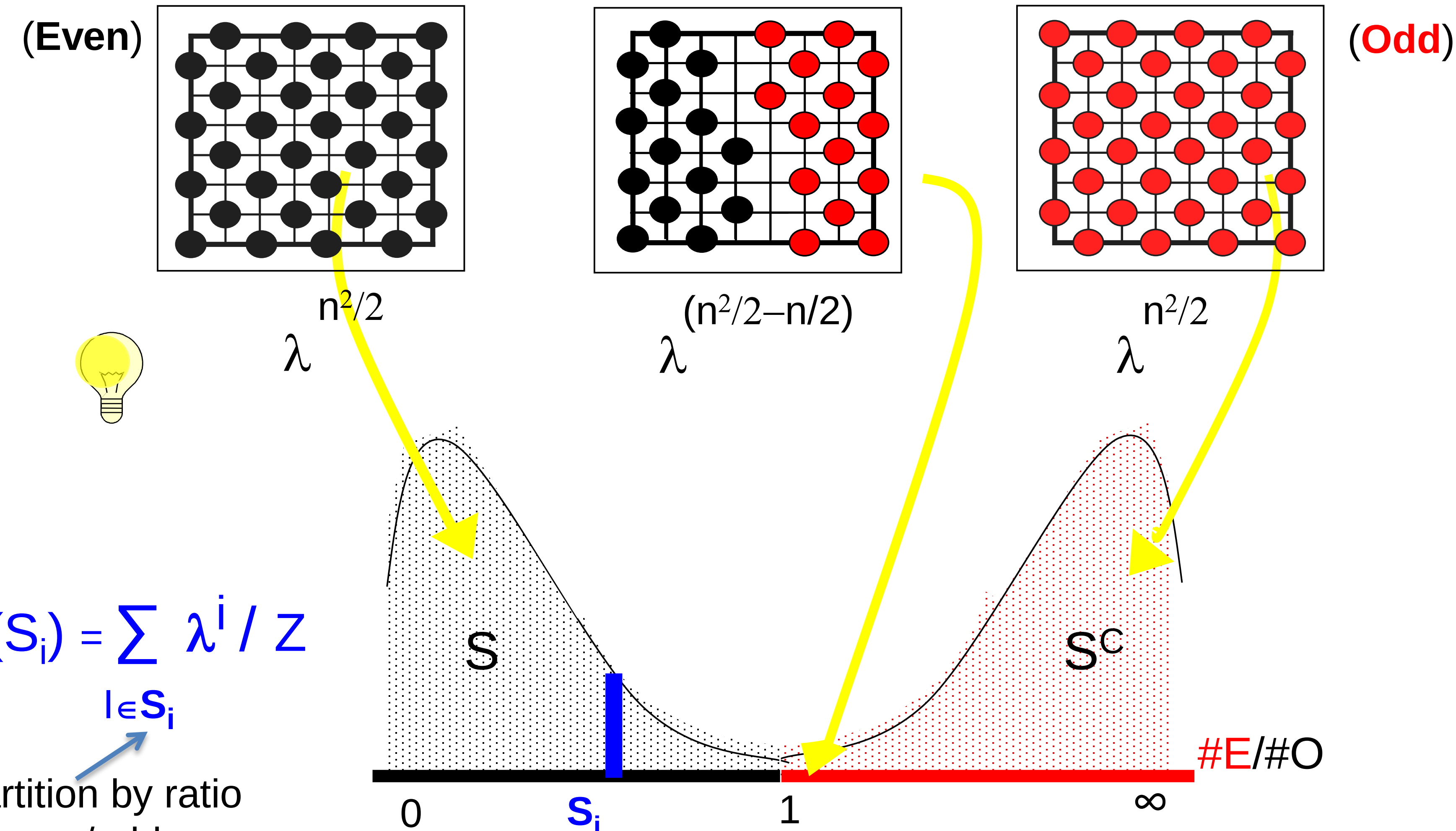
Slow mixing of MC_{IND} (large λ)



$\pi(S_i) = \sum_{I \in S_i} \lambda^i / Z$

Partition by ratio of even/odd

Slow mixing of MC_{IND} (large λ)



$$\pi(S_i) = \sum_{I \in S_i} \lambda^{|I|} / Z$$

Partition by ratio of even/odd

λ large $\rightarrow$ there is a "bad cut,"

. . . so MC_{IND} is slowly mixing. ☒

Mixing times for *local* algorithms

1. Independent sets on Z^2:

Conj: $\lambda_c = 3.79$: fast for $\lambda < \lambda_c$; slow for $\lambda > \lambda_c$.

Thms: Fast for $\lambda < 2.48$; slow for $\lambda > 5.396$

2. Ising model on Z^2:

Thms: Fast for $\lambda < \lambda_c$; slow for $\lambda > \lambda_c$.

Fast for $\lambda = \lambda_c$. [Lubetzky, Sly '10]

3. 3-colorings on Z^d:

Thms: Fast in Z^2. [Luby, R., Sinclair; Goldberg, Martin, Patterson]

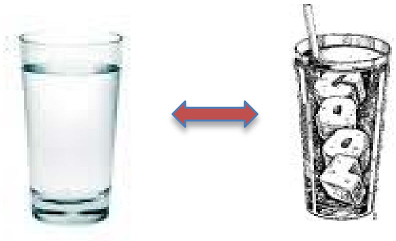
Slow in Z^d for large d. [Galvin, Kahn, R., Sorkin; Peled]

…Sometimes suggests new (fast) approaches.

Outline

- Basics of Sampling
  - Independent Sets on $\mathbf{Z}^2$
- Applications
  - Physics
  - Colloids
- Harnessing Phase Transitions
  - Self-Organizing Particle Systems

Physics

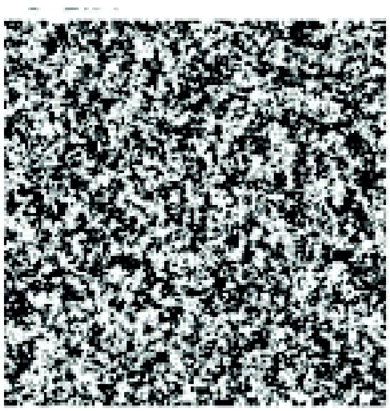


Phase transitions:

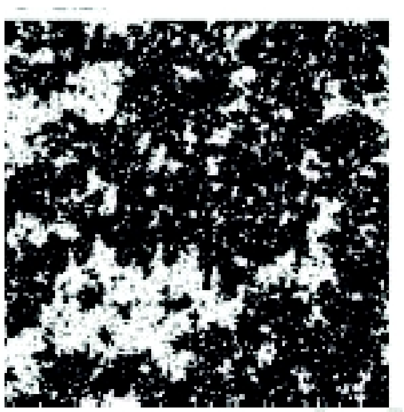
Macroscopic changes to the system due to a microscopic change to some parameter.

e.g.: gas/liquid/solid, spontaneous magnetization

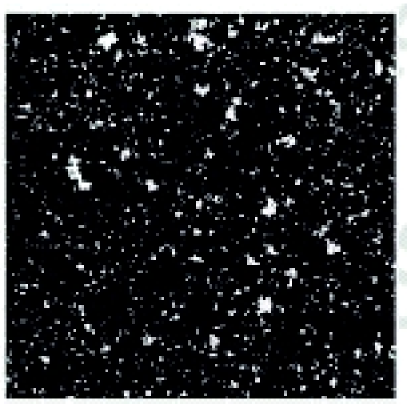
Simulations of the Ising model



High temperature



Criticality



Low temperature

Physics

Given: A physical system $\Omega = \{\sigma\}$

Define: A Gibbs measure as follows:

$H(\sigma)$ (the Hamiltonian),

$\beta = 1/kT$ (inverse temperature, and k is Boltzmann's constant)

$\pi(\sigma) = e^{-\beta H(\sigma)} / Z,$

where $Z = \sum_\tau e^{-\beta H(\tau)}$ (the partition function)

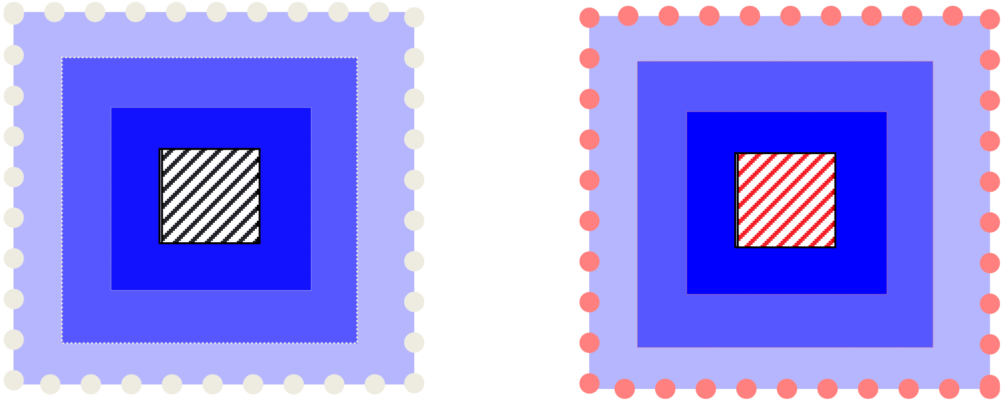
- Independent sets: $H(\sigma) = -|I|$

  If $\lambda = e^{\beta}$ then $\pi(\sigma) = \lambda^{|I|} / Z.$

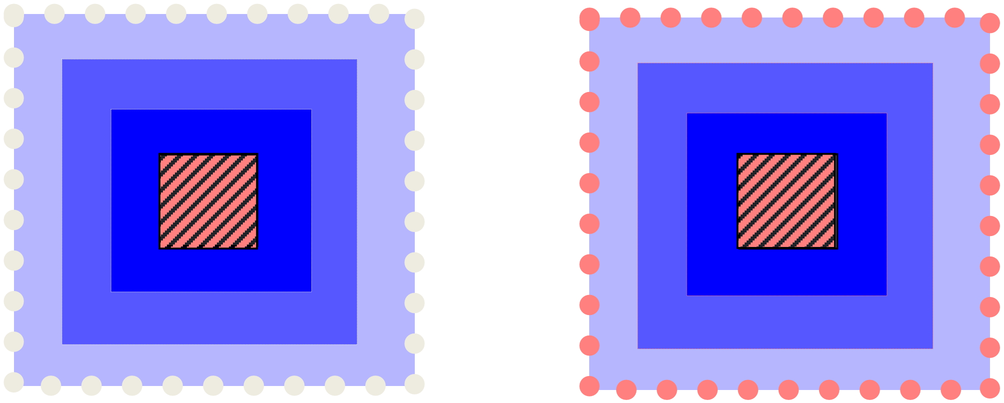
- Ising model: $H(\sigma) = - \sum_{(u,v) \in E} \sigma_u \sigma_v$

  If $\nu = e^{2\beta}$ then $\pi(\sigma) = \nu^{|E_=|} / Z.$

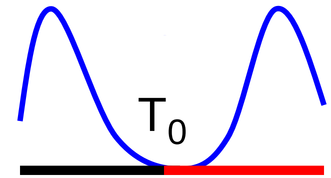
Physics perspective (cont.)

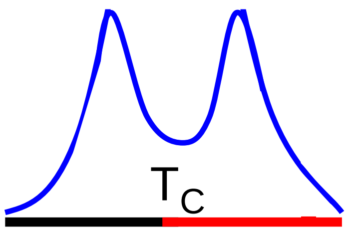


Low temperature: long range effects

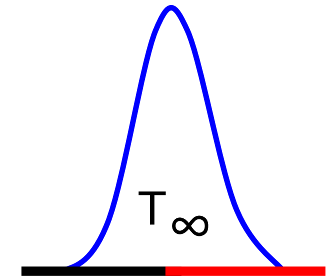


High temperature: boundary effects die out





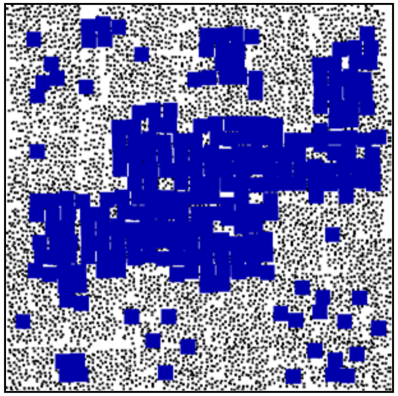
(T_C indicates a "phase transition.")



For the "hard core model" the best rigorous results are

$2.48 < \lambda_c < 5.396.$

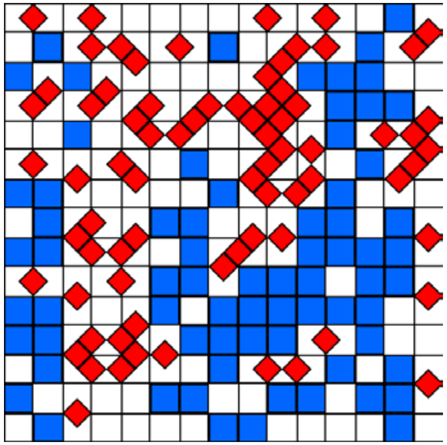
Other Models: Colloids



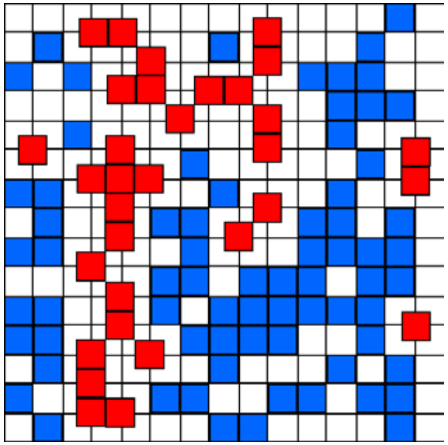
Clustering for a class of interfering binary mixtures
[Miracle, R., Streib]

Thm: Low density: models won't cluster.
High density: models will cluster.

Including:



Ising Model

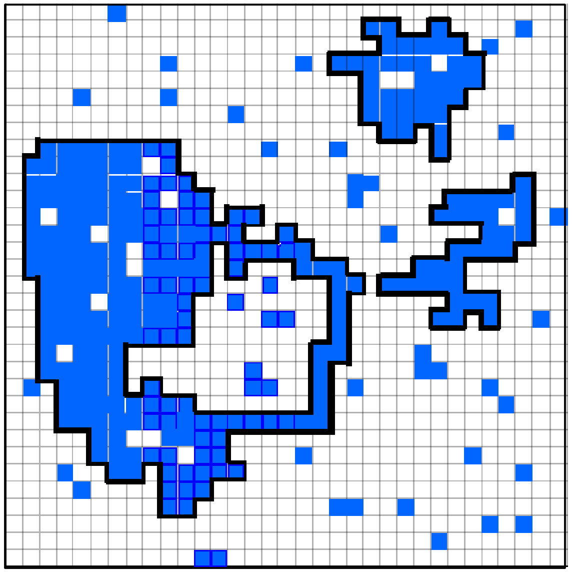


Independent Sets

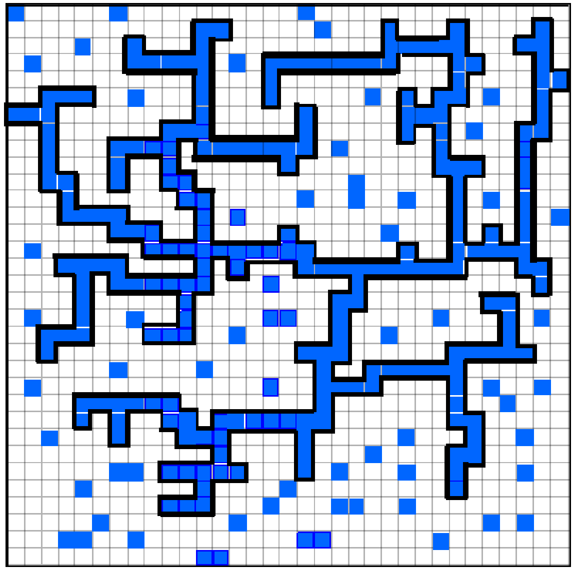
The "Clustering Property"

What does it mean for a configuration to "cluster"?

- There is a region with large area and small perimeter
- that is dense with one kind of tile
- and the complement of the region is sparse



Clustering

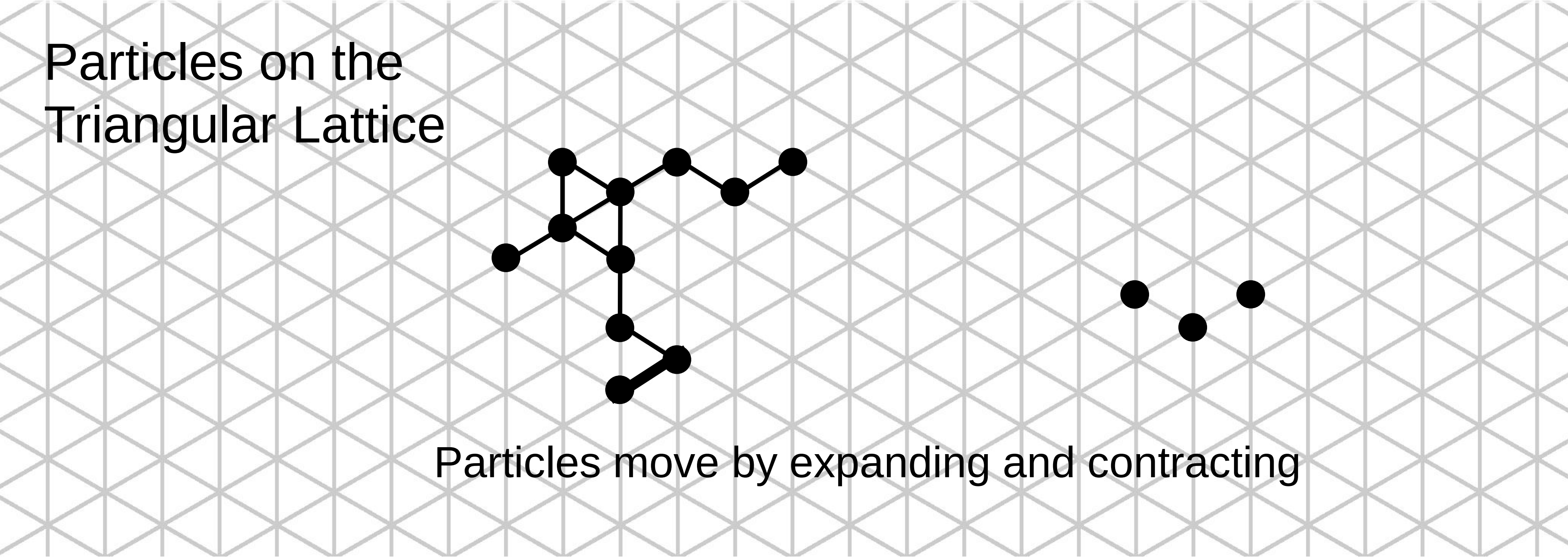


No Clustering

Outline

- Basics of Sampling
  - Independent Sets on $\mathbf{Z}^2$
- Applications
  - Physics
  - Colloids
- Harnessing Phase Transitions
  - Self-Organizing Particle Systems

Geometric Amoebot Model



Assumptions/Requirements:

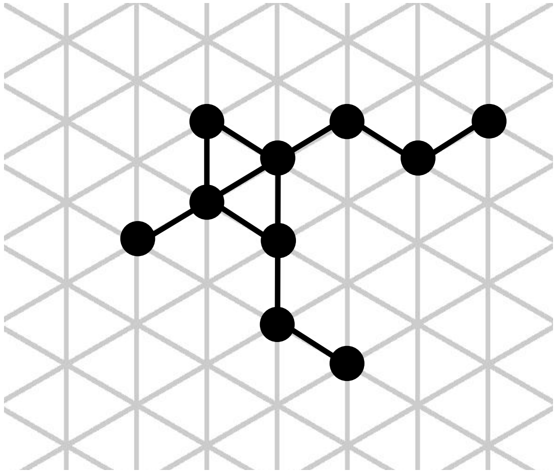
- Particles have constant size memory
- Particles can communicate only with adjacent particles
- Particles have no common orientation, only common chirality
- Require that particles stay connected

Previous Work in Amoebot Model

Algorithms exist for:

- Leader election
- Shape formation (triangle, hexagon)
- Infinite object coating

(Rida A. Bazzi, Zahra Derakhshandeh, Shlomi Dolev, Robert Gmyr, Andréa W. Richa, Christian Scheideler, Thim Strothmann, Shimrit Tzur-David)



We were interested in the **compression** problem: to gather the particles together as tightly as possible.

- Often found in natural systems
- Our approach is decentralized, self-stabilizing, and oblivious (no leader necessary)
- Use a Markov chain that rewards internal edges for the local algorithm

MC converges to: $\Pi(\sigma) = \lambda^{e(\sigma)} / Z$, where $e(\sigma)$ is # internal edges.

Results for Compression

[Cannon, Daymude, R., Richa '16]

Thm: (compression) For any $\lambda > 2 + 2^{1/2}$, there is a constant $\alpha > 1$ s.t. particles will be α-compressed almost surely.

Thm: (non-compression) For any $\lambda < 2.17$, for any $\alpha > 1$, the probability that particles are α-compessed is exp. small.

Defn: A particle configuration is α-compressed if its perimeter is at most α times the minimum perimeter (i.e., $\Theta(n^{1/2})$).

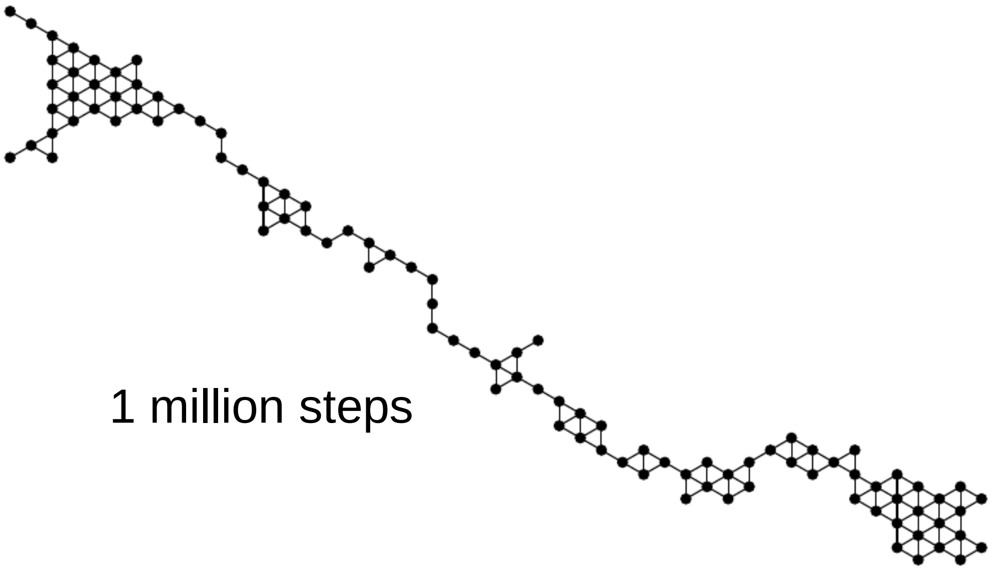
Pf. idea: * "Peierls arguments" based on information theory

* Use of combinatorial identities:

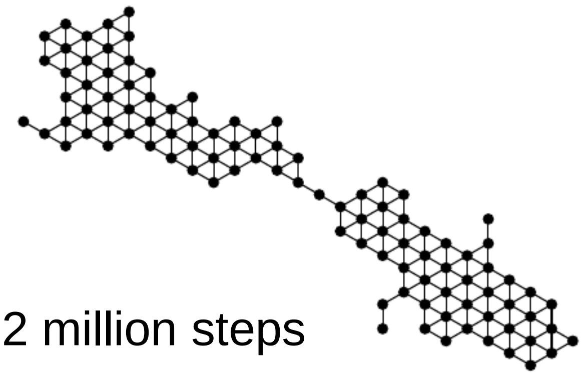
e.g., $\lambda = 2 + 2^{1/2}$ is the "connective constant" for self-avoiding walks on the hexagonal lattice

Simulations: $\lambda = 4$

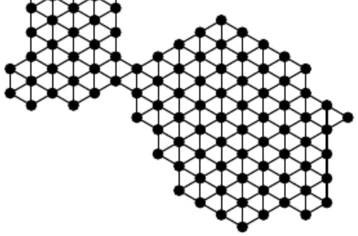
Start: 100 particles in a line



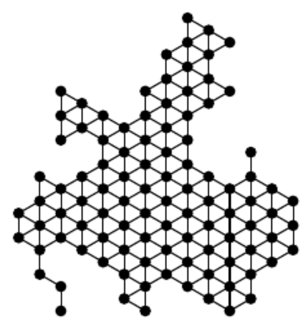
1 million steps



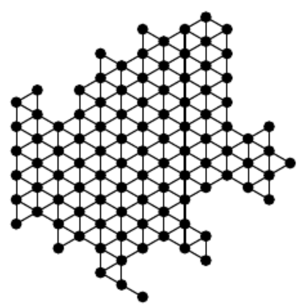
2 million steps



3 million steps



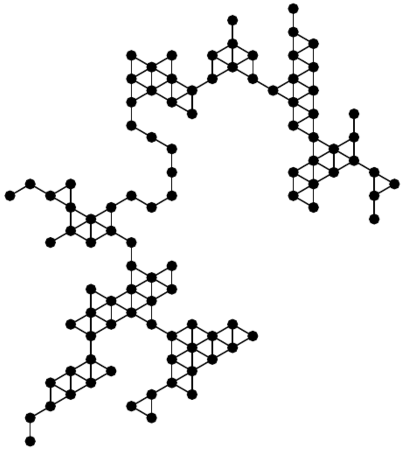
4 million steps



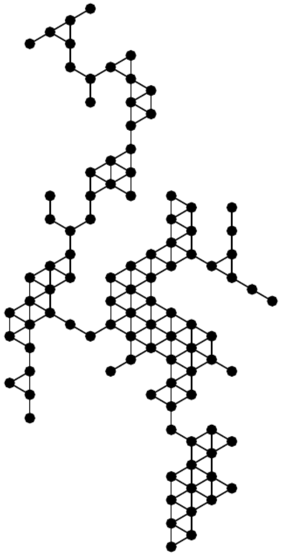
5 million steps

Simulations: $\lambda = 2$

Start: 100 particles in a line



10 million steps

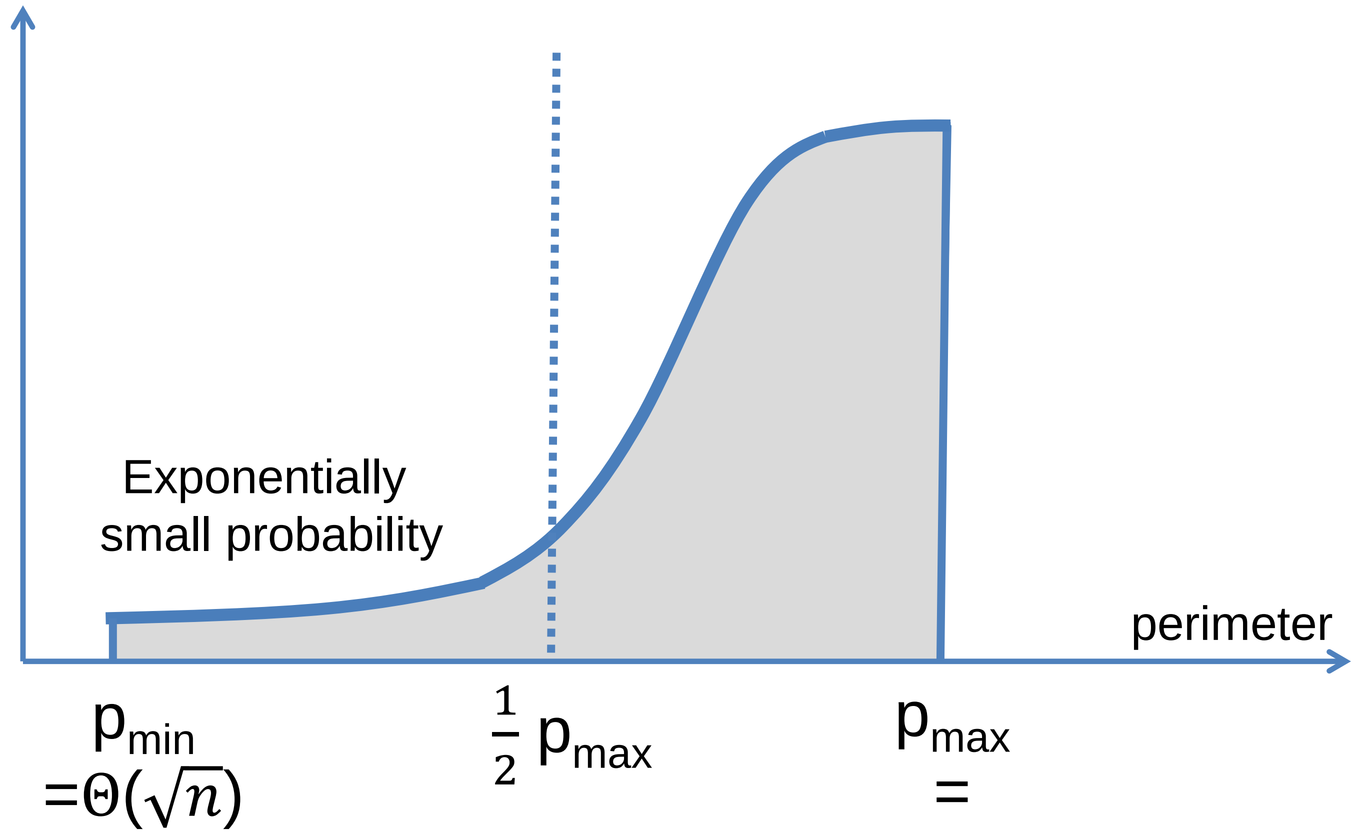


20 million steps

Why not compression for all $\lambda > 1$?

$\Pi(\sigma) = \lambda^{e(\sigma)} / Z$

$\lambda = 1$

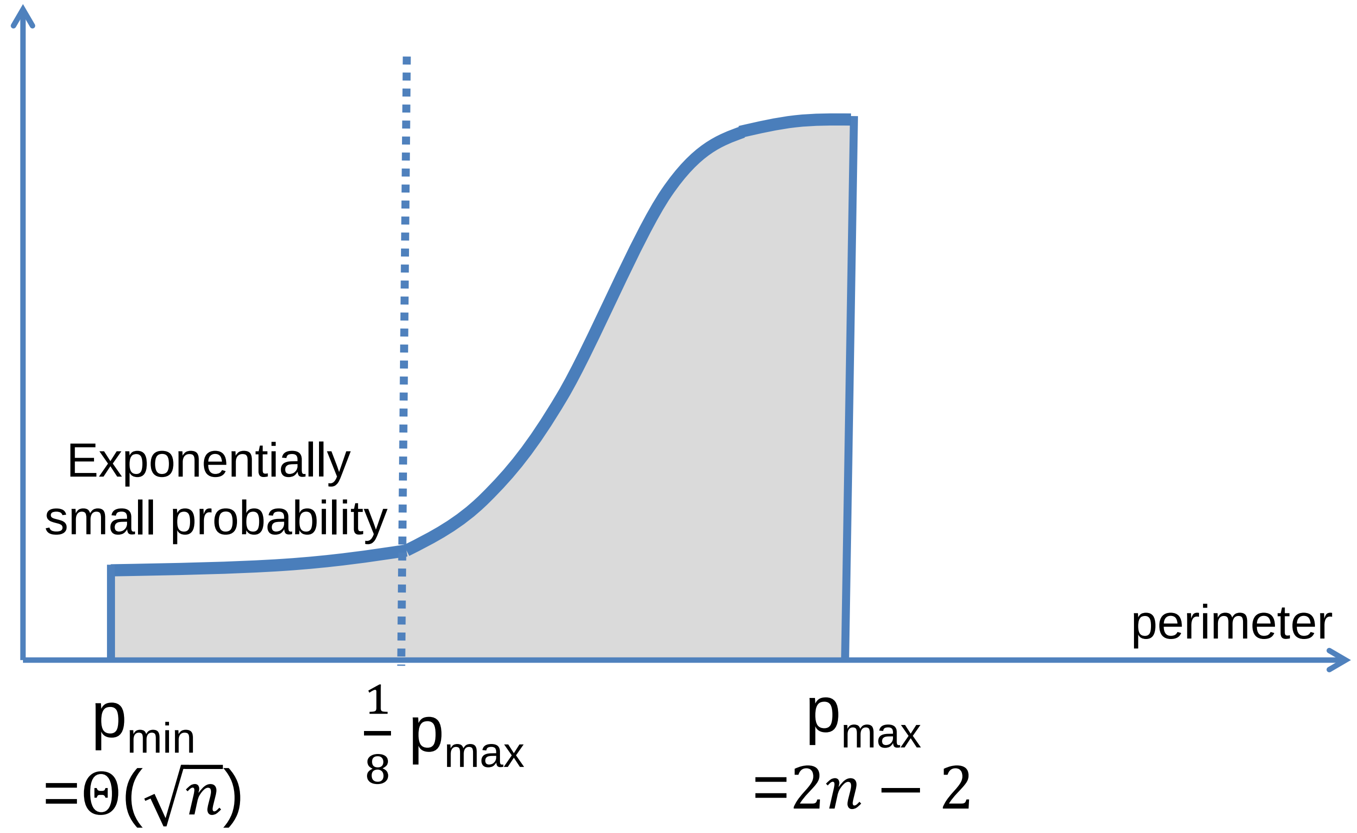


A graph of perimeter vs. stationary probability when all configurations have equal weight.

Why not compression for all $\lambda > 1$?

$\Pi(\sigma) = \lambda^{e(\sigma)} / Z$

$\lambda = 2$

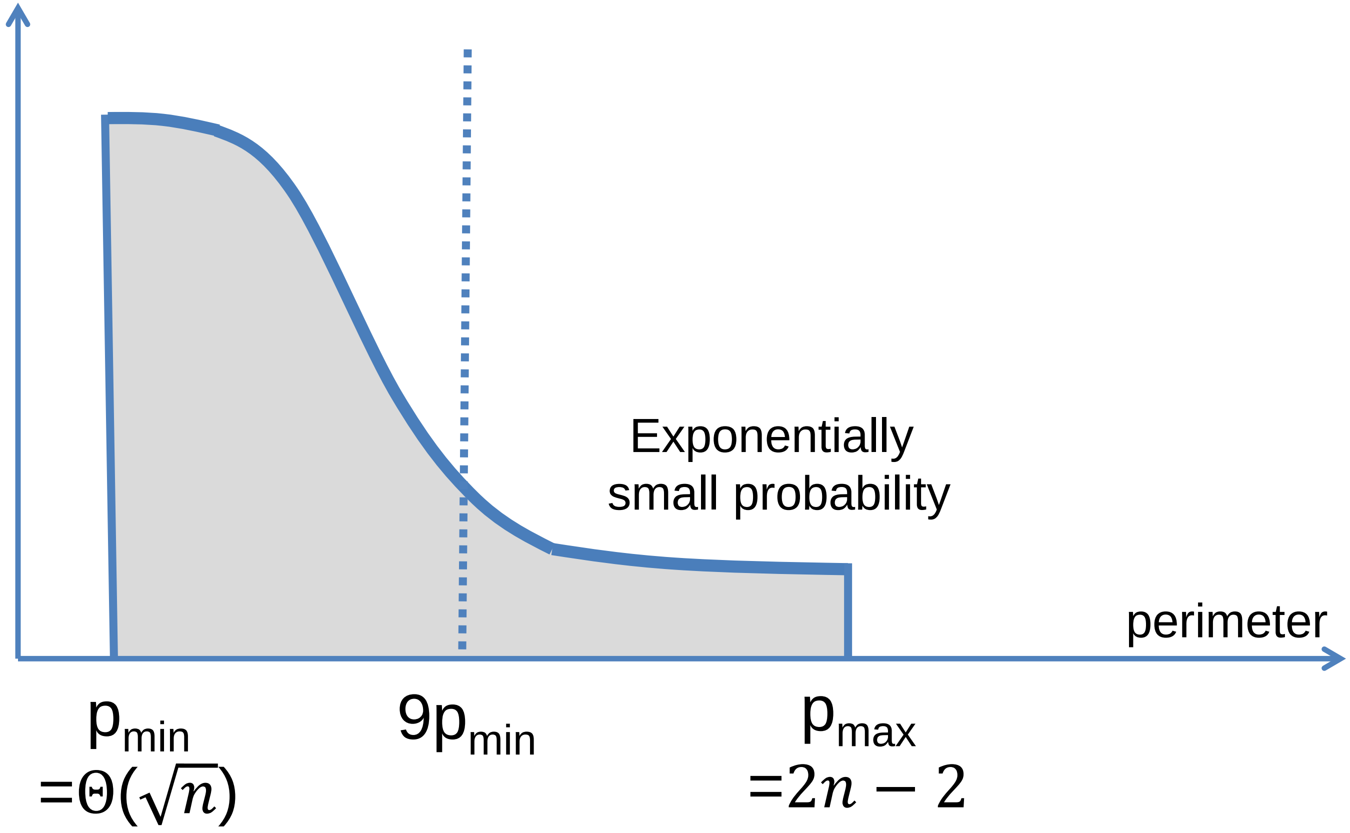


Perimeter vs. stationary probability when configurations with more internal edges have higher probability

Why not compression for all $\lambda > 1$?

$\Pi(\sigma) = \lambda^{e(\sigma)} / Z$

$\lambda = 4$



Perimeter vs. stationary probability when configurations with more internal edges have higher probability

Challenges and Opportunities

- **Algorithms:** How an we use knowledge of phase trans to design efficient algorithms at all temps?

- **Applications:** Can we develop better methods to confirm phase changes without rigorous proofs?

- **Data:** Does stat. phys. play an increasing role as "n" becomes huge and algs are forced to be simpler?

Thank you!