Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program
605-447/625-438
Dr. Mark Fleischer

Module 6.1: Other Optimization Techniques—Theoretical Foundations of the Simulated Annealing Algorithm

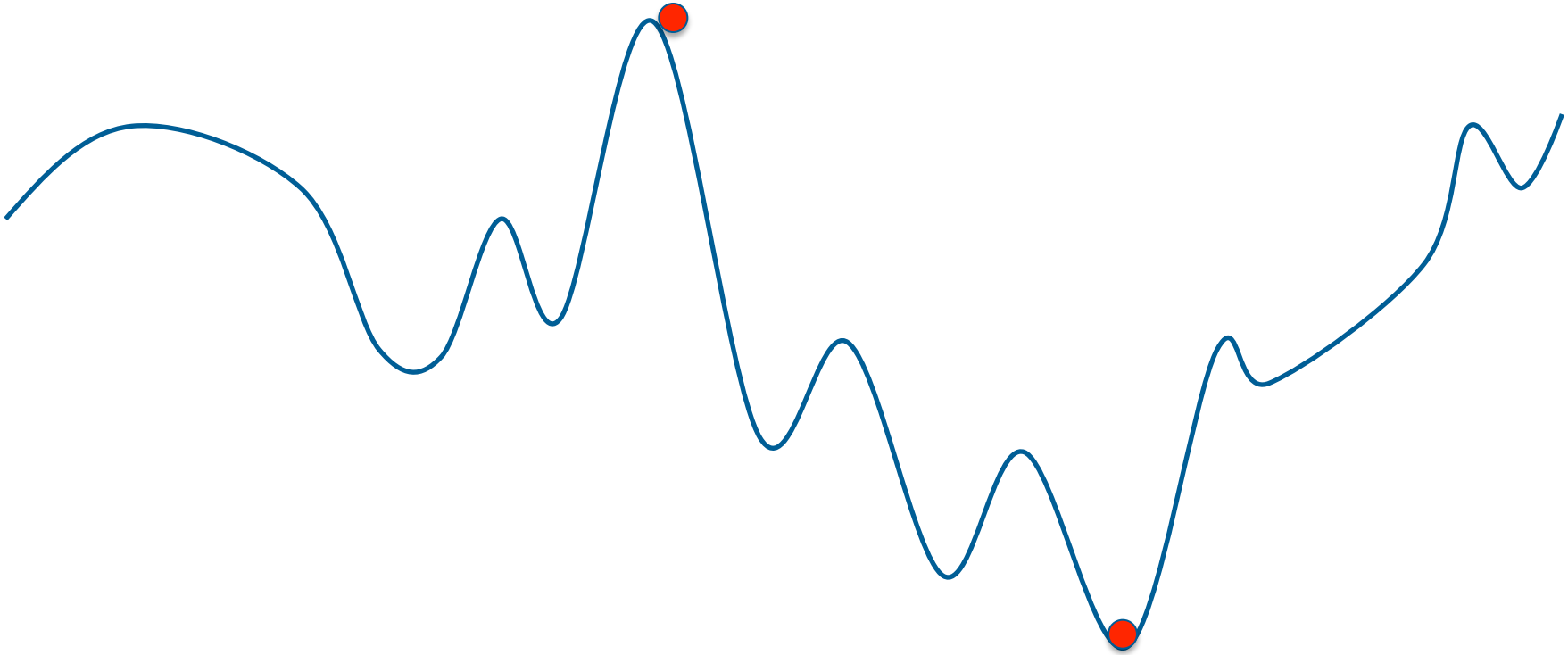
This Sub-Module Covers …

- Describe the history and mathematical foundations of the Simulated Annealing Program.
- Provides a foundation for study of stochastic neural networks which we will cover in later modules.
- Subsequent sub-modules will describe implementation issues and other optimization techniques.

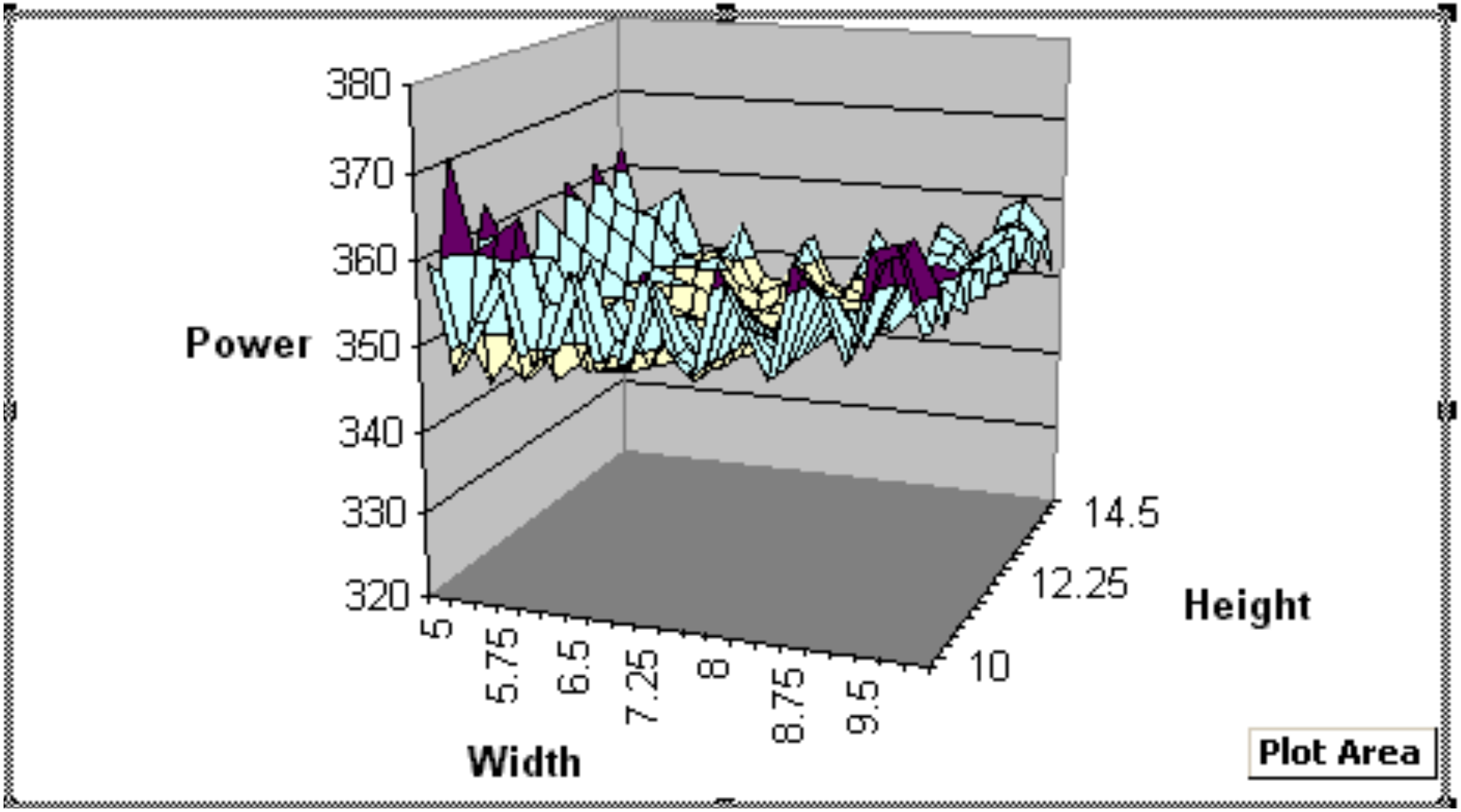
Overview

- Why Global Optimization?
- Metaheuristics
  - Simulated Annealing
  - Genetic Algorithms
  - Others as well
- Research Issues

A Nasty Landscape



A Really Nasty Landscape



What is Visible to an Algorithm?

- Can't see the entire landscape.
  - Doing so equivalent to total enumeration.
- Only a neighborhood is visible. Why?
  - A tractable computational effort required.

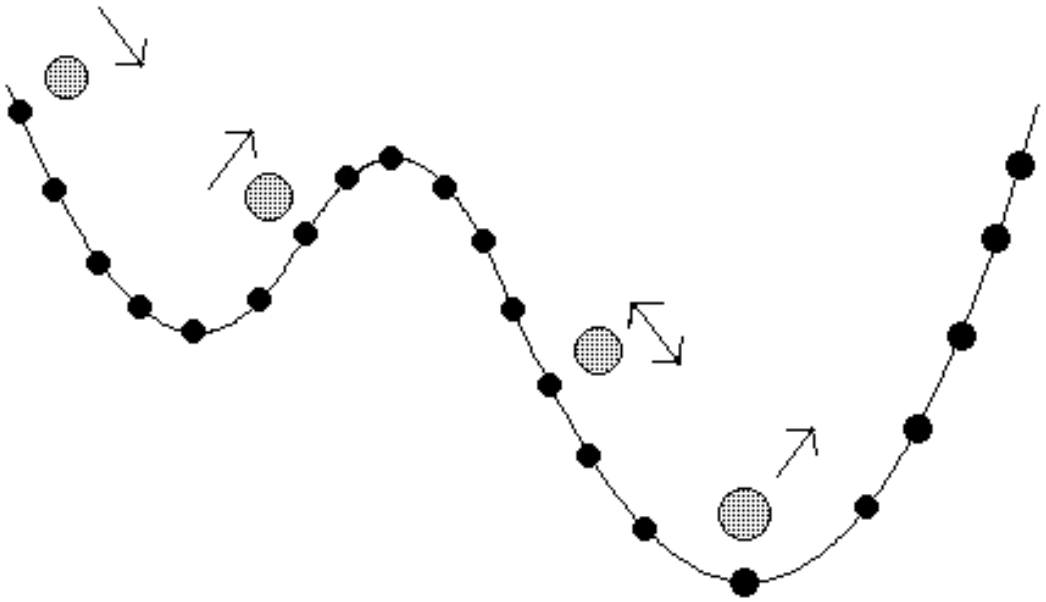
Search Themes

- Directed *e.g.* gradient based.
  - Not applicable to most real-world problems. Why?
- Deterministic.
- Random.
- Random with some determinism.
- Memory dependent – avoid where we have already been.

Search Paradigms
Rules for Path-Finding

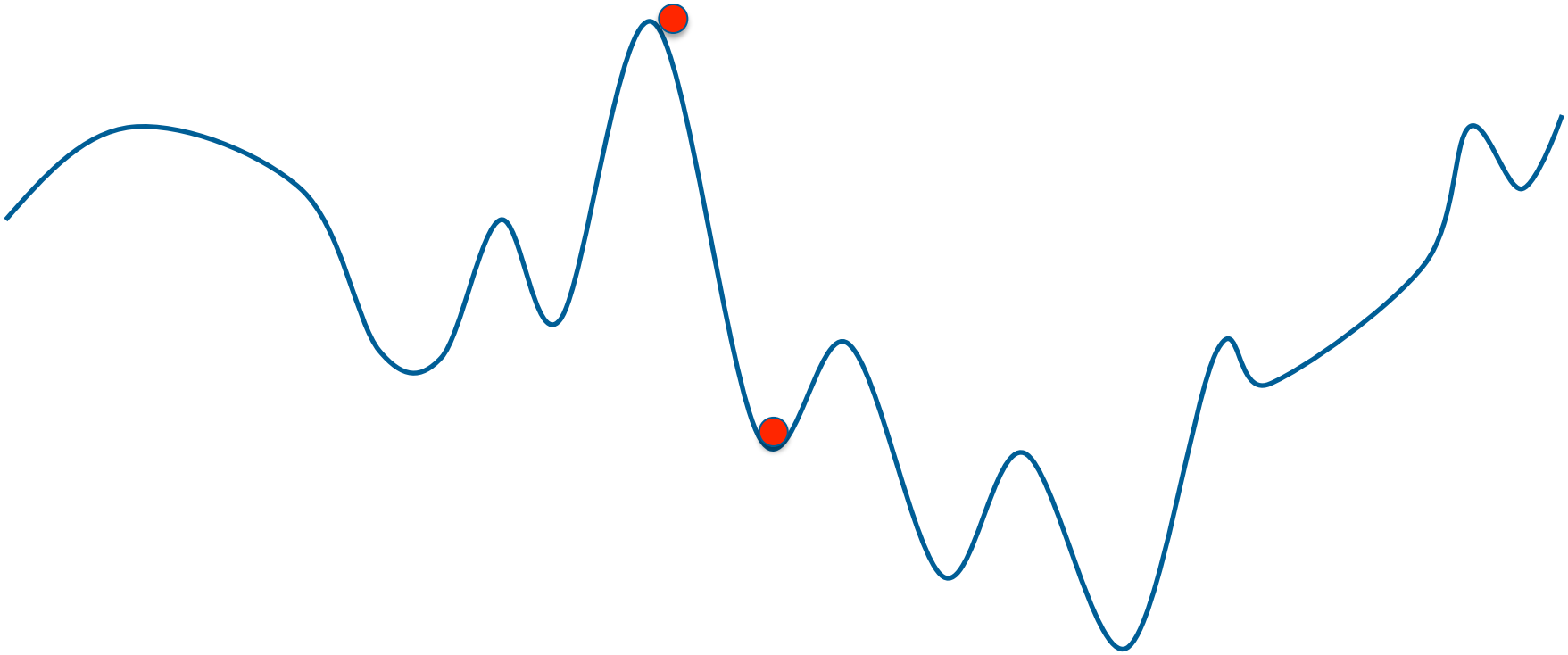
- Thermodynamics
  - Simulated Annealing is based on the laws of thermodynamics
- Biological
  - Genetic algorithms are based on the paradigms of evolution: natural selection, survival of the fittest, etc.
- Common-Sense/Intelligent

Simulated Annealing



Hill Climbing ---> Getting to the Global Optima

A Nasty Landscape



Modeling Thermodynamic Systems

- a large (really large) population (ensemble) of physical entities, *e.g.,* molecules of a gas, liquid, or solid.
- obeys the laws of thermodynamics and statistical mechanics.
- entities subject to collisions.
- entities have mass, position, velocity (kinetic energy)

Statistical Mechanics

Diffusion, entropy, temperature

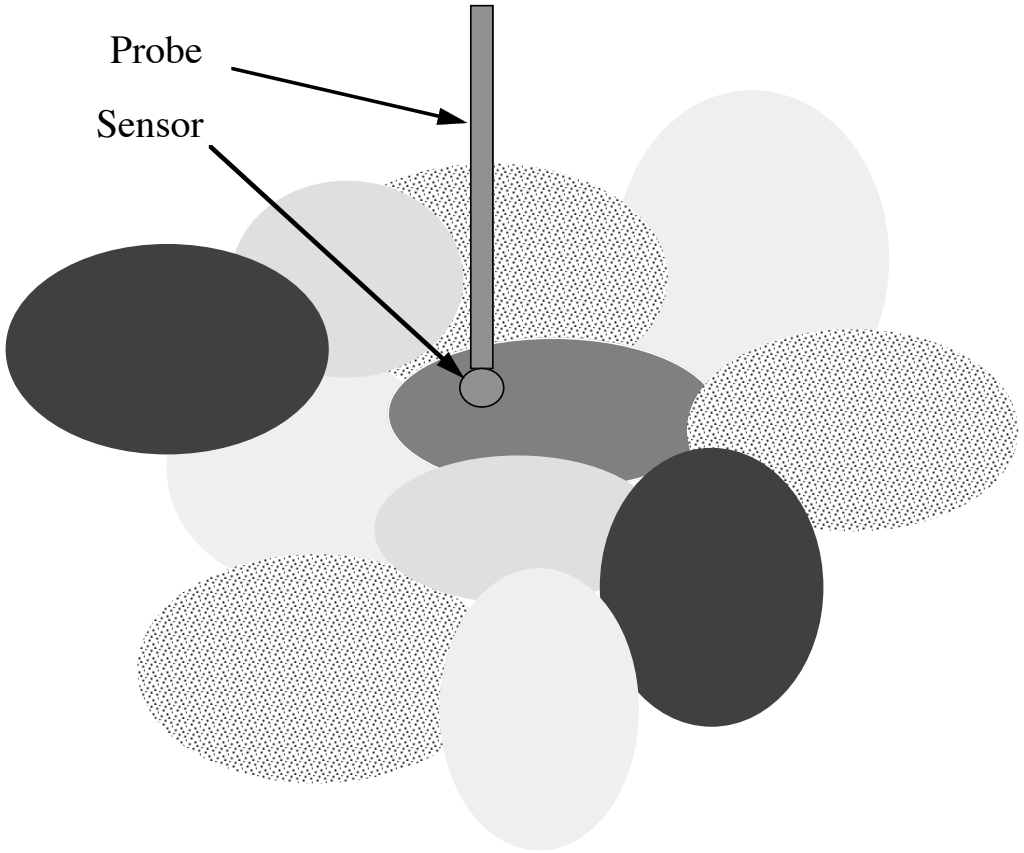
Thermodynamics/Statistical Mechanics

- Particles change energy levels sporadically because of collisions.
- Energy level of system proportional to the **mean distance** between particles.
- Want to facilitate a simulation of such a system!

Thermodynamics/Statistical Mechanics

- Kinetic energy of particles is virtually a **continuous function** because of the number of particles involved.
- Define a **finite** set of states *i* each with an associated energy level E_i .
- Consider the evolution of "states" at a given temperature.
  - Some regions become more/less dense randomly over time.
  - Some regions gain/lose energy over time.

Gas with Varying Densities



Temperature at sensor fluctuates.
Energy states change over time.

Definition of Terms

t = an estimate of the mean kinetic energy of a substance, gas, etc.

$\pi_i(t)$ the stationary probability of energy state *i* at temperature *t.*

E_i the energy level of partition *i*.

Modeling the System of Particles

- An ensemble of particles seeks the maximum state of disorder.
- Statistical Mechanics translates this to a probability distribution that maximizes entropy.

What is Entropy?

A measure of the total ‘uncertainty’ associated with an ensemble of possibilities.

A single, non-negative scalar value that is associated with uncertainty.

We want it to encompass many different possible events.

We want it to be additive.

$$S_i \propto \frac{1}{p_i}, \quad S_{ij} \propto S_i + S_j \qquad S_i = k \log \frac{1}{p_i} = -k \log p_i$$

$$S_{ij} = k \log \left(\frac{1}{p_i p_j} \right) = k \log \frac{1}{p_i} + k \log \frac{1}{p_j}$$

What is 'Uncertainty'

A	B
0.01	0.99

Uncertainty = Expected Surprisal

$$U = p_A S_A + p_B S_B$$

$$U = -k \sum_i p_i \log p_i$$

A	B
0.5	0.5

An NLP

$$\text{Max} \quad -k \sum_{i=1}^{n} \pi_i(t) \log \pi_i(t)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(t) E_i = k\,t$$

$$\sum_{i=1}^{n} \pi_i(t) = 1$$

$$t, \pi_i(t), E_i \geq 0 \ \forall i$$

The Boltzmann Distribution

$$\pi_i(t) = \frac{e^{-E_i/kt}}{B(t)}$$

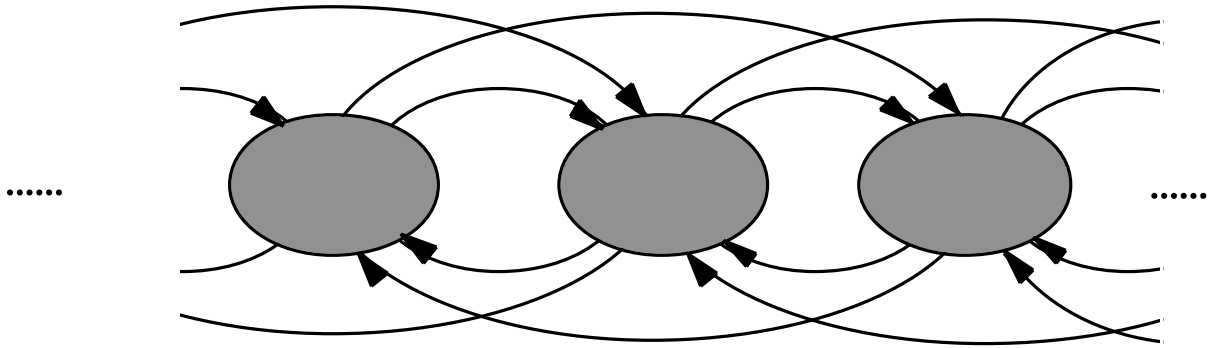
$$B(t) \equiv \sum_{i=1}^{n} e^{-E_i/kt}$$

State Transitions

(Metropolis, *et al.* 1953)

Detailed balance requires that

$$\pi_i(t)p_{ij}(t) = \pi_j(t)p_{ji}(t)$$

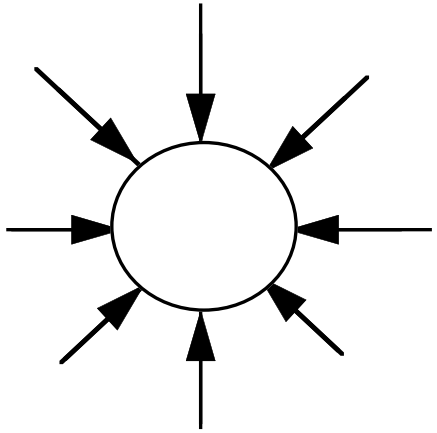


State Transitions

(Metropolis, *et al.* 1953)

Global balance requires that

$$\pi_i(t) = \sum_j \pi_j(t) p_{ji}(t)$$



Equations of State

(Metropolis Acceptance Criterion)

- Regions (states) move spontaneously to different energy levels.
- Given current energy level *i*,
  - if $\Delta E = E_j - E_i \leq 0$ then that region moves to the new energy level *j*.
  - if $\Delta E = E_j - E_i > 0$ then that region moves to the new energy level *j* with the following acceptance probability

$$\Pr\{\text{Accept state } j\} = e^{-\Delta E / t}$$

Transition Probability

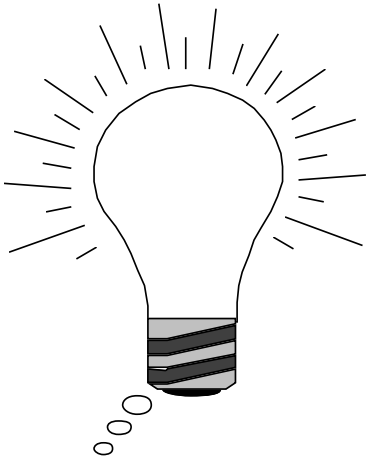
Balance conditions lead to the form for the transition probabilities:

$$p_{ij}(t) = \begin{cases} G_{ji} e^{-\Delta E^{+} / t} & \Delta E_{ji} > 0, j \neq i \\ G_{ji} & \Delta E_{ji} \leq 0, j \neq i \\ 1 - \sum_{k} p_{ik}(t) & j = i \\ 0 & \text{otherwise} \end{cases}$$

Kirkpatrick, et al. (1983), Cerny (1985)

The annealing process lowers
the energy content of a solid.

Energy content of a solid ⇔ Objective Function of an optimization problem



The Simulated Annealing Algorithm

Mathematical Properties of SA

Convergence in Probability (stationary probability)

$$\lim_{k \to \infty} \Pr\{S_k \in S_{\text{OPT}}\} = 1$$

Convergence in Distribution

$$\lim_{t \to 0} \pi_i(t) = \begin{cases} \dfrac{1}{|S_{\text{OPT}}|} & \text{if } i \text{ is in the set of optima} \\ 0 & \text{otherwise} \end{cases}$$

Summarizing Simulated Annealing

- The Simulated Annealing is based on modeling a thermodynamic system.
  - As in thermodynamics, we model large-scale behavior instead of modeling the dynamics of each involved entity.
  - Metropolis formulated 'Equations of State' as a non-linear mathematical program.
  - Led to the mathematical form of transition probabilities between states.
- Provides equivalent state probabilities as in thermodynamic systems.
- Kirkpatrick and Cerny found a way to apply these equations of state to simulate the annealing process.

Conclusion

- More metaphors exist
  - Tabu Search, Adaptive Memory Programming and Genetic Algorithms
- Active area of research
- Many overlooked/ignored possibilities
- Balance accuracy, search time
- Structure, properties, computer architecture.