

## B. Stochastic Neural Networks

(in particular, the stochastic Hopfield network)

10/5/09 1

### Trapping in Local Minimum

The diagram shows a black curve representing an energy landscape with two valleys. A blue dot representing a particle is positioned on the slope of the left valley. Two green arrows point downwards from the dot, indicating the direction of the local gradient, which leads to a local minimum rather than the global one on the right.

10/5/09 2

### Escape from Local Minimum

The diagram shows the same energy landscape as slide 2. A blue dot is on the slope of the left valley. A red arrow points upwards and to the right from the dot, indicating a move against the local gradient to escape the valley.

10/5/09 3

### Escape from Local Minimum

The diagram shows the same energy landscape. A blue dot is on the slope of the left valley. A red arrow points up and right, followed by a series of green arrows that zig-zag down the slope towards the right valley, representing a stochastic path that eventually reaches the global minimum.

10/5/09 4

## Motivation

- **Idea:** with low probability, go against the local field
  - move up the energy surface
  - make the “wrong” microdecision
- **Potential value for optimization:** escape from local optima
- **Potential value for associative memory:** escape from spurious states
  - because they have higher energy than imprinted states

10/5/09 5

## The Stochastic Neuron

Deterministic neuron:  $s'_i = \text{sgn}(h_i)$

$\Pr\{s'_i = +1\} = \Theta(h_i)$   
 $\Pr\{s'_i = -1\} = 1 - \Theta(h_i)$

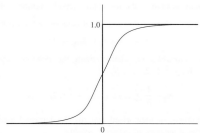
Stochastic neuron:  
 $\Pr\{s'_i = +1\} = \sigma(h_i)$   
 $\Pr\{s'_i = -1\} = 1 - \sigma(h_i)$

The graph shows the logistic sigmoid function  $\sigma(h) = \frac{1}{1 + \exp(-2h/T)}$ . The horizontal axis is labeled  $h$  and the vertical axis is labeled  $\sigma(h)$ . The curve is an S-shape passing through (0, 0.5). A red vertical line is drawn at  $h=0$ , and a green horizontal line is drawn from the intersection of the red line and the curve to the y-axis.

Logistic sigmoid:  $\sigma(h) = \frac{1}{1 + \exp(-2h/T)}$

10/5/09 6

### Properties of Logistic Sigmoid

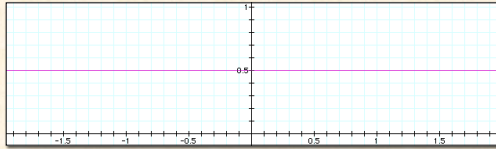


$$\sigma(h) = \frac{1}{1 + e^{-2h/T}}$$

- As  $h \rightarrow +\infty$ ,  $\sigma(h) \rightarrow 1$
- As  $h \rightarrow -\infty$ ,  $\sigma(h) \rightarrow 0$
- $\sigma(0) = 1/2$

10/5/09 7

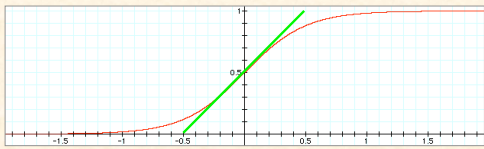
### Logistic Sigmoid With Varying $T$



$T$  varying from 0.05 to  $\infty$  ( $1/T = \beta = 0, 1, 2, \dots, 20$ )

10/5/09 8

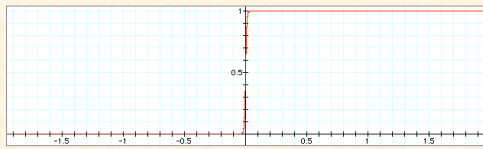
### Logistic Sigmoid $T = 0.5$



Slope at origin =  $1 / 2T$

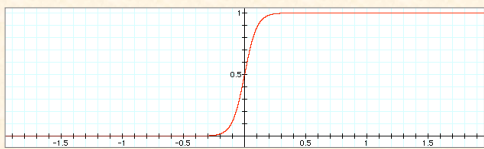
10/5/09 9

### Logistic Sigmoid $T = 0.01$



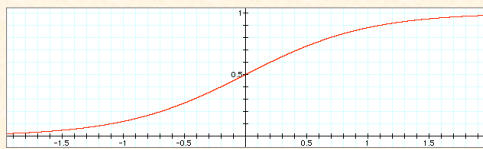
10/5/09 10

### Logistic Sigmoid $T = 0.1$

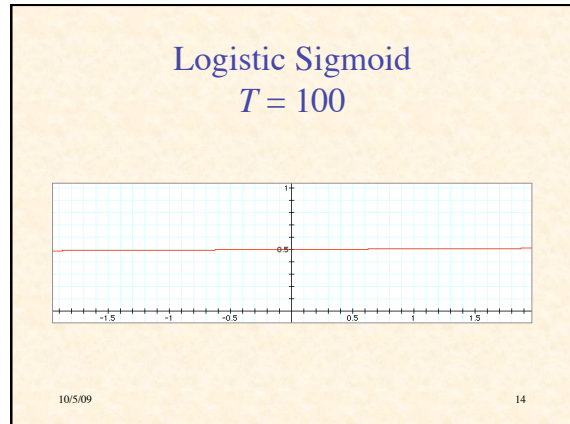
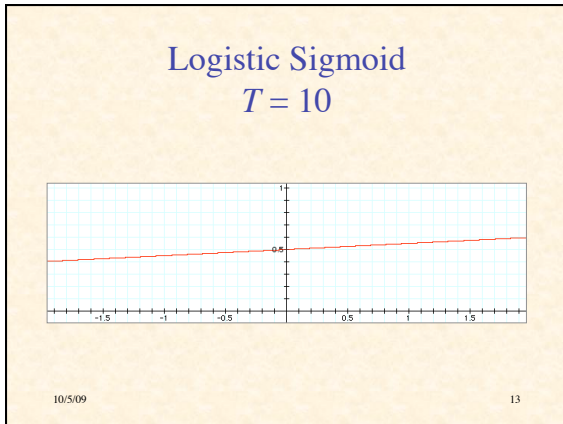


10/5/09 11

### Logistic Sigmoid $T = 1$



10/5/09 12



- ### Pseudo-Temperature
- Temperature = measure of thermal energy (heat)
  - Thermal energy = vibrational energy of molecules
  - A source of random motion
  - Pseudo-temperature = a measure of nondirected (random) change
  - Logistic sigmoid gives same equilibrium probabilities as Boltzmann-Gibbs distribution
- 10/5/09 15

### Transition Probability

Recall, change in energy  $\Delta E = -\Delta s_k h_k$   
 $= 2s_k h_k$

$$\Pr\{s'_k = \pm 1 | s_k = \mp 1\} = \sigma(\pm h_k) = \sigma(-s_k h_k)$$

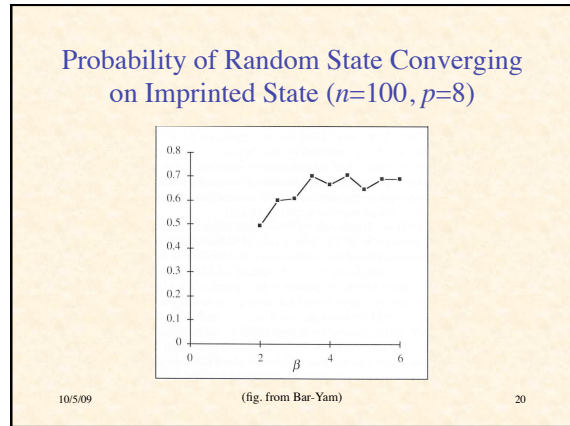
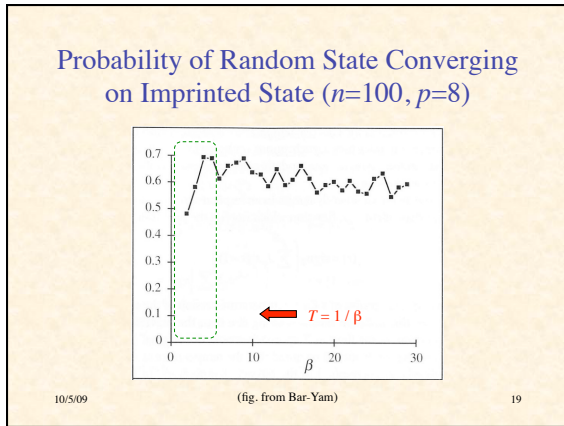
$$\Pr\{s_k \rightarrow -s_k\} = \frac{1}{1 + \exp(2s_k h_k / T)}$$

$$= \frac{1}{1 + \exp(\Delta E / T)}$$

10/5/09 16

- ### Stability
- Are stochastic Hopfield nets stable?
  - Thermal noise prevents absolute stability
  - But with symmetric weights:  
 average values  $\langle s_i \rangle$  become time - invariant
- 10/5/09 17

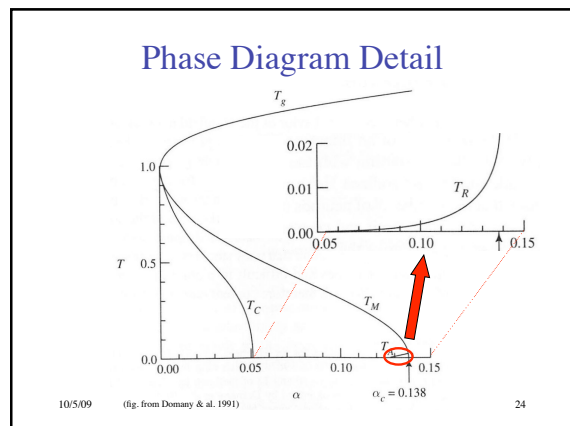
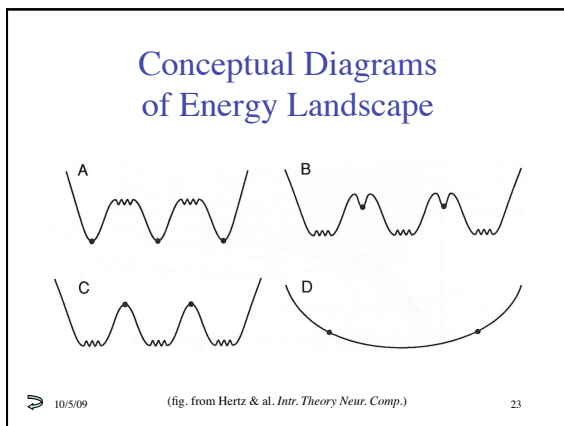
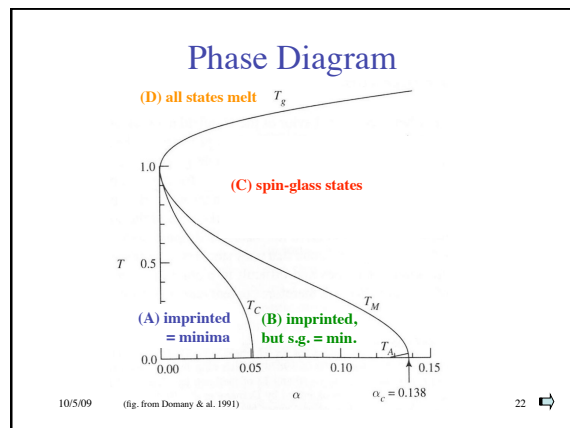
- ### Does “Thermal Noise” Improve Memory Performance?
- Experiments by Bar-Yam (pp. 316-20):
    - $n = 100$
    - $p = 8$
  - Random initial state
  - To allow convergence, after 20 cycles set  $T = 0$
  - How often does it converge to an imprinted pattern?
- 10/5/09 18



### Analysis of Stochastic Hopfield Network

- Complete analysis by Daniel J. Amit & colleagues in mid-80s
- See D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks*, Cambridge Univ. Press, 1989.
- The analysis is beyond the scope of this course

10/5/09 21



## Simulated Annealing

(Kirkpatrick, Gelatt & Vecchi, 1983)

10/5/09 25

## Dilemma

- In the early stages of search, we want a high temperature, so that we will explore the space and find the basins of the global minimum
- In the later stages we want a low temperature, so that we will relax into the global minimum and not wander away from it
- **Solution:** decrease the temperature gradually during search

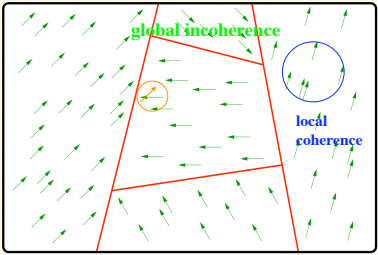
10/5/09 26

## Quenching vs. Annealing

- **Quenching:**
  - rapid cooling of a hot material
  - may result in defects & brittleness
  - local order but global disorder
  - locally low-energy, globally frustrated
- **Annealing:**
  - slow cooling (or alternate heating & cooling)
  - reaches equilibrium at each temperature
  - allows global order to emerge
  - achieves global low-energy state

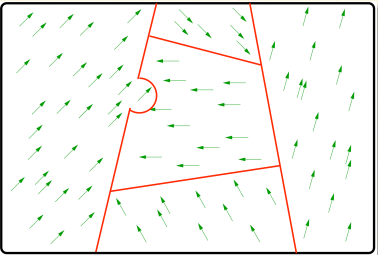
10/5/09 27

## Multiple Domains



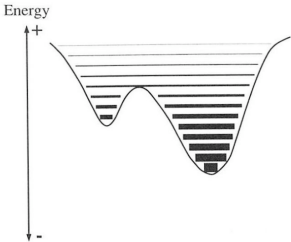
10/5/09 28

## Moving Domain Boundaries



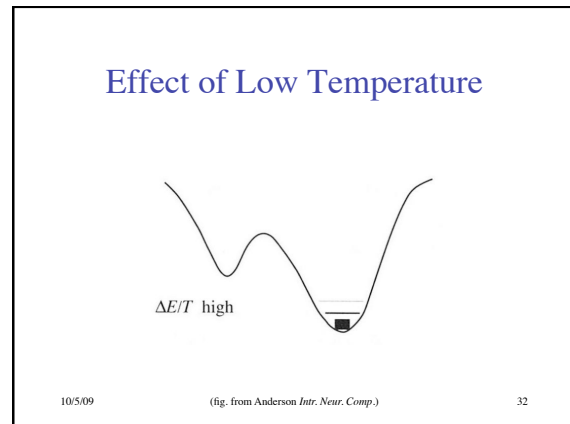
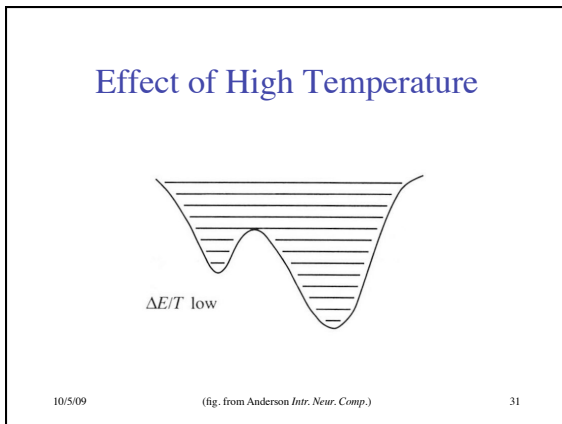
10/5/09 29

## Effect of Moderate Temperature



10/5/09 30

(fig. from Anderson *Intr. Neur. Comp.*)



### Annealing Schedule

- Controlled decrease of temperature
- Should be sufficiently slow to allow equilibrium to be reached at each temperature
- With sufficiently slow annealing, the global minimum will be found with probability 1
- Design of schedules is a topic of research

10/5/09 33

### Typical Practical Annealing Schedule

- **Initial temperature**  $T_0$  sufficiently high so all transitions allowed
- **Exponential cooling:**  $T_{k+1} = \alpha T_k$ 
  - typical  $0.8 < \alpha < 0.99$
  - at least 10 accepted transitions at each temp.
- **Final temperature:** three successive temperatures without required number of accepted transitions

10/5/09 34

### Summary

- Non-directed change (random motion) permits escape from local optima and spurious states
- Pseudo-temperature can be controlled to adjust relative degree of exploration and exploitation

10/5/09 35

### Hopfield Network for Task Assignment Problem

- Six tasks to be done (I, II, ..., VI)
- Six agents to do tasks (A, B, ..., F)
- They can do tasks at various rates
  - A (10, 5, 4, 6, 5, 1)
  - B (6, 4, 9, 7, 3, 2)
  - etc
- What is the optimal assignment of tasks to agents?

10/5/09 36


NetLogo Implementation of  
Task Assignment Problem

[Run TaskAssignment.nlogo](#)

10/5/09 37

Additional Bibliography

1. Anderson, J.A. *An Introduction to Neural Networks*, MIT, 1995.
2. Arbib, M. (ed.) *Handbook of Brain Theory & Neural Networks*, MIT, 1995.
3. Hertz, J., Krogh, A., & Palmer, R. G. *Introduction to the Theory of Neural Computation*, Addison-Wesley, 1991.

10/5/09  38