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

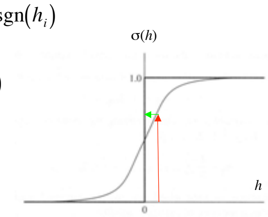
The Stochastic Neuron

Deterministic neuron: $s_i' = \mathrm{sgn}(h_i)$

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

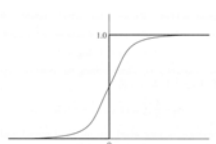
Stochastic neuron:

$$\Pr\{s_i' = +1\} = \sigma(h_i)$$
$$\Pr\{s_i' = -1\} = 1 - \sigma(h_i)$$



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

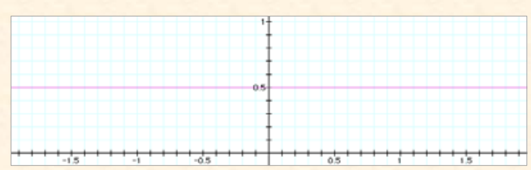
Properties of Logistic Sigmoid



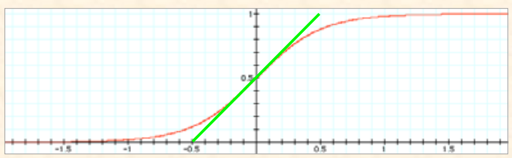
$$\sigma(h) = \frac{1}{1 + e^{-2h/T}}$$

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

Logistic Sigmoid With Varying *T*

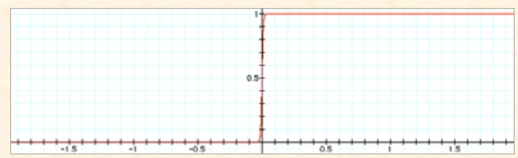


Logistic Sigmoid
$T = 0.5$

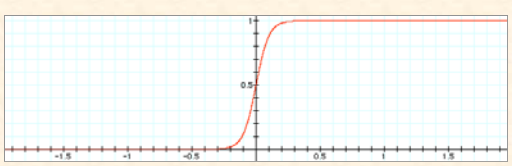


Slope at origin = $1 / 2T$

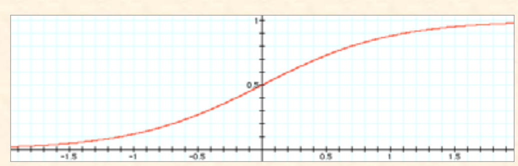
Logistic Sigmoid
$T = 0.01$



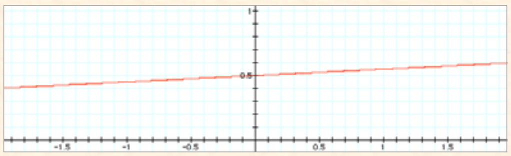
Logistic Sigmoid
$T = 0.1$



Logistic Sigmoid
$T = 1$

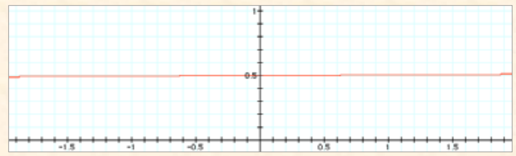


Logistic Sigmoid
$T = 10$



11/13/04 9

Logistic Sigmoid
$T = 100$



11/13/04 10

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

11/13/04 11

Transition Probability

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

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

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

11/13/04 12

Stability

- Are stochastic Hopfield nets stable?
- Thermal noise prevents absolute stability
- But with symmetric weights:

  average values $\langle s_i \rangle$ become time-invariant

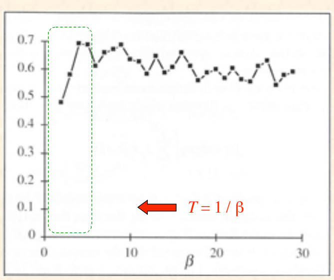
11/13/04 13

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?

11/13/04 14

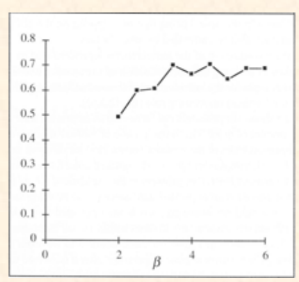
Probability of Random State Converging on Imprinted State (n=100, p=8)



(fig. from Bar-Yam)

11/13/04 15

Probability of Random State Converging on Imprinted State (n=100, p=8)



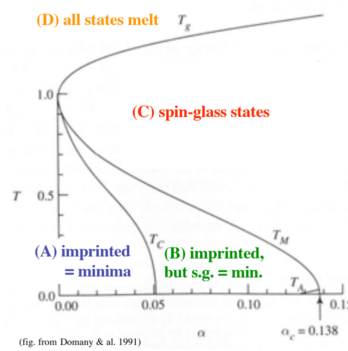
(fig. from Bar-Yam)

11/13/04 16

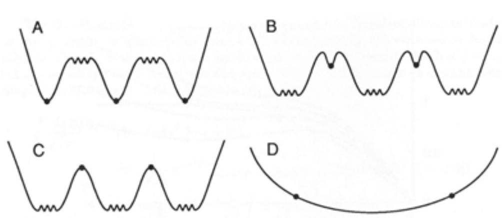
Analysis of Stochastic Hopfield Network

- 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

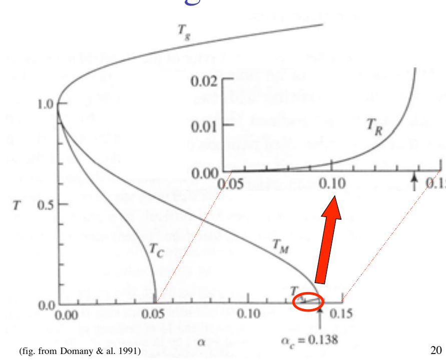
Phase Diagram



Conceptual Diagrams of Energy Landscape



Phase Diagram Detail



Simulated Annealing

(Kirkpatrick, Gelatt & Vecchi, 1983)

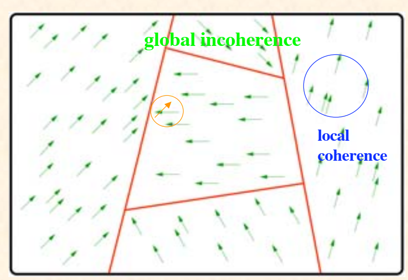
Dilemma

- 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

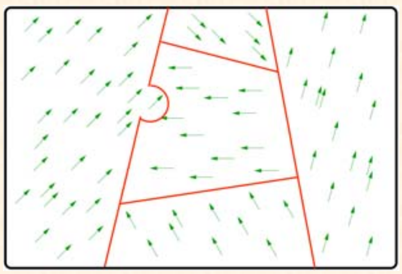
Quenching vs. Annealing

- 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

Multiple Domains

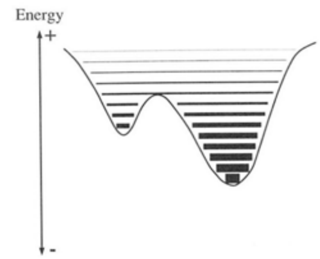


Moving Domain Boundaries



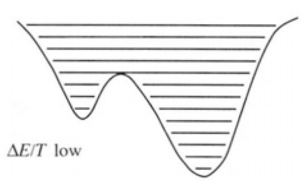
11/13/04 25

Effect of Moderate Temperature



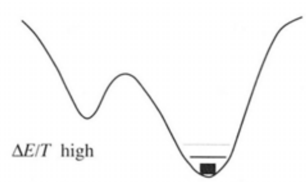
11/13/04 (fig. from Anderson *Intr. Neur. Comp.*) 26

Effect of High Temperature



11/13/04 (fig. from Anderson *Intr. Neur. Comp.*) 27

Effect of Low Temperature



11/13/04 (fig. from Anderson *Intr. Neur. Comp.*) 28

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

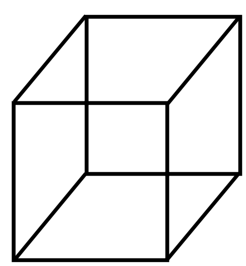
Typical Practical Annealing Schedule

- 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

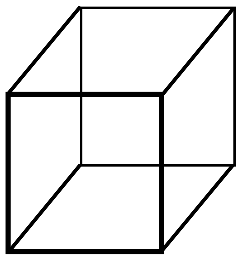
Demonstration of Boltzmann Machine & Necker Cube Example

Run ~mclennan/pub/cube/cubedemo

Necker Cube



Biased Necker Cube



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

Additional Bibliography

1. Kandel, E.R., & Schwartz, J.H. *Principles of Neural Science*, Elsevier, 1981.
2. Peters, A., Palay, S. L., & Webster, H. d. *The Fine Structure of the Nervous System*, 3rd ed., Oxford, 1991.
3. Anderson, J.A. *An Introduction to Neural Networks*, MIT, 1995.
4. Arbib, M. (ed.) *Handbook of Brain Theory & Neural Networks*, MIT, 1995.
5. Hertz, J., Krogh, A., & Palmer, R. G. *Introduction to the Theory of Neural Computation*, Addison-Wesley, 1991.