UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR/1003 - 12

SUMMER COLLEGE IN CONDENSED MATTER ON
" STATISTICAL PHYSICS OF FRUSTRATED SYSTEMS "

(28 July - 15 August 1997)

" Spin-glass mean field theory and replicas" (PART III)

presented by:

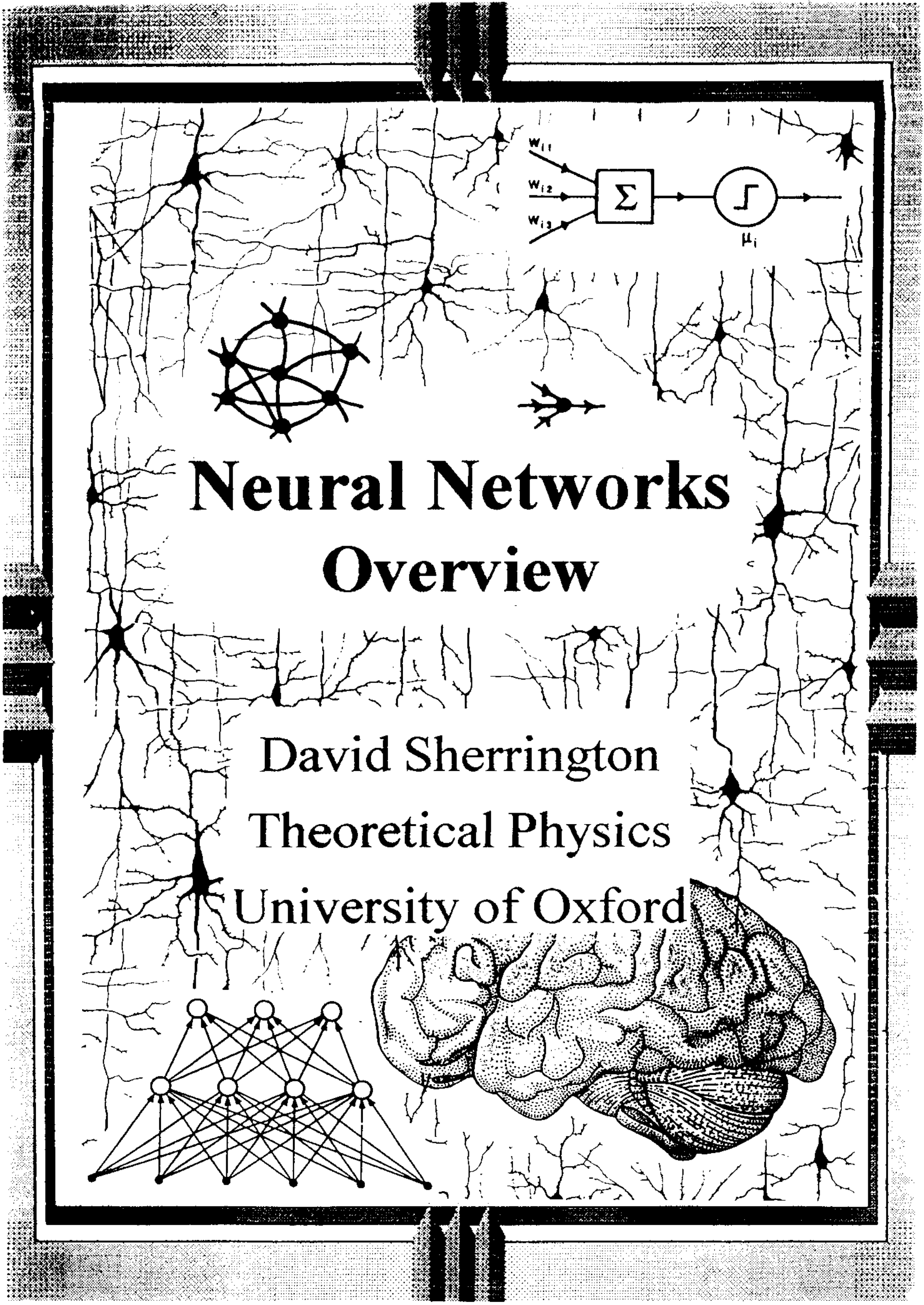
D. SHERRINGTON
University of Oxford
Department of Physics
Theoretical Physics
1 Keble Road
Oxford OX1 3NP
United Kingdom

Statistical Physics of Neural Networks

by David Sherrington

- **Neural networks**
  - Biology
  - Computer Science
  - Physics

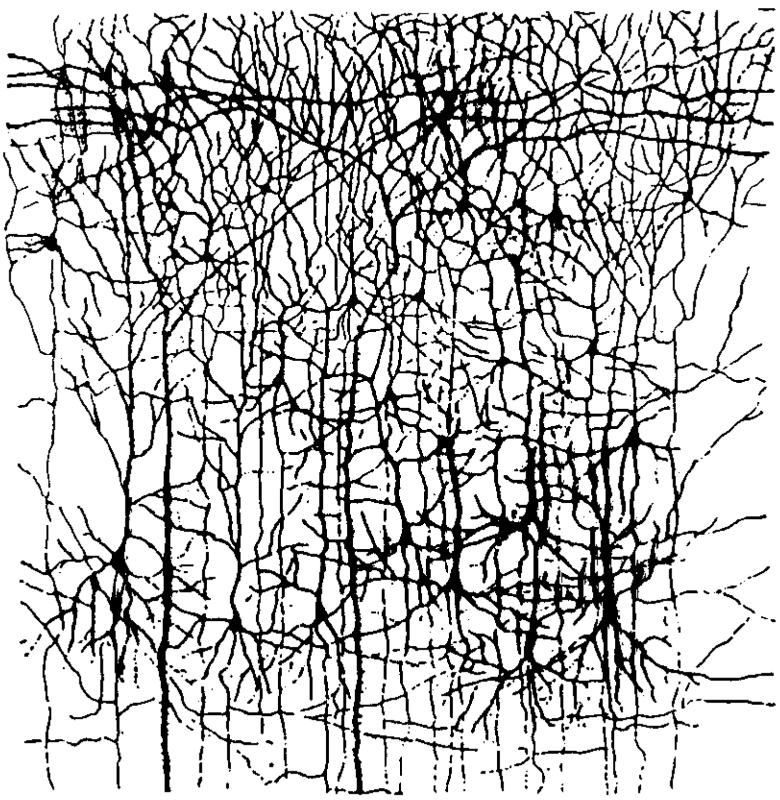
- **Statistical physics**
  - Statistical mechanics (eq^n + non-$equil^m$)
    - Neurodynamics → effective thermal stochastic dynamics
      - memory retrieval + generalization
    - Learning from examples
      - optimization + dynamical $implement^n$
  - Statistical relevance
    - 'typical' behaviour



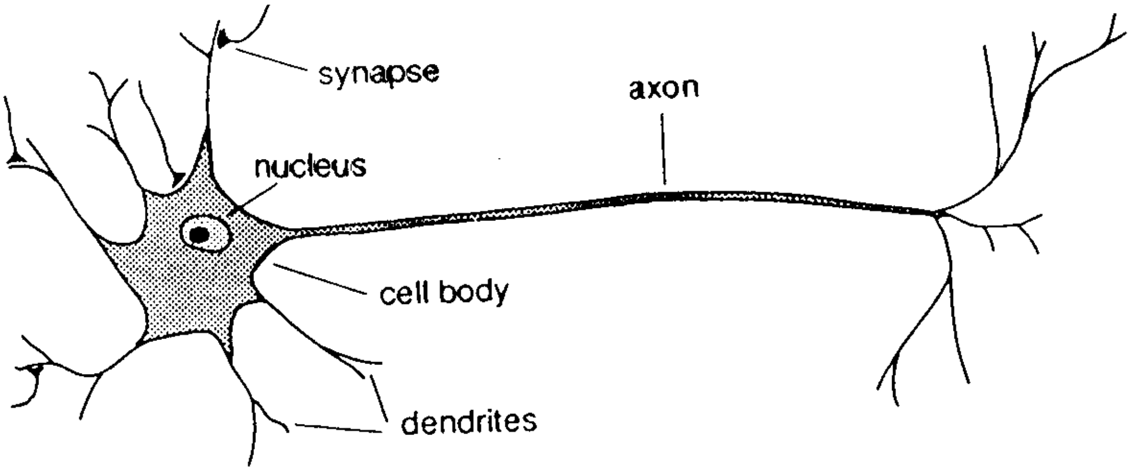
Neural Networks
Overview

David Sherrington
Theoretical Physics
University of Oxford

Cerebral cortex



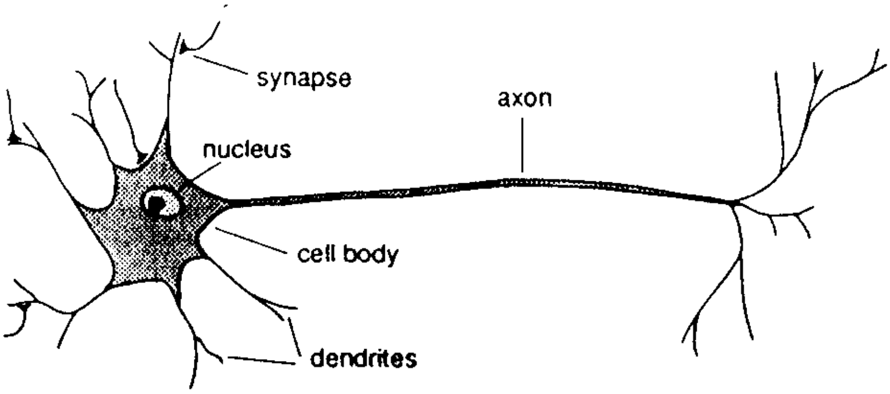
Typical neuron

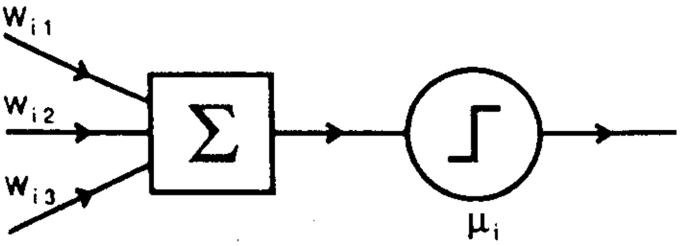


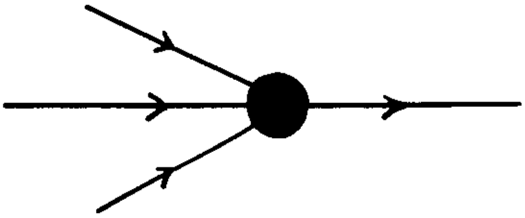
Observations about the brain

- Many neurons ~ 10^{11}
- Many connections ~ 10^5 per neuron
- Connections of all ranges
- Individual neurons relatively simple
- Can learn, retrieve & process a lot of data
- Operates quickly despite relatively slow elements
- Robust
- Capabilities complementary to conventional computers

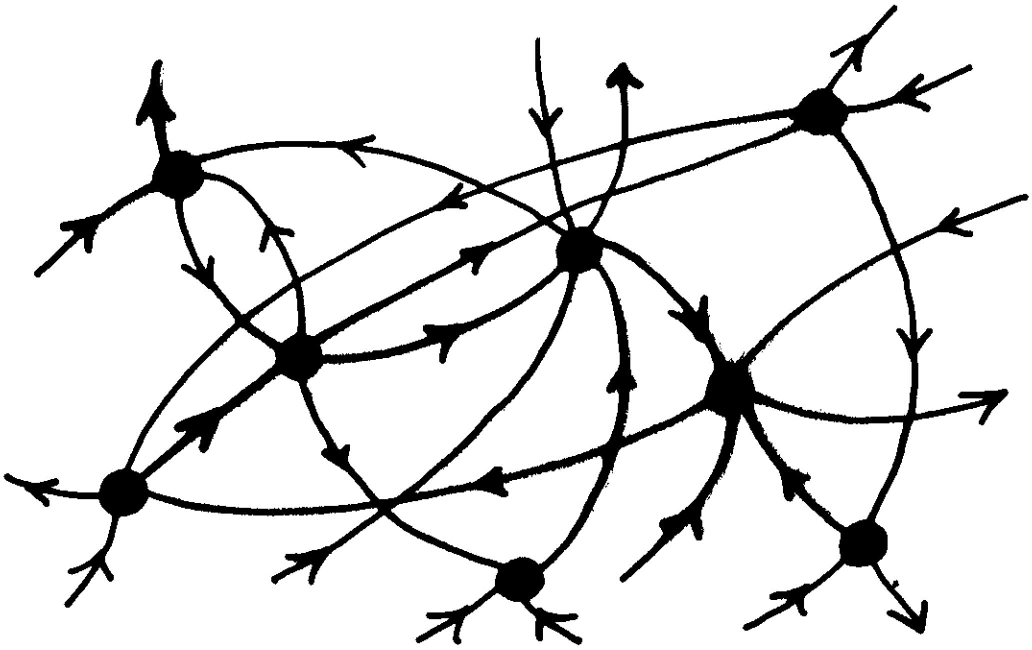
Schematise


synapse
axon
nucleus
cell body
dendrites


W_{i1}
W_{i2}
W_{i3}
Σ
μ_i



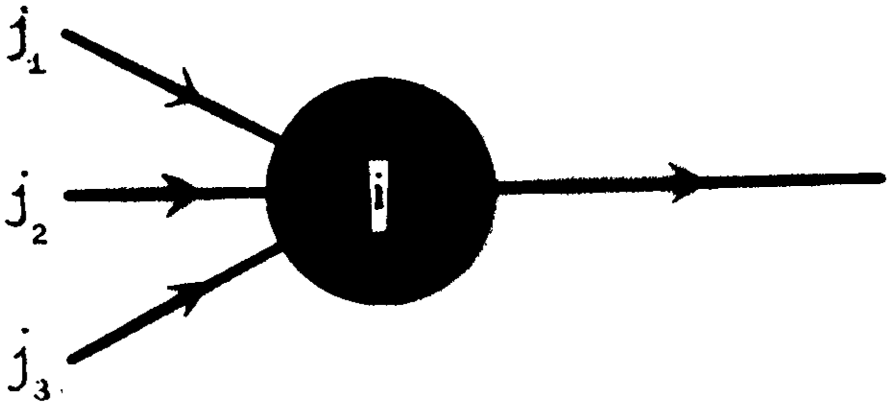
Recurrent networks



- Feedforward and backward
- ‘cerebral cortex’

Mathematical modelling

- Idealise



- Neuronal activity: V_i
- Synaptic weight: J_{ij}

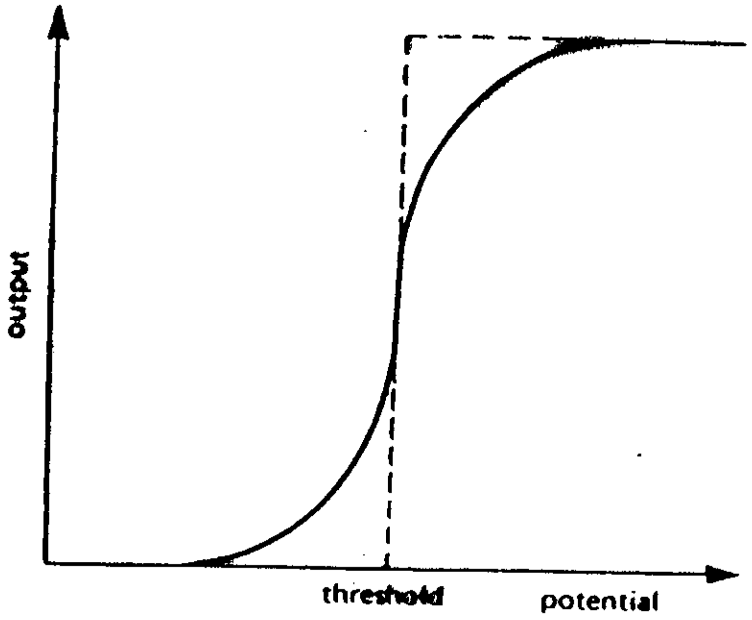
Excitatory: $J_{ij} > 0$ Inhibitory: $J_{ij} < 0$

- Total synaptic input to neuron i

$$U_i = \sum_j J_{ij} V_j$$

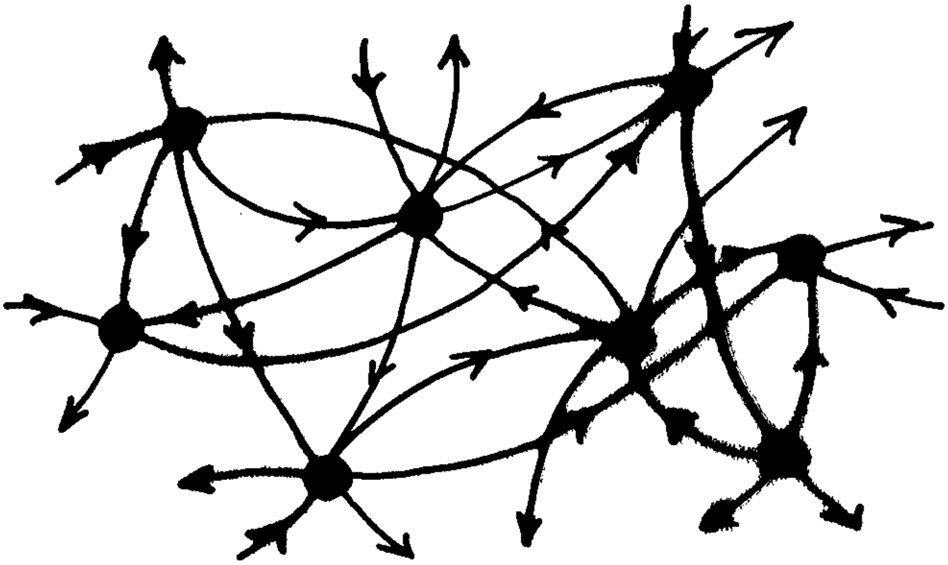
Consequence of input 'potential'

- **Output** activity of **neuron** *i*



- and so on through the network

Recurrent network

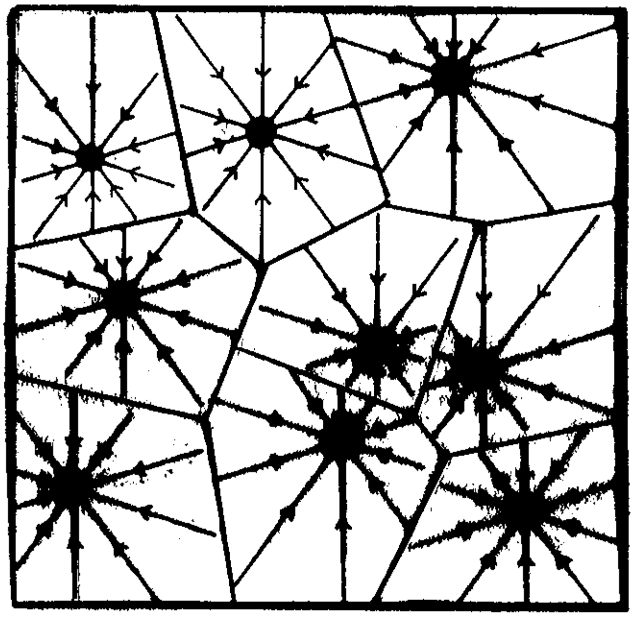


- dynamically stable/quasi-stable global firing pattern/attractor
- sequential global attractor
- chaos
- determined by synaptic weights, thresholds, starting state.

Concepts for neural memory

- **Patterns to be memorized**
  - Particular global activity states or sequences
- **Recall**
  - Retrieval of memorized pattern from distorted initial state / Association
  - Generalization
- **Learning**
  - Modification of local rules/ synaptic strengths/ thresholds

Attractors



- single state or more complex
- for associative memory,
  attractors ~ memorized patterns
- for many memorized patterns one requires competition/frustration

Note: There may also be attractors which do not correspond to stored patterns.

Issues

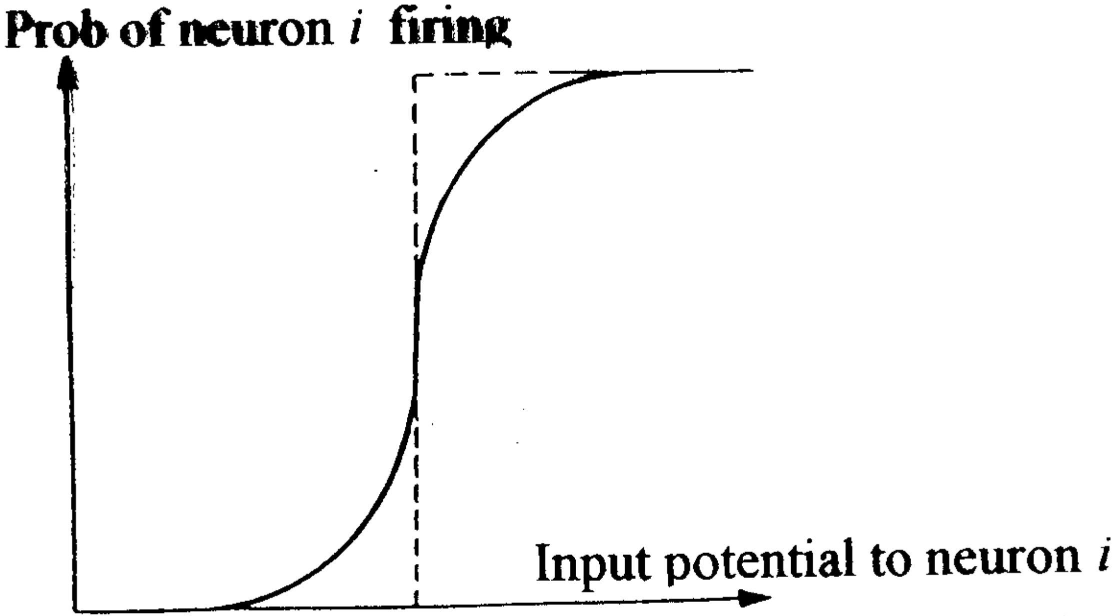
- Given an architecture, local rules and algorithms for synapses and thresholds
- determine performance

- Given information to learn
- determine what is optimally acheivable for particular aspects
- determine algorithms to acheive it
- determine consequences for other questions

Binary neurons

(McCulloch-Pitts)

- Idealise to two neural states
  - Firing: $V_i = 1\,;\quad \sigma_i = +1$
  - Non-firing: $V_i = 0\,;\quad \sigma_i = -1$
- Probabilistic update rule

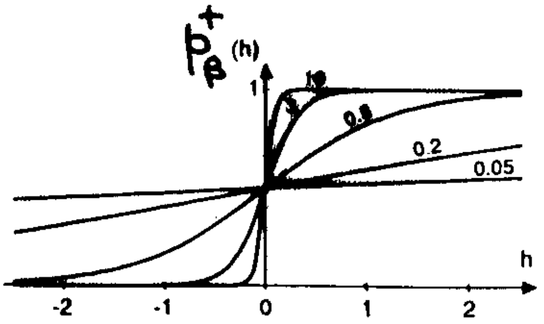


- Analogous to stochastic dynamics of Ising model of magnetism → statistical physics

Glauber dynamics in statistical physics

- Ising spin system
- Temperature T
- **Update** probability

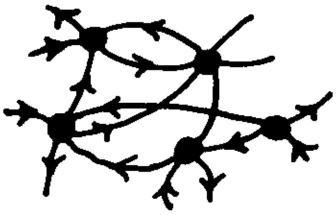
$$p(\sigma_i') = \frac{1}{2}\left\{1+\sigma_i' \tanh(\beta(\sum_j J_{ij}\sigma_j)\right\}$$



- 'Gain' = Inverse of temperature $\beta = T^{-1}$

Statistical mechanics of memory retrieval

- Recurrent networks



- Binary neurons : $\sigma_i = \pm 1$: firing / non-firing
- Patterns : $\{\xi_i^\mu = \pm 1\}$: $\mu = 1, \cdots p = \alpha N$
- Synapses : $\{J_{ij}\}$: code/store pattern info^n
- Dynamics : stochastic, determined by local fields $h_i = \sum_j J_{ij}\sigma_j - W_i$

  effective temperature T

- Measure of retrieval : overlap

$$m^\mu(t) = N^{-1}\sum_i \sigma_i(t)\,\xi_i^\mu$$

$\rightarrow O(1)$ retrieval / association

$\rightarrow O(N^{-1/2})$ non-retrieval

- Steady state (equilibrium)
- Non-equilibrium dynamics

Statistical relevance

- Typical properties

  eg. patterns $\{\xi_i^\mu\}$; $\mu = 1 \ldots p$

  p patterns chosen randomly from some distribution

  - distribution could be biased or not
  - concentrate on unbiased

- In principle
  - solve for particular randomly chosen set of p patterns
  - average over choice

- In practice (often)
  - use replica theory to average formally over choice of p patterns – average observables
  - study resulting effective replicated system with higher order 'interactions'

I. Steady state equilibrium : $t \to \infty$

- Simplification 1 :
  - Detailed balance* → Boltzmann distⁿ

$$p_{t\to\infty}(\underline{\sigma}) \sim \exp\left\{-\beta H(\underline{\sigma})\right\} ; \beta = T^{-1}$$

  - Random sequential dynamics

$$\to \quad H(\underline{\sigma}) = -\frac{1}{2}\sum_{ij} J_{ij}\sigma_i\sigma_j \quad - \sum_i W_i\sigma_i$$

cf. Ising model of magnetism

but with non-trivial exchange $\{J_{ij}\}$

  - Spin glasses

Frustration / competition in $\{J_{ij}\}$

→ many non-equivalent attractors for $T < T_g$

→ • Competition between excitatory + inhibitory synapses → many attractors / memory storage

---

* Symmetric synapses $J_{ij} = J_{ji}$ } Hopfield

No self-terms $J_{ii} = 0$

Retrieval in a recurrent network

• explicit analysis.

1). Hopfield model

$$H = -\sum_{(ij)} J_{ij}\, \sigma_i \sigma_j$$

$$\uparrow$$

Hebb synapses

$$J_{ij} = N^{-1} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu}$$

(i). p = 1 → Mattis model

↓

equivalent to simple ferromag^t

$$H = -N^{-1} \sum_{(ij)} \underset{\substack{\uparrow \\ \text{quenched} \pm 1}}{\xi_i \xi_j}\ \underset{\substack{\uparrow \\ \text{annealed} \pm 1 \\ \text{variables}}}{\sigma_i \sigma_j}$$

Gauge tranf: $\tau_i = \xi_i \sigma_i = \pm 1$

$$\rightarrow \quad H = -N^{-1} \sum_{(ij)} \tau_i \tau_j$$

= ∞-ranged ferromagnet.

$\langle \sigma_i \rangle = \underset{\uparrow}{m}\, \xi_i$ ←→ trivial to solve

magnetisation of ferromagnet : fn. of T

(ii) General p

$$H = N^{-1} \sum_{(ij)} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j$$

Gauge transf w.r.t. specified pattern

$$\tau_i = \xi_i^\nu \sigma_i$$

$$H = N^{-1} \sum_{(ij)} \tau_i \tau_j + N^{-1} \sum_{(ij)} \sum_{\substack{\mu \\ \neq\nu}} \xi_i^\mu \xi_j^\mu \overset{\xi_i^\nu \xi_j^\nu}{\wedge} \tau_i \tau_j$$

$\swarrow$

'signal'
trying to retrieve
$\langle \sigma_i \rangle \sim \xi_i^\nu$

$\downarrow$

'noise'
from other patterns
competes with
'ferromagnetic' term

$\rightarrow$ Signal / noise analysis

D. Sherrington
HK

• 'Thermodynamics'

$$Z = \underset{\underset{\mathrm{Tr}}{\downarrow}}{\sum_{\{\sigma\}}} \exp\{-\underset{\underset{\underset{\text{retrieval temperature.}}{\uparrow}}{\beta = 1/T}}{\underset{\uparrow}{\beta}} H(\sigma)\}$$

Hopfield-Hebb → mean field soln exact.

$$H = N^{-1} \sum_{(ij)} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j$$

$$= \frac{1}{2N} \sum_{\mu} \Big[\underset{\underset{\text{complete square}}{\uparrow}}{\Big(\sum_i \xi_i^{\mu} \sigma_i\Big)^2} \; \underset{\underset{\text{irrelevant const.}}{\uparrow}}{-p/2} \Big]$$

- Valuable identity } $\exp(\lambda a^2) = \int \frac{dx}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} + (2\lambda)^{1/2} a x\right)$

Hubbard-Stratonovich

↓

change of quadratic a^2 to linear a in exponent

$$a^{\mu} = \sum_i \xi_i^{\mu} \sigma_i$$

$$\Rightarrow \exp\Big(\cdot\cdot\big(\underset{\underset{\text{interacting spins}}{\uparrow}}{\sum_i} \xi_i \sigma_i\big)^2\Big) \rightarrow \exp\Big(\cdots \underset{\underset{\text{separated spins}}{\uparrow}}{\sum_i} \xi_i \sigma_i\Big)$$

6.

$$\rightarrow Z = \int \left(\prod_{\mu=1}^{p} dm^{\mu} \left(\frac{\beta N}{2\pi} \right)^{1/2} \right)$$

$$\mathrm{Tr} \exp \left\{ \sum_{\mu=1}^{p} \left(-\frac{N\beta}{2} (\tilde{m}^{\mu})^{2} - \beta \tilde{m}^{\mu} \sum_{i} \xi_{i}^{\mu} \sigma_{i} \right) - \frac{\beta p}{2} \right\}$$

$$= \int \left(\prod_{\mu} d\tilde{m}^{\mu} \left(\frac{\beta N}{2\pi} \right)^{1/2} \right) \exp(-N\beta f(\{M^{\mu}\}))$$

where

$$f(\{M^{\mu}\}) = \sum_{\mu=1}^{p} \frac{(m^{\mu})^{2}}{2}$$

$$-N\beta^{-1} \sum_{i} \ln \left[2 \cosh \beta \sum_{\mu=1}^{p} (m^{\mu} \xi_{i}^{\mu}) \right]$$

$$+ \frac{p}{2N}$$

(a) Intensive number of patterns

p finite : $\lim_{N \to \infty} (p/N) = 0$

$N \to \infty$ → partition function dominated by m^{μ} corresponding to maximum of $f(\{m^{\mu}\})$

↓

Steepest descents.

Extrema of $f(\{m^\mu\})$

- Self-consistency eqn.

$$\tilde{m}^\mu = N^{-1} \sum_i \xi_i^\mu \tanh\left(\beta \sum_\nu \tilde{m}^\nu \xi_i^\nu\right)$$

- **Strict thermodynamics.**
  - **Only $\tilde{m}^\mu$ corresponding to absolute minima of $f(m)$ are relevant.**

- **Neural attractors**
  - All minima of $f(m)$ are relevant*

- Interpretation of $\tilde{m}^\mu$

$$\tilde{m}^\mu = m^\mu \equiv N^{-1} \sum_i \underset{\substack{\uparrow \\ \text{pattern overlap.}}}{\xi_i^\mu} \langle \sigma_i \rangle$$

---

* possible to study dynamics directly.

(in principle: not so easy in practice).

Dynamics of ∞-range fm in field b

$$m(t+1) = N^{-1} \sum_i \sigma_i(t+1)$$

Update rule: $\sigma_i(t) \rightarrow \sigma_i(t+1)$

with prob. $\frac{1}{2}\left[1 - \sigma_i(t+1) \tanh \beta h_i(t)\right]$

and $h_i(t) = \frac{J_0}{N} \sum_j \sigma_j(t) + b$

$= J_0 m(t) + b$

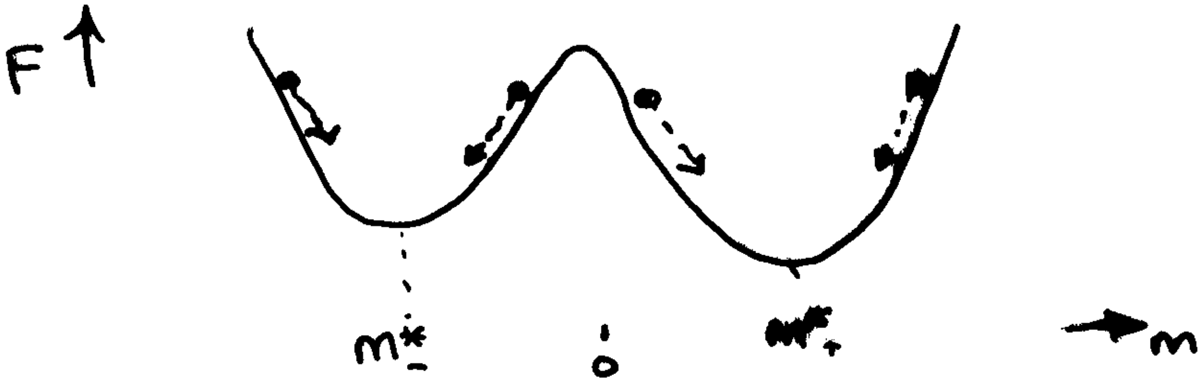
→ $m(t+1) = \tanh(\beta(J_0 m(t) + b))$

→ attractors

$$m^* = \tanh(\beta(J_0 m^* + b))$$

$m(0) > 0 \rightarrow m^*_+ > 0$

$m(0) < 0 \rightarrow m^*_- < 0$



$m^*_+ > |m^*_-|$ for $b \neq 0$

Finite # patterns

$$m_\mu(t) = N^{-1} \sum_i \xi_i^\mu \sigma_i(t)$$

$$m_\mu(t+1) = N^{-1} \sum_i \xi_i^\mu \tanh(\beta h_i(t))$$

$$h_i(t) = \sum_j J_{ij} \sigma_j(t)$$

$$\uparrow \; J_{ij} = N^{-1} \sum_\nu \xi_i^\nu \xi_j^\nu$$

$$= \sum_\nu \xi_i^\nu m^\nu(t)$$

$$\therefore \boxed{m_\mu(t+1) = N^{-1} \sum_i \xi_i^\mu \tanh\Big(\sum_\nu \xi_i^\nu m^\nu(t)\Big)}$$

Correct only if no of patterns $< O(\sqrt{N})$.

3. HK.

Attractor solutions : p intensive : $N \to \infty$

Average over choice of $\{\xi\}$.

1) $T=0$

(i) All embedded patterns $\{\xi^\mu\}$ are solutions

ie. p solutions $m^\mu = 1$: one pattern
$= 0$: rest

→ RETRIEVAL

(ii) Mixed solutions with more than one $m^\mu \neq 0$

→ SPURIOUS MIXTURES

2) $0.46 > T > 0$

(i) Solutions $m^\mu = m \sim O(1)$: one pattern
$= 0$: rest

→ (IMPERFECT) RETRIEVAL

(ii) Mixed solutions

3) $1 > T > 0.46$

Only type (i) solutions survive
extensive barriers

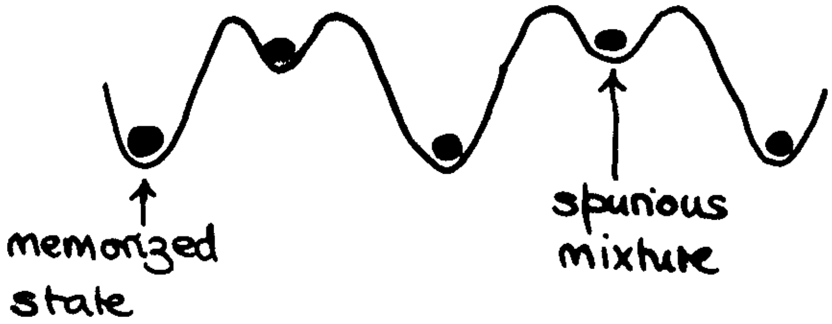
stochastic noise removes confusion.

4) $T > 1$

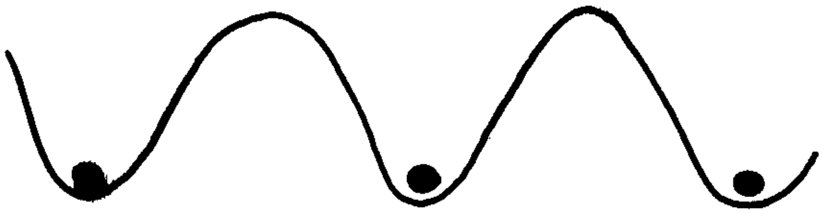
Only paramagnetic solution $m^\mu = 0$: all μ

Schematically (for appropriate $\{J_{ij}\}$)

- $T=0$



- $0 < T < T_c$



Spurious states smoothed out.
Memories remain

- $T > T_c$

Too much stochastic 'noise' destroys recall.

(b) Extensive number of patterns

$$p = \underset{\substack{\uparrow \\ \text{storage capacity}}}{\alpha} N$$

Cannot assume $f(\{m^{\mu}\})$ intensive.

→ spin glass techniques

- Concentrate on basin of attraction of one pattern.
- Average over other pattern choices.
- Note relevant quantity is $\ln Z(\{\xi\})$
- Use replica theory to do average correctly.

- Several manipulations lead to coupled eqns (RS)

$$m = f(m, q)$$

$$q = g(m, q)$$

where m = overlap with pattern of interest.

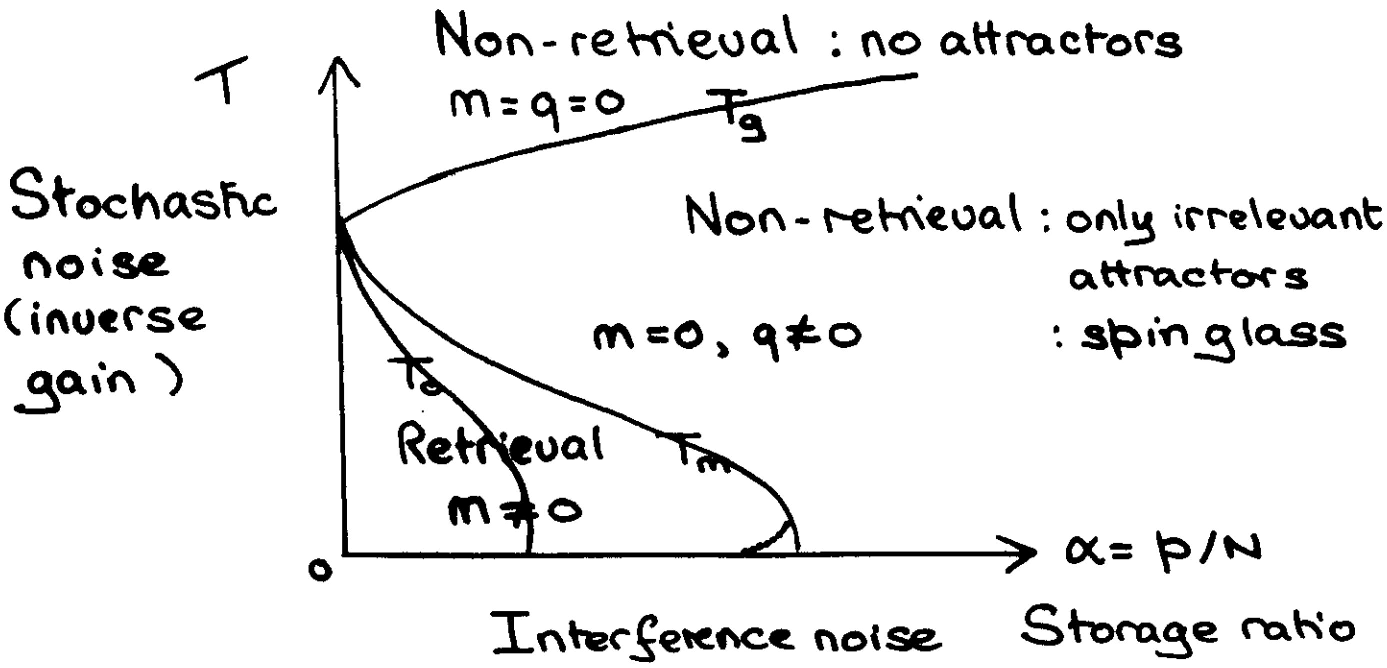
$q = \overline{\langle \sigma_i \rangle^2}$; 'spin glass parameter'

- RSB. small effect near $T = 0$ + $\alpha = \alpha_{max}$.

Attractor phase diagram

Hopfield model: $J_{ij} = N^{-1} \sum_{\mu=1}^{p=\alpha N} \xi_i^\mu \xi_j^\mu$

Random ξ.



Order parameters: overlap : $m^\mu = N^{-1} \sum_i \xi_i^\mu \langle \sigma_i \rangle$

spin-glass : $q = N^{-1} \sum_i \langle \sigma_i \rangle^2$

Note
- # of patterns stored is proportional to connectivity
- Above diagram is for full connectivity

  for connectivity c per site → total capicity ~ c

  ∴ Desire for high capacity implies high connectivity

(Also high barriers, good resilience to damage).

RSB. Subtleties found in region marked by /

(ii) More details (Hopfield model · AGS)

Recall $$Z \sim \int \left(\prod_{\mu=1}^{p} dm^{\mu}\right) \mathrm{Tr}\, \exp\left\{\sum_{\mu}\left(-N\beta \frac{m^{\mu 2}}{2} - \beta m^{\mu} \sum_i \xi_i^{\mu} \sigma_i\right) - \frac{\beta p}{2}\right\}$$

⇒ Replicating

$$Z^n \sim \int \left(\prod_{\mu=1}^{p} \prod_{\alpha=1}^{n} dm^{\mu\alpha}\right) \mathrm{Tr}\, \exp\left\{\sum_{\mu}\sum_{\alpha}\left(-N\beta \frac{m^{\mu\alpha 2}}{2} - \beta m^{\mu\alpha} \sum_i \xi_i^{\mu} \sigma_i^{\alpha}\right) - \frac{\beta n p}{2}\right\}$$

Average over $\{\xi\}$: random uncorrelated patterns. *

$$\langle Z^n \rangle_{\{\xi\}} \sim \exp\left(-\frac{np\beta}{2}\right) \sum_{\{\sigma^\alpha\}} \int \prod_{\mu=1}^{p} \prod_{\alpha=1}^{n} dm^{\mu\alpha}$$

$$\times \exp\left\{-N \sum_{\mu}\left[\beta \sum_{\alpha} (m^{\mu\alpha})^2/2 + N^{-1} \sum_i \ln\cosh\left(\beta \sum_{\alpha} m^{\mu\alpha} \sigma_i^{\alpha}\right)\right]\right\}$$

↑

Now: Consider just one pattern condensed out, labelled 1. All other m^{μ} small → expand ln cosh to second order

$$\prod_{\mu>1} \exp\left[-\sum_i \ln\cosh\left(\beta \sum_{\alpha} m^{\mu\alpha} \sigma_i^{\alpha}\right)\right]$$

$$\to \prod_{\mu>1} \exp\left[\frac{\beta^2}{2} \sum_{\alpha\beta} m^{\mu\alpha} m^{\mu\beta} \sum_i \sigma_i^{\alpha} \sigma_i^{\beta}\right]$$

Decouple $\sum_i \sigma_i^\alpha \sigma_i^\beta$ from m's by introducing

spin-glass like order parameter $q^{\alpha\beta}$

$$1 = \int dq^{\alpha\beta}\, \delta\Big(q^{\alpha\beta} - N^{-1}\sum_i \sigma_i^\alpha\sigma_i^\beta\Big)$$

and integral representation of δ-fn

$$= \int dq^{\alpha\beta}\int \frac{d\tilde r^{\alpha\beta}}{2\pi}\, \exp\Big(i\tilde r^{\alpha\beta}\big(q^{\alpha\beta} - N^{-1}\sum_i\sigma_i^\alpha\sigma_i^\beta\big)\Big)$$

$$\to \int \prod_{(\alpha\beta)} dq^{\alpha\beta}\,\frac{d\tilde r^{\alpha\beta}}{2\pi}\, \exp\Big\{ i\sum_{(\alpha\beta)} \tilde r^{\alpha\beta}\big(q^{\alpha\beta}-N^{-1}\sum_i \sigma_i^\alpha\sigma_i^\beta\big)$$
$$+ N\beta^2\sum_{\mu>1}\sum_{(\alpha\beta)} q^{\alpha\beta} m^{\mu\alpha}m^{\mu\beta}$$
$$+ \frac{N\beta^2}{2}\sum_{\mu>1}\sum_\alpha (m^{\mu\alpha})^2\Big\}$$

↗ for expression at bottom of last transparency

In $\langle Z^n\rangle$ now have σ in form separable in i

$$\sim \exp\Big\{ \underset{\substack{\uparrow\\ \sim N}}{\sum_i} \Big(\ln\cosh\big(\beta\sum_\alpha \underset{\substack{\uparrow\\ \text{overlap with condensed pattern}}}{m^\alpha}\sigma_i^\alpha\big) \underbrace{- iN^{-1}\underset{\substack{\uparrow\\ \text{scales as } p=\alpha N}}{\sum_{(\alpha\beta)}}\tilde r^{\alpha\beta}\sigma_i^\alpha\sigma_i^\beta\Big)\Big\}}_{O(1)\text{ w.r.t. }N}$$

$\underbrace{\qquad\qquad}$

extremal dominance

$$\langle Z^n\rangle \sim \int \prod dm^\alpha \prod dq^{\alpha\beta}\, d\tilde r^{\alpha\beta}\, \exp(-N\beta \underset{\substack{\uparrow\\ \text{intensive}}}{\Phi})$$

$$\langle Z^n \rangle \sim \int \prod dm^\alpha \prod dq^{\alpha\beta}\; dr^{\alpha\beta} \exp(-N\beta\Phi)$$

$$\Phi = \frac{np}{2N} + \frac{1}{2}\sum_\alpha (m^\alpha)^2 + \frac{(p-1)}{2\beta N}\,\mathrm{Tr}\ln\{(1-\beta)\delta_{\alpha\beta} - \beta q^{\alpha\beta}\}$$

$$+ \alpha\beta \sum_{(\alpha\beta)} r^{\alpha\beta} q^{\alpha\beta}$$

$$- \beta^{-1} \ln \underset{\substack{\uparrow \\ \text{single site}}}{\sum_{\{\sigma^\alpha\}}} \exp\Big\{ \ln\cosh\Big(\beta \sum_\alpha m^\alpha \sigma^\alpha\Big) + \underset{\substack{\uparrow \\ \text{storage ratio}}}{\alpha}\,\underset{\substack{\uparrow \\ \text{inverse temp}}}{\beta} \sum_{\underset{\uparrow\ \text{replica labels}}{(\alpha\beta)}} r^{\alpha\beta}\sigma^\alpha\sigma^\beta \Big\}$$

[illegible]

Replica symmetric ansatz.

$$\begin{aligned} m^\alpha &= m \\ q^{\alpha\beta} &= q \\ r^{\alpha\beta} &= r \end{aligned} \quad \Rightarrow \quad \exp\{\ \} \updownarrow$$

$$\frac{1}{2}\sum_{\xi=\pm 1} \exp\Big\{ \beta m \xi \sum_\alpha \sigma^\alpha + \frac{\alpha\beta r}{2}\Big(\sum_\alpha \sigma^\alpha\Big)^2 - \frac{\alpha\beta r}{2} \Big\}$$

$\uparrow$ complete square

$\therefore$ Hubbard-Stratonovich to linear $\sum_\alpha \sigma^\alpha$

$\downarrow$

straightforward $n \to 0$

m, q, r given by extrema of

$$\Psi = \frac{\alpha}{2} + \frac{m^2}{2} + \alpha\beta r\frac{(1-q)}{2}$$

$$+(\alpha/2\beta)\left[\ln(1-q(1-q)) - \beta q/(1-\beta(1-q))\right]$$

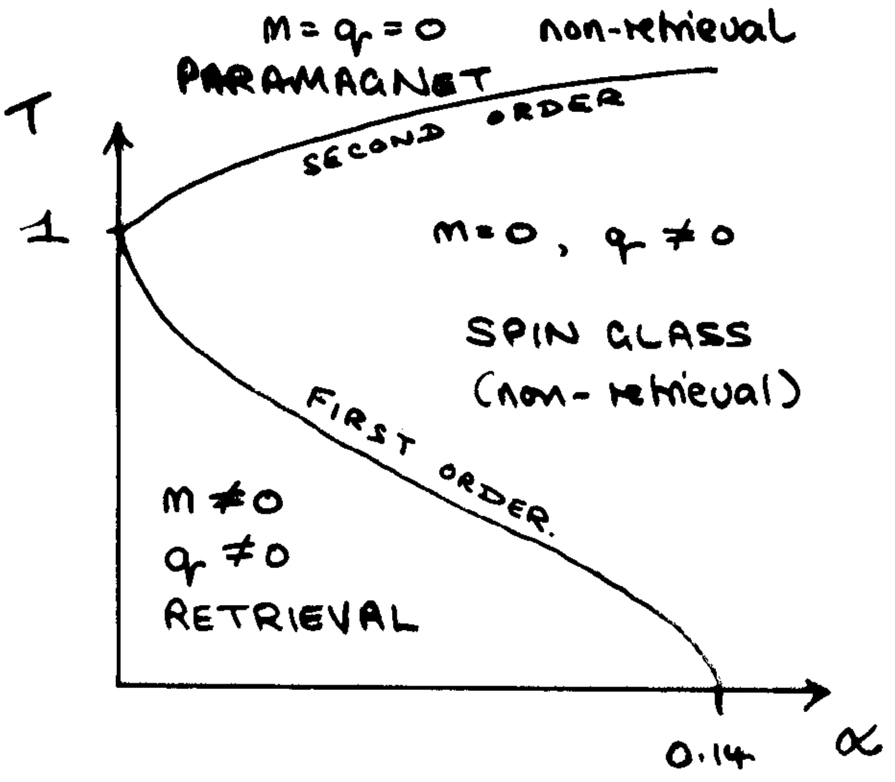
$$-\beta^{-1}\int dz e^{-z^2/2}\langle \ln[2\cosh\beta(z\sqrt{\alpha r} + m\xi)]\rangle_{\xi=\pm 1}$$

$$\Rightarrow \quad m = \int\frac{dz}{\sqrt{2\pi}}\exp(-z^2/2)\tanh[\beta(z\sqrt{\alpha r} + m)]$$

$$q = \int\frac{dz}{\sqrt{2\pi}}\exp(-z^2/2)\tanh^2[\beta(z\sqrt{\alpha r} + m)]$$

$$r = q(1-\beta(1-q))^{-2}$$

⇒ Phase diagram



overlap : $m^1 = N^{-1}\sum_i \xi_i^1\langle\sigma_i\rangle$: arbitrary nominated pattern

spin-glass : $q = N^{-1}\sum_i \langle\sigma_i\rangle^2$

NB Can extend to look at mixture states eg allow for 3 patterns to be macroscop. condensed

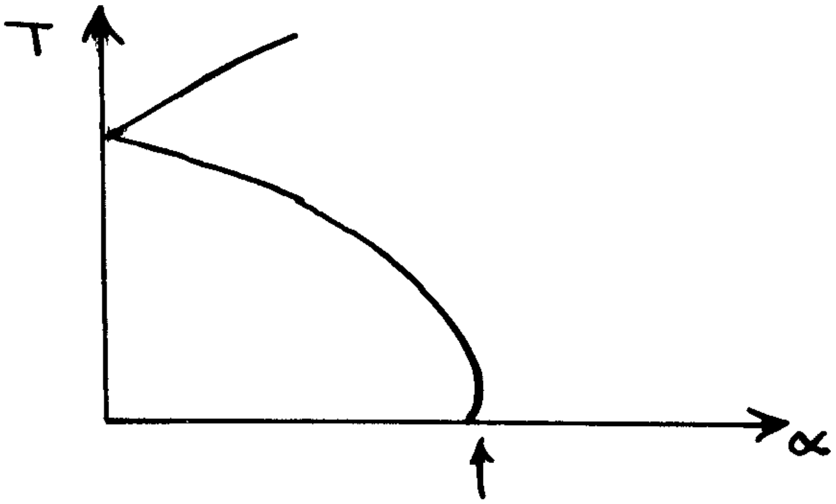
$$m^1, m^2, m^3 \sim O(1)$$

$$m^{p>3} \sim O(N^{-1})$$

NB For dynamic stability we want self-consistently stable solns to eqns for m : not necc. minima of $F = -kT \ln Z$.

Note

RS theory actually give re-entrance



↑

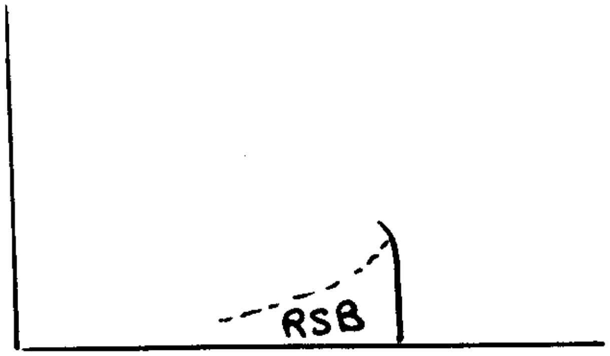
artifice of RS

↓

need RSB / Parisi th

→ vertical end.

But RSB not great.



NN

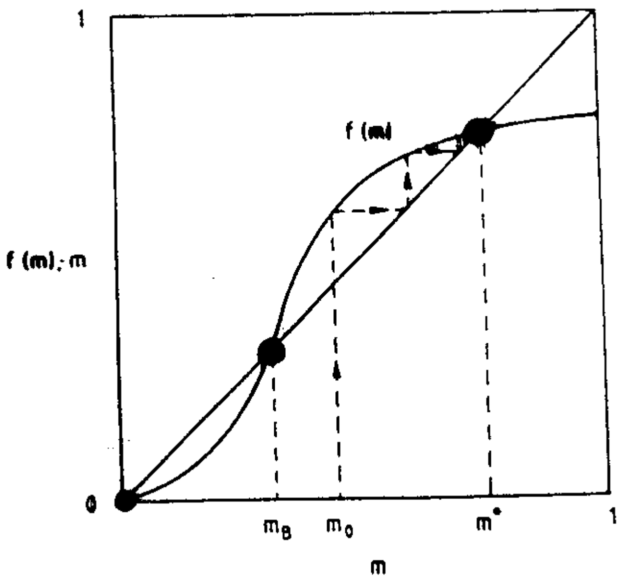
Randomly connectd network

- Simplification 2 (alternative)
  - Dilute connectivity : $c \ll \ln N$
  - Asymmetric synapses : $J_{ij} \neq J_{ji}$
    ↑ At least one = 0

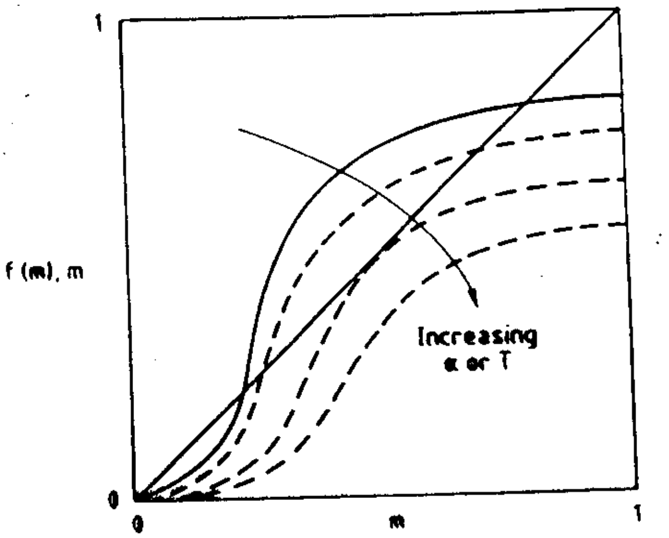
→ Retrieval as iterative map

$$\frac{dm}{dt} = \underset{\substack{\uparrow \\ \text{depends on } T,\ \alpha = p/c}}{f(m)} - m \quad \text{or} \quad m(t+1) = f(m(t))$$

- Fixed points →
  - retrieval quality
  - size of attractor basin
  - retrieval phase diagram
- Continuous or discontinuous transitions (2nd or 1st order)



← →
retrieval



→ phase diag^m.

Application of statistical physics to neural networks

- (i) • Given architecture + algorithm for $\{J_{ij}\}$

  determine retrieval / performance properties

- (ii) • Given examples find $\{J_{ij}\}$ to optimize

  some performance measures.

  - What is achievable ?
  - How to achieve it ?
  - Consequences of choice for other behavioural measures

- Statistically relevant answers
  - Typical rather than worst-case.

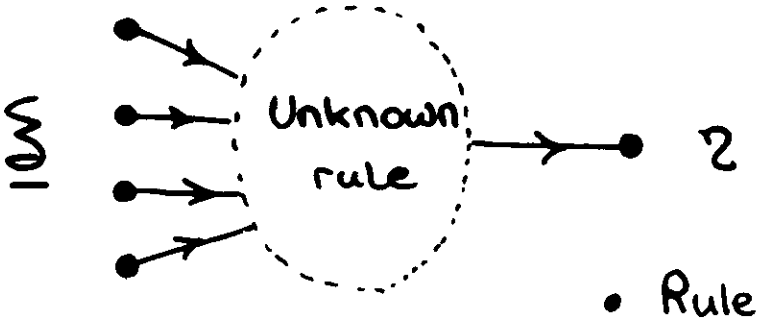
History

(i) Hopfield (1982), Amit et. al. (1985) .....

(ii) Gardner (1987), Wong + S (1990), Sompolinsky et. al (1990), ...

Learning a 'rule' from examples

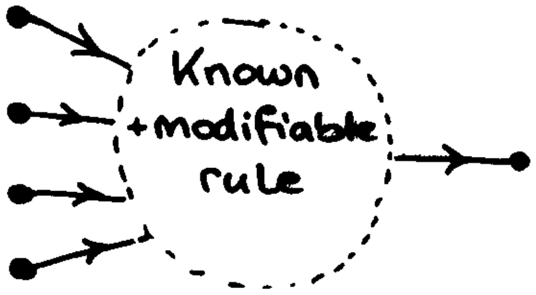
Rule



$$\eta = V(\{\underline{\xi}\})$$

- Rule unknown to student
- Teacher provides examples

Network trying to learn rule (student)



$$\eta = B(\{\underline{\xi}\})$$

↑ Modify to improve performance

? optimization ?

Error measures

- Training error

$$E_t = \sum_{\mu=1}^{p} e\left(B(\{\xi^\mu\}), V(\{\xi^\mu\})\right)$$

p = size of training set.

training set of examples; $\mu = 1, \dots p$

correct output (given by teacher)

output of 'student' network

error on pattern/association μ

$e = 0$ if correct

> 0 if incorrect

eg. $e(x) = x^2$

- Fractional training error

$$\mathcal{E}_t = E_t / p$$

- Generalization error

$$\mathcal{E}_g = \left\langle e\left(B(\{\xi\}), V(\{\xi\})\right)\right\rangle$$

averaged over all possible examples

- Optimization   Minimization of $\mathcal{E}$

Optimization as a problem in statistical physics

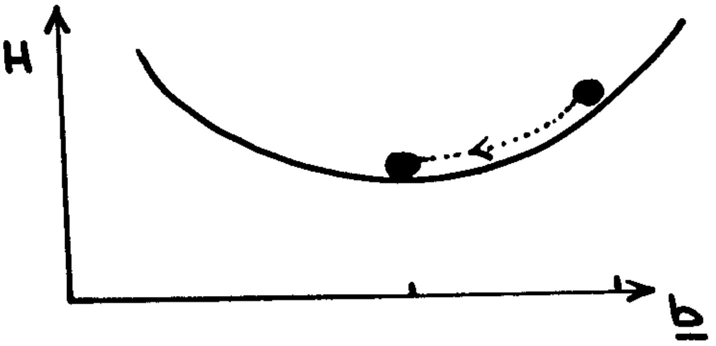
- To minimize $H_{\{a\}}(\{b\})$

$$\underset{\text{fixed}}{\{a\}} \qquad \underset{\text{variable}}{\{b\}}$$

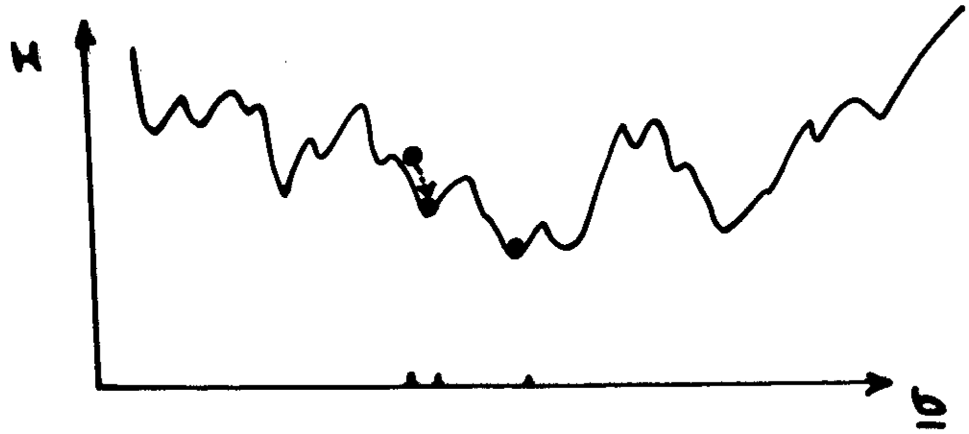
(i) Try all values of $\{b\}$: impractical if many choices

(ii) Iterative improvement :

- 'Random' start for $\underline{b}$
- Find slope of $H(\underline{b})$
- Change $\underline{b}$ in direction of reducing $H(\underline{b})$
- OK for



- Problems for



Hard optimization problem.

eg travelling salesman, training n.n.

- **Simulated annealing** (probalistic hill-climbing)
  - Microscopic dynamics (computer sim^n)

    on variables $\underline{b}$ at fixed $\underline{a}$

    - Stochasticity parameter T_A (annealing temp.)

      $T_A = 0$ accept only moves reducing H

      $T_A > 0 \to$ prob to move uphill $\sim e^{-\Delta H / T_A}$

  - Gradually reduce T_A

    Argue that infinitely slow cool $\to \underset{\underline{b}}{\mathrm{Min}}\, H_{\underline{a}}(\underline{b})$

- $\to$ **Quasi-statistical mechanics** / dynamics

  eg. Langevin dynamics

$$\frac{d\underline{b}}{dt} = -\nabla_{\underline{b}} H_{\underline{a}}(\underline{b}) + \underset{\substack{\uparrow \\ \text{white noise of} \\ \text{variance } T_A}}{\eta(t)}$$

  or Metropolis

$$\to p_{\infty\,\underline{a}}(\underline{b}) \sim \exp\left(-H_{\underline{a}}(\underline{b}) / T_A\right)$$

  As $T_A \to 0$ this prob $\to \underset{\underline{b}}{\min}\, H_{\underline{a}}(\underline{b})$

Statistical physics of optimization (analytic)

- distrib^n: $p(\underline{b}) \sim \exp(-H(\underline{b})/T_A)$
- 'Thermal average cost':

$$\langle H \rangle_{T_A} = \sum_{\underline{b}} H(\underline{b})\, p(\underline{b})$$

- Minimum cost:

$$H_{min} = \lim_{T_A \to 0} \langle H \rangle_{T_A}$$

- Typical behaviour: average over allowed $\{\underline{a}\}$

$$\langle H_{min} \rangle_{\underline{a}} = \lim_{T_A \to 0} \langle\langle H \rangle_{T_A} \rangle_{\underline{a}}$$

↑

analagous problem to that experienced in s.m. of retrieval or thermodyn. of spin glass (but diff. variables + functions).

NB. various constraints.

$$\langle H \rangle_{T_A} = -\frac{\partial}{\partial \beta_A} \ln Z_A \qquad ; \ \beta_A = T_A^{-1}$$

$$Z_A = \sum_{\underline{b}} \exp(-\beta_A H(\underline{b}))$$

$$F_{T_A} = -T_A \ln Z_A \qquad ; \ H_{min} = \lim_{T_A \to 0} F_{T_A}$$

Optimization

- **Cost** to minimize

$$H_{\underset{\substack{\uparrow \\ \text{fixed parameters}}}{\{a\}}}(\{\underset{\substack{\uparrow \\ \text{variables to adjust}}}{b}\})$$

eg.

$\{a\}$ includes example associations given by teacher

$\{b\}$ parameters of student network

Typical n.n. optimization problem

Cost (to minimize)

$$E = -\sum_{\mu=1}^{p} g(\Lambda^{\mu})$$

↕ stability field

$$\Lambda^{\mu} = \eta_0^{\mu} \sum_i J_i \xi_i^{\mu}$$

↑

p training associations (chosen from much larger set of possible assoc^{ns})

↑

eg N binary input neurons → 2^N possible input states

How well can one do if $p=\alpha N$?

Constraints

eg. $\sum_i J_i^2 = N$: spherical

$J_i = \pm 1$: Ising.

Statistical typicality

Average minimum E over choice of the p inputs from the full set of possibilities

Method (neural net training optimization)

- Want $\lim_{T_A \to 0} \left(-T_A \langle \ln Z_{\{\xi\}} \rangle_{\{\xi\}} \right)$

average over special choice of p examples

partition function for fixed set of examples

$$Z_{\{\xi\}} = \sum_{\substack{\{J\} \\ \text{with constraints}}} \exp\left\{ \beta_A \sum_{\mu} g(\Lambda^{\mu}) \right\}$$

Replica theory

$$\langle \ln Z \rangle = \lim_{n \to 0} \frac{1}{n} \left\{ \langle Z^n \rangle - 1 \right\}$$

hard to average

average over pattern choices.

n replicas / extra labels all with same $\{\xi\}$

Easier to average formally

$$\langle Z^n \rangle \equiv Z_{\text{eff}}$$

Partition fn of effective system of higher-dimensional units J_i^{α}, $\alpha = 1..n$ interacting through more complicated but non-disordered interactions.

Solve + take limit $n \to 0$

Possible, in principle, for $N \to$ large

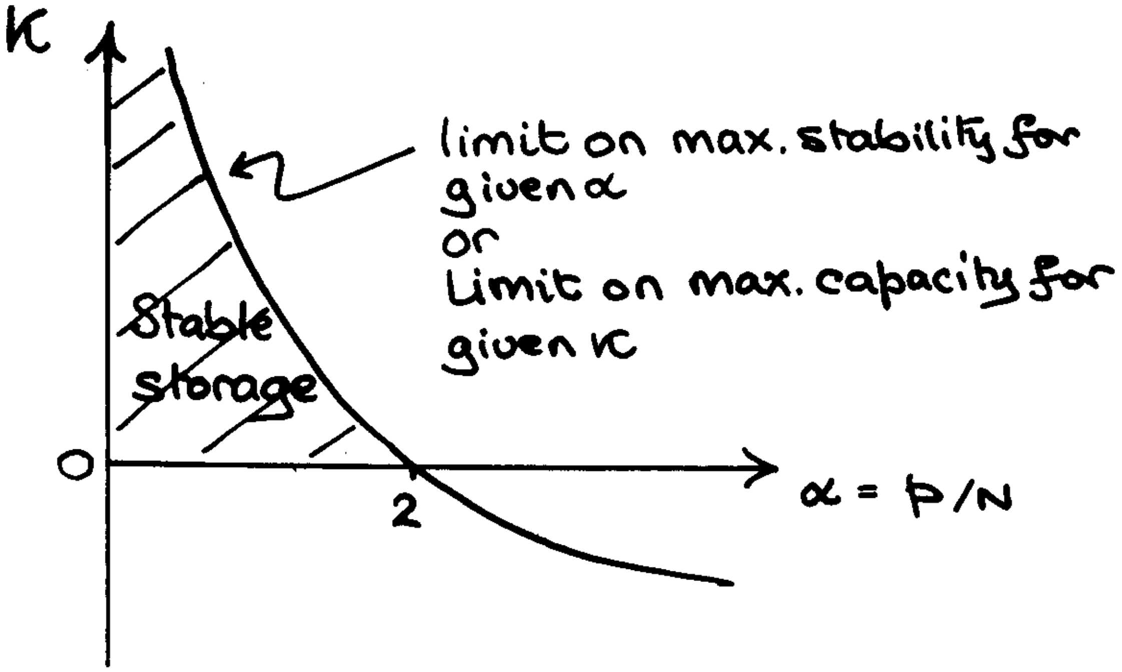
Training a perceptron for maximum stability

Storing random patterns ; (ξ_j^μ, η^μ) .

Cost example μ: $g(\Lambda^\mu) = +\Theta(\kappa - \Lambda^\mu)$: $\Lambda^\mu = \sum_j \eta^\mu J_j \xi_j^\mu$

Total cost : $H = \sum_{\mu=1}^{p} g(\Lambda^\mu)$; $p = \underset{\substack{\uparrow \\ \text{storage ratio}}}{\alpha} \underset{\substack{\uparrow \\ \text{no of input neuron:}}}{N}$

Stability phase diagram



$$\text{Also} \rightarrow \underset{\substack{\uparrow \\ \text{stability field dist}^n}}{\rho(\Lambda)} \rightarrow \underset{\substack{\uparrow \\ \text{retrieval map fn}}}{f(m)} \rightarrow$$ retrieval dynamics for low connectivity network.

$$\rho(\Lambda) = p^{-1} \sum_{\mu=1}^{p} \delta(\Lambda - \Lambda^\mu)$$

Minimization of $E_{\{\xi\}}(\{J\}) = -\sum_{\mu} g(\Lambda^{\mu})$

$; \; \Lambda^{\mu} = \xi_0^{\mu} \sum_{j=1}^{N} J_j \, \xi_j^{\mu}$

$; \quad \sum J_j^2 = c$

$\rightarrow \quad Z = \int \prod dJ \, \delta(\sum J^2 - c) \exp(\beta_A \sum_{\mu} g(\Lambda^{\mu}))$

↓

exponentiate

$\int d\varepsilon \exp(i\varepsilon(\sum J^2 - c))$

$\Lambda^{\mu} \to \lambda^{\mu} : \int d\lambda^{\mu} \; \delta(\lambda^{\mu} - \Lambda^{\mu})$

↓

exponentiate

<u>Replicate</u>

Average on patterns etc. $\langle \; \rangle_{\xi}$

$\int \prod dJ$

→ eventually

$\int dx^{\cdot\cdot} dq^{\cdot\cdot} d\varepsilon^{\cdot} \; \exp(-c(\text{fn of } x, q, \varepsilon))$

↑

extremize.

pins / neurons (driven by instantaneous interactions)

g Glauber dynamics

$$p(\sigma_i \to \sigma_i') = \frac{1}{2}\{1 + \sigma_i' \tanh \beta \tilde{h}_i\}$$

$$h_i = \sum_j J_{ij}\sigma_j + h_i^{ext.}$$

Interactions / synapses (driven by spin corr^{n}s)

$$\tau \frac{dJ_{ij}}{dt} = N^{-1}\Big\{ \underset{\substack{\uparrow \\ \text{average over} \\ \text{spin dynamics} \\ \text{(Hebb)}}}{\langle \sigma_i \sigma_j \rangle} + \underset{\substack{\uparrow \\ \text{ext. bias}}}{K_{ij}^{ext}} \Big\} - \underset{\substack{\uparrow \\ \text{Limiting term}}}{\mu J_{ij}} + N^{-1/2} \underset{\substack{\uparrow \\ \text{stochastic} \\ \text{noise}}}{\eta_{ij}(t)}$$

dist^n

$$\langle \eta_{ij}(t)\, \eta_{kl}(t') \rangle = 2\tilde{\beta}^{-1}\tau\, \delta_{ij,kl}\, \delta(t-t')$$

Note 2 'temperatures' ; $\underset{\substack{\uparrow \\ \text{spins}}}{\beta^{-1}}$, $\underset{\substack{\uparrow \\ \text{interactions}}}{\tilde{\beta}^{-1}}$

ssume : Spin dynamics much faster than interaction dynamics
(cf. Born-Oppenheimer , Adiabaticity)

→ spin dynamics → instantoneous Gibbs dist^n

$$p(\{\sigma\}) \sim \exp(-\beta H)$$

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_j h_j^{ext}\sigma_j$$

→ corr^n fn relevant to interaction dynamics

$$\langle\sigma_i\sigma_j\rangle = \beta^{-1}\frac{\partial}{\partial J_{ij}} \ln Z_\beta$$

$$Z_\beta = \operatorname*{Tr}_{\{\sigma\}} \exp(-\beta H)$$

∴ Synapse/interaction dynamics

$$\tau\frac{dJ_{ij}}{dt} = -\frac{\partial}{\partial J_{ij}}\left\{ \frac{-1}{\beta N}\ln Z_\beta - \frac{1}{N}\sum J_{ij}K_{ij}^{ext} + \frac{\mu}{2}\sum J_{ij}^2 \right\}$$

$$+ N^{-1/2}\eta_{ij}(t)$$

$$= -\frac{\partial}{\partial J_{ij}}\underset{\uparrow}{\tilde{H}} + N^{-1/2}\eta_{ij}(t)$$

Effective Hamiltonian

→ 'Langevin' dynamics

Long time : → eqm dist^n ~ $\exp(-\tilde{\beta}\tilde{H})$

$$\tilde{Z}_{\tilde\beta} = \int \prod_{i<j} dJ_{ij}\, [Z_\beta]^{\tilde\beta/\beta} \exp\Big\{\tilde\beta \sum J_{ij} K_{ij}^{ext}$$

$$- \frac{1}{2}\tilde\beta \mu N \sum_{i<j} J_{ij}^2 \Big\}$$

$$\sim \int \prod_{i<j} dJ_{ij}\, \underset{\uparrow}{P(J_{ij})} \quad Z_\beta^{\,n} \quad ; \quad n = \tilde\beta/\beta$$

$$\sim \exp\Big\{- \sum \frac{(J_{ij}-\tilde J_0)^2}{2\tilde J^2} \Big\}$$

cf. Replica theory for $\overline{Z^n}$

but now with physical meaning for n

$n \to 0$: $\beta \gg \tilde\beta$: interactions uncorrelated with spins → spin glass

$n \to \infty$: $\tilde\beta \gg \beta$: interactions slaved to spins → no frustration.

Analysis

- Replica theory analysis
- Spin glass like order parameters
    - Replica-symmetric ansatz
    - Parisi-like RSB

↑ If needed → complexity (many non-equiv thermo states. structure in $P(q)$)

- interpret

$$P(q) = \frac{1}{\tau^2}\int_0^\tau dt\,dt'\,\delta(q - N^{-1}\Sigma\sigma_i(t)\sigma_i(t'))$$

but actually calc from 'eqm' thermodyn.

$$\begin{cases} \bullet\ \delta\text{-fn} : \rightarrow \text{one thermodyn state} \\ \bullet\ \text{structure} : \rightarrow \text{many non-equiv states} \end{cases}$$

Results : phase diagram.

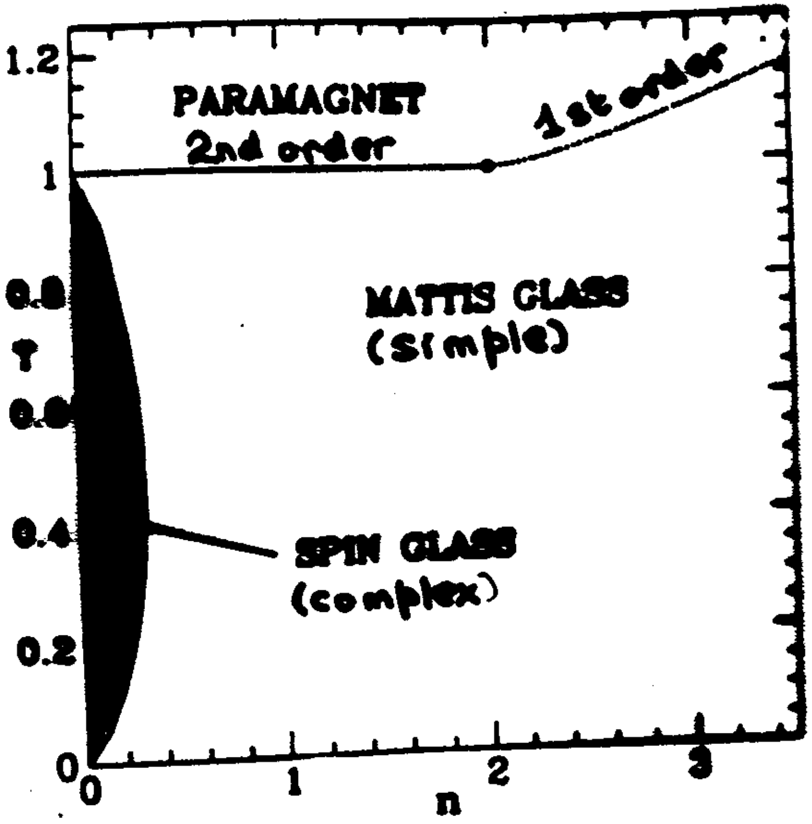


Figure 1: The phase diagram for the coupled dynamical model. Dotted (solid) lines denote first (higher) order transitions. The curve separating the Mattis and spin glass phases coincides with the de Almeida-Thouless instability.

- $n \to \infty$: $\tilde{\beta} \gg \beta$

  synapses slaved to spins
  (no frustration)

- $n \to 0$; $\tilde{\beta} \ll \beta$

  synapses random, uncorr. with spins
  (strong frustration)

Predictions for P(q) in complex region

- Not borne out by preliminary sim^ns of microdynamics

- → Casts doubt on assumption that

  τ_{trial} (neurons) $\ll$ τ_{trial} (synapses)

  implies

  τ_{equil} (neurons) $\ll$ τ_{trial} (synap[illegible])

- Rather, as complex region is approached perhaps, $\lim\limits_{N\to\infty} \tau_{eqm}$ (neurons) $\to \infty$

  → Locking on the 'edge of chaos' ?
  'edge of complexity' ?

- Complete dynamics still to be analyzed.

MACRODYNAMICS OF DISORDERED AND FRUSTRATED SYSTEMS

D. SHERRINGTON, A.C.C. COOLEN AND S.N. LAUGHTON
Theoretical Physics, University of Oxford, 1 Keble Road,
Oxford, OX1 3NP, England
E-mail: sherr@thphys.ox.ac.uk, coolen@thphys.ox.ac.uk, stephen@thphys.ox.ac.uk

ABSTRACT

It is shown how the macroscopic non-equilibrium dynamics of systems whose microscopic stochastic dynamics involves disordered and frustrated interactions can be well described by closed deterministic flow equations; this requires an appropriate choice of order parameter/function and ansätze.

1. Introduction

One would often wish to describe the macroscopic non-equilibrium dynamics of systems of many stochastically interacting microscopic units in terms of a closed set of deterministic flow equations. This is a non-trivial exercise for systems which are highly disordered and frustrated, but below we describe a procedure which works very well for two such problems, the Sherrington-Kirkpatrick spin glass[1] and the Hopfield neural network[2], and, we believe, provides a framework for more general application.

It is based on an appropriate choice of macroscopic order parameters together with two ansätze. Here we outline the philosophy and the essentials of the technique and present a few results to illustrate the degree of success. Further details can be found elsewhere[3,4,5].

2. The problem

In general terms, we are concerned with systems whose microscopic state is described by a set of N variables $\boldsymbol{S}_i$ at 'sites' $i = 1...N$ and which obey random stochastic microdynamics leading to known master equations for the time-dependence of the microstate probability distributions $p_t(\{\boldsymbol{S}\})$, involving only the instantaneous time t but with disorder and frustration in the local and intersite controlling elements. Our objective is to devise a description of the macrostate dynamics in terms of closed equations for few-parameter sets of macrovariables $\Omega_\mu(\{\boldsymbol{S}\}); \mu = 1...n$.

More specifically, we concentrate on systems in which the variables $\boldsymbol{S}$ are Ising spins $\{\sigma = \pm 1\}$ and obey random sequential Glauber stochastic dynamics via local effective fields determined through pairwise exchange interactions with other spins and external stimuli. The evolution of the microstate distribution satisfies the master equation

$$\frac{d}{dt}p_t(\boldsymbol{\sigma}) = \sum_{k=1}^{N}[p_t(F_k\boldsymbol{\sigma})W_k(F_k\boldsymbol{\sigma}) - p_t(\boldsymbol{\sigma})W_k(\boldsymbol{\sigma})] \tag{1}$$

where F_k is the spin-flip operator $F_k\Phi(\boldsymbol{\sigma}) = \Phi(\sigma_1, ..., -\sigma_k, ..., \sigma_N)$, the transition rates and local fields are

$$W_k(\boldsymbol{\sigma}) = \frac{1}{2}[1 - \sigma_k \tanh(\beta h_k(\boldsymbol{\sigma}))] \qquad h_k(\boldsymbol{\sigma}) = \sum_{\ell \neq k} J_{k\ell}\sigma_\ell + \theta_k, \tag{2}$$

β is the inverse temperature and we are now using the vector notation $\boldsymbol{\sigma} = (\sigma_1, ...\sigma_N)$. From (1) we may derive an equation for the evolution of the macrovariable probability distribution

$$P_t[\boldsymbol{\Omega}] = \sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]; \quad \boldsymbol{\Omega} \equiv (\Omega_1, ...\Omega_n) \tag{3}$$

in the form

$$\frac{d}{dt}P_t[\boldsymbol{\Omega}] = \sum_{\ell \geq 1} \frac{(-1)^\ell}{\ell!} \sum_{k_1=1}^{n} .. \sum_{k_\ell=1}^{n} \frac{\partial^\ell}{\partial\Omega_{k_1}...\partial\Omega_{k_\ell}} \{P_t[\boldsymbol{\Omega}]F^{(\ell)}_{k_1..k_\ell}[\boldsymbol{\Omega};t]\} \tag{4}$$

where

$$F^{(\ell)}_{k_1...k_\ell}[\boldsymbol{\Omega};t] = \langle \sum_{j=1}^{N} W_j(\boldsymbol{\sigma})\Delta_{jk_1}(\boldsymbol{\sigma})...\Delta_{jk_\ell}(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t} \quad \Delta_{jk}(\boldsymbol{\sigma}) \equiv \Omega_k(F_j\boldsymbol{\sigma}) - \Omega_k(\boldsymbol{\sigma}) \tag{5}$$

and the notation $\langle\rangle_{\boldsymbol{\Omega};t}$ refers to a sub-shell average

$$\langle f(\boldsymbol{\sigma})\rangle_{\boldsymbol{\Omega};t} \equiv \frac{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]f(\boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta[\boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\sigma})]} \quad . \tag{6}$$

In several cases of interest and for finite times only the first term on the right hand side of (4) survives in the limit $N \to \infty$, yielding the deterministic flow

$$\frac{d}{dt}\boldsymbol{\Omega}_t = \langle \sum_i W_i(\boldsymbol{\sigma})[\boldsymbol{\Omega}(F_i\boldsymbol{\sigma}) - \boldsymbol{\Omega}(\boldsymbol{\sigma})]\rangle_{\boldsymbol{\Omega};t} \quad . \tag{7}$$

In general this does not yet constitute a closed set of equations due to the appearance of $p_t(\boldsymbol{\sigma})$ in the sub-shell average. This requires an appropriate choice of $\boldsymbol{\Omega}$ and possibly further ansätze.

3. The specific physical systems

Here we concentrate on the two specific model systems, the Sherrington-Kirkpatrick (SK) spin glass[1] and the Hopfield neural network[2]. In the SK model the $\{\sigma\}$ represent true magnetic spins and the $\{J_{ij}\}$ are chosen randomly from a Gaussian distribution

$$J_{ij} = J_o/N + Jz_{ij}/\sqrt{N} \qquad z_{ij} = z_{ji} \quad \langle z_{ij}\rangle = 0 \quad \langle {z_{ij}}^2\rangle = 1 \tag{8}$$

In the Hopfield model the $\{\sigma\}$ represent states of McCulloch-Pitts neurons, $\sigma = \pm 1$ corresponding to firing/non-firing, and the $\{J_{ij}\}$ provide for the storage and retrieval of random patterns $\{\xi_i^\mu = \pm 1\}$; $\mu = 1...p = \alpha N$, via the Hebb rule $J_{ij} = N^{-1} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$. Concentrating for simplicity on the region of phase space within the basin of attraction of one pattern, $\mu = 1$, it is convenient to apply the gauge transformation $\sigma_i \rightarrow \sigma_i \xi_i^1$, $J_{ij} \rightarrow \xi_i^1 \xi_j^1 J_{ij}$, $\theta_i \rightarrow \xi_i^1 \theta_i$ to re-write this in the form of (8) with

$$z_{ij} = \frac{1}{\sqrt{p}} \sum_{\mu > 1}^{p} \xi_i^1 \xi_i^\mu \xi_j^1 \xi_j^\mu \quad J_o = 1 \quad J = \sqrt{\alpha} \quad \langle \xi_i \rangle = 0 \ . \tag{9}$$

The symmetry $z_{ij} = z_{ji}$ provides for the steady state distribution $p_\infty(\boldsymbol{\sigma})$ to be expressible in the Boltzmann form $p_\infty(\sigma) \sim \exp(-\beta H)$ with the Hamiltonian

$$H = -\sum_{i<j} J_{ij} \sigma_i \sigma_j - \sum_i \theta_i \sigma_i \tag{10}$$

Further specializing to $\theta_i = \theta$, all i, the Hamiltonian is expressible in terms of the two macroscopic parameters

$$m(\boldsymbol{\sigma}) = N^{-1} \sum_i \sigma_i \qquad r(\boldsymbol{\sigma}) = N^{-3/2} \sum_{i<j} \sigma_i z_{ij} \sigma_j \tag{11}$$

$$H(\boldsymbol{\sigma})/N = [-\frac{1}{2} J_o m^2(\boldsymbol{\sigma}) + \theta m(\boldsymbol{\sigma})] - J r(\boldsymbol{\sigma}) + 0(N^{-1}) \tag{12}$$

where the term in [ ] is disorder-independent with all the disorder effects in the $Jr(\boldsymbol{\sigma})$ term. Hence m, r suffice as order parameters to describe the equilibrium probability distribution and represent a minimum choice for the set $\boldsymbol{\Omega}$ consistent with needing m to study magnetization or overlap with the condensed pattern.

Thus we shall first discuss an attempt to find a non-equilibrium macrodynamics in terms of m, r, alone, and show that it provides a reasonable but imperfect description. We shall then go on to a more sophisticated theory in terms of a generalized order function which provides a very good fit to the results of microscopic simulation.

4. The simple version of the theory

In this section we choose the minimal form

$$\boldsymbol{\Omega}^s(\boldsymbol{\sigma}) \equiv (\Omega_1(\boldsymbol{\sigma}), \Omega_2(\boldsymbol{\sigma})) = (m(\boldsymbol{\sigma}), r(\boldsymbol{\sigma})) \tag{13}$$

The resultant $P_t[\boldsymbol{\Omega}^s]$ does indeed satisfy a Liouville equation in the thermodynamic limit, yielding the deterministic flow equations

$$\frac{dm}{dt} = \int dz D_{m,r;t}(z) \tanh \beta(J_o m + J z + \theta) - m \tag{14}$$

$$\frac{dr}{dt} = \int dz D_{m,r;t}(z) z \tanh \beta(J_o m + Jz + \theta) - 2r \tag{15}$$

where $D_{m,r;t}(z)$ is the sub-shell averaged distribution of the disorder contributions to the local fields

$$D_{m,r;t}(z) = \lim_{N\to\infty} \frac{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta(m - m(\boldsymbol{\sigma}))\delta(r - r(\boldsymbol{\sigma}))N^{-1}\sum_i \delta(z - z_i(\boldsymbol{\sigma}))}{\sum_{\boldsymbol{\sigma}} p_t(\boldsymbol{\sigma})\delta(m - m(\boldsymbol{\sigma}))\delta(r - r(\boldsymbol{\sigma}))} \tag{16}$$

$$h_i(\boldsymbol{\sigma}) = J_o m(\boldsymbol{\sigma}) + J z_i(\boldsymbol{\sigma}) + \theta + 0(N^{-1}) \qquad z_i(\boldsymbol{\sigma}) = N^{-1/2}\sum_j z_{ij}\sigma_j \tag{17}$$

As yet, because of the $p_t(\boldsymbol{\sigma})$ in (16), equations (14), (15) are not closed except in the disorder-free case $J = 0$. To close the equations we introduce two simple ansätze: (i) we assume that the evolution of the macrostate (m, r) is self-averaging with respect to the specific microscopic realization of the disorder $\{z_{ij}\}$, (ii) as far as evaluating $D(z)$ is concerned we assume equipartitioning of the microstate probability $p_t(\boldsymbol{\sigma})$ within each (m, r) shell. The first of these ansätze is well borne out by computer simulations of the microscopic dynamics and permits averaging over pattern choices. The second, which can only be judged *a posteriori*, eliminates memory effects beyond their reflection in m, r and removes explicit time-dependence from D. Together they give

$$D_{m,r;t}(z) \rightarrow D_{m,r}(z) = \left\langle \frac{\sum_{\boldsymbol{\sigma}} \delta(m - m(\boldsymbol{\sigma}))\delta(r - r(\boldsymbol{\sigma}))N^{-1}\sum_i \delta(z - z_i(\boldsymbol{\sigma}))}{\sum_{\boldsymbol{\sigma}} \delta(m - m(\boldsymbol{\sigma}))\delta(r - r(\boldsymbol{\sigma}))} \right\rangle_{\{z_{ij}\}} \tag{18}$$

whose insertion into (14) and (15) yields the required closure.

The actual evaluation of $D_{m,r}(z)$ from (18) remains a non-trivial exercise, but one which is amenable to solution by replica theory as developed for the investigation of local field distributions in spin glasses[6]. After several manipulations D can be expressed in the form

$$D_{m,r}(z) = \lim_{n\to 0} \int \prod_{i,j} \prod_{\alpha,\beta=1\dots n} dx_i^\alpha dy_j^{\alpha\beta} \exp[-N\Phi(m, r, z; \{x_i^\alpha\}, \{y_j^{\alpha\beta}\})] \tag{19}$$

where the number of indices i, j is finite and Φ is $O(N^0)$. Because of the exponential scaling as N, the integral can be evaluated by steepest descents.

The extremization of (19) is discussed in detail elsewhere[4] but we note that it involves a parameter $q^{\alpha\beta}$ which is a dynamical analogue of the usual (static) spin-glass order parameter and has a similar, but now dynamically constrained, interpretation in terms of overlaps; specifically, the disorder-averaged probability distribution for the mutual overlap between microscopic configurations confined to the same macroscopic (m, r) sub-shell is given by

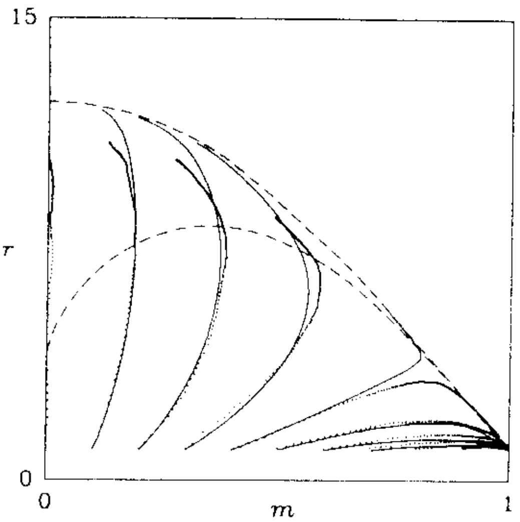


Fig. 1. Macroscopic flow trajectories for a Hopfield model with storage capacity $\alpha = 0.1$ and deterministic microscopic dynamics ($\beta = \infty$); dots indicate simulations ($N = 32000$), solid lines indicate analytic RS theory. The outer dashed line is the boundary predicted by RS theory; the inner dashed line indicates the onset of instability against RS-breaking fluctuations.

$$
\begin{aligned}
P_{m,r}(q) \;\; &\equiv \\
&\left\langle \frac{\sum_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}\delta(q - N^{-1}\sum_i \sigma_i\sigma_i')\delta(m - m(\boldsymbol{\sigma}))\delta(m - m(\boldsymbol{\sigma}'))\delta(r - r(\boldsymbol{\sigma}))\delta(r - r(\boldsymbol{\sigma}'))}{\sum_{\boldsymbol{\sigma},\boldsymbol{\sigma}'}\delta(m - m(\boldsymbol{\sigma}))\delta(m - m(\boldsymbol{\sigma}'))\delta(r - r(\boldsymbol{\sigma}))\delta(r - r(\boldsymbol{\sigma}'))} \right\rangle_{\{z_{ij}\}} \\
&= \;\; \lim_{n\to 0} \frac{1}{n(n-1)} \sum_{\alpha\neq\beta} \delta(q - q^{\alpha\beta})
\end{aligned}
\tag{20}
$$

The steady state condition $dm/dt = dr/dt = 0$ is satisfied by $m, r, \{q^{\alpha\beta}\}$ which obey the usual self-consistency equations as obtained from equilibrium analysis[1,7,8,9], including all replica-symmetry breaking aspects.

For general m, r the explicit extremization and limiting procedure for (19) is greatly simplified within the replica-symmetric (RS) ansatz, which already yields a non-Gaussian form for $D_{m,r}(z)$ in qualitative accord with the results of simulation[3,4]. Substituting into (14) and (15) there result closed equations of the form

$$
dm/dt = F(m,r) \qquad dr/dt = G(m,r) \tag{21}
$$

where F, G are complicated functions taking the form of integrals over other functions involving parameters self-consistently determined from m, r via non-linear equations. Figs 1 and 2 show examples of their predictions for the Hopfield model compared with microscopic simulations[10]. These figures show several points of note; first,

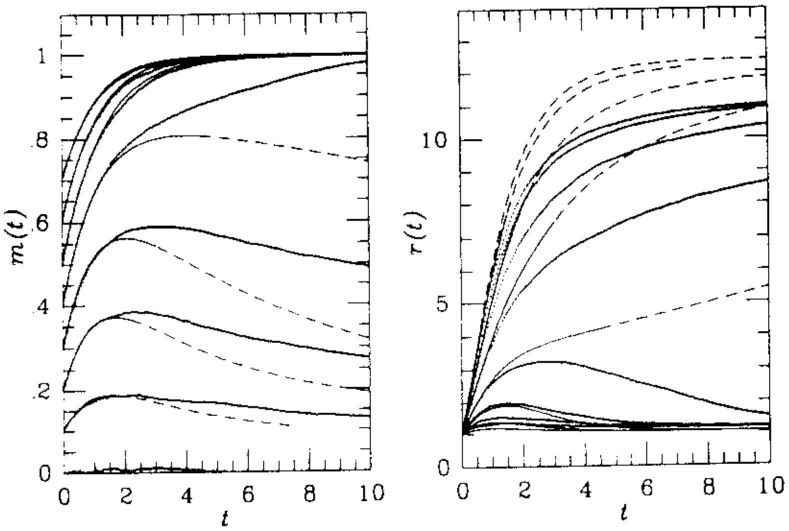


Fig. 2. Temporal dependence of the order parameters for a Hopfield model with storage $\alpha = 0.1$ and zero-temperature dynamics ($\beta = \infty$); dots indicate simulations ($N = 32000$), the other lines indicate RS theory shown with solid lines where stable, dashed lines where unstable. Time is measured in Monte Carlo steps per spin.

they demonstrate the concept of basins of attraction – there is a critical locus of $(m(0), r(0))$ which separates flows which retrieve ($m(\infty) \sim \mathcal{O}(1)$) from those which do not ($m(\infty) \to 0$); second, the critical $m_c(0)$ depends on $r(0)$; third, the present version of the theory and ansätze yield qualitatively reasonable results for r versus m along flow lines away from the critical locus, but miss the slowing-down of $m(t), r(t)$ seen in simulations of non-retrieving flows. Fig. 1 also shows (dashed) two other special loci; the upper one is a theoretical limit of the boundary of the physical region within the RS ansatz and corresponds to the limit of maximum $q = 1$; the lower dashed locus is the upper boundary of the limit of stability of the RS ansatz for calculations of D, a dynamical analogue of the de Almeida-Thouless[11] (AT) line.

Although the crossing of the AT line could be expected to herald slowing-down due to replica-symmetry breaking, the figures show this occurs earlier, suggesting a different origin for the discrepancy between theory and simulation. Results of a similar quality are obtained for finite temperature and for the SK spin glass[4]. A study of a related RS toy model[12] further emphasises the qualitative usefulness but quantitative incompleteness of the simple theory.

5. The sophisticated version of the theory: order function dynamics

To improve on the theory requires broadening the range of order parameters. For a qualitative improvement we anticipate the need for a qualitative change in that range and so, instead of two order parameters, we consider a continuous order function.

Since the microscopic dynamics is formulated entirely in terms of the states of the

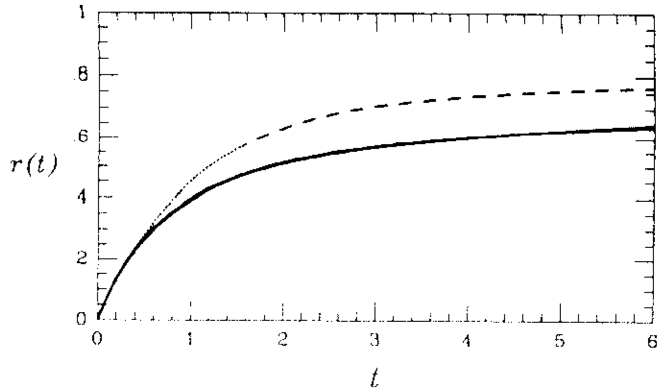


Fig. 3. Evolution of the binding energy of the Sherrington-Kirkpatrick spin glass ($J_o = 0$) from a random microscopic start. Comparison of simulations ($N = 8000$, solid line) and predictions of the simple two-parameter (m, r) theory of section 4 (RS stable, dotted; RS unstable, dashed) and of the advanced order-function theory of section 5 (solid), for $\beta = \infty$. Note that the two solid lines are almost coincident.

spins σ_i and the fields h_i we choose for $\Omega(\boldsymbol{\sigma})$ the joint distribution

$$\mathcal{D}[\zeta, h; \boldsymbol{\sigma}] = N^{-1} \sum_i \delta_{\zeta \sigma_i} \delta(h - h_i(\boldsymbol{\sigma})) \tag{22}$$

This choice automatically includes the positive features of the simple version, since m, r follow straightforwardly from $\mathcal{D}$ and hence a formulation in terms of $\mathcal{D}$ will automatically be correct asymptotically for systems whose Hamiltonians can be expressed completely in terms of m, r. It has further advantages in being applicable also to systems without detailed balance (and therefore not expressible by Boltzmann statistical equilibrium) and of being extendable to analogue spins.

As discussed in §2 we consider the evolution of the probability distribution for the order function $P_t[\mathcal{D}]$. Provided we first discretize the h-distribution at a finite number of values, only passing later to the continuum limit, we do indeed find a Liouville form corresponding to deterministic evolution of $\mathcal{D}_t[\zeta, h]$ in the thermodynamic limit. The equation of motion for $\mathcal{D}_t[\zeta, h]$ involves a further noise distribution over a sub-shell with $\mathcal{D}[\zeta, h; \boldsymbol{\sigma}] = D_t[\zeta, h]$, again weighted by $p_t(\boldsymbol{\sigma})$. Thus once more we make an equipartitioning ansatz to eliminate $p_t(\boldsymbol{\sigma})$, as well as self-averaging, but now the relevant sub-shell is much more restricted and contains much more memory information. This provides a closed equation for the evolution of the order function $\mathcal{D}_t[\zeta, h]$. To evaluate the 'noise-term' again requires replica analysis and even within replica-symmetric theory yields a complicated self-consistent set of equations to determine the flow. The results, however, are impressive, as fig. 3 illustrates for the flow of the binding energy of the SK model at $T = 0^{5,13}$; while there is clear difference between the predictions of the simple theory and the slower evolution obtained in computer simulations, within the numerical accuracy shown $\mathcal{D}_t[\zeta, h]$ captures the slowing-down well, although clearly it cannot reproduce perfectly asymptotically the equilibrium

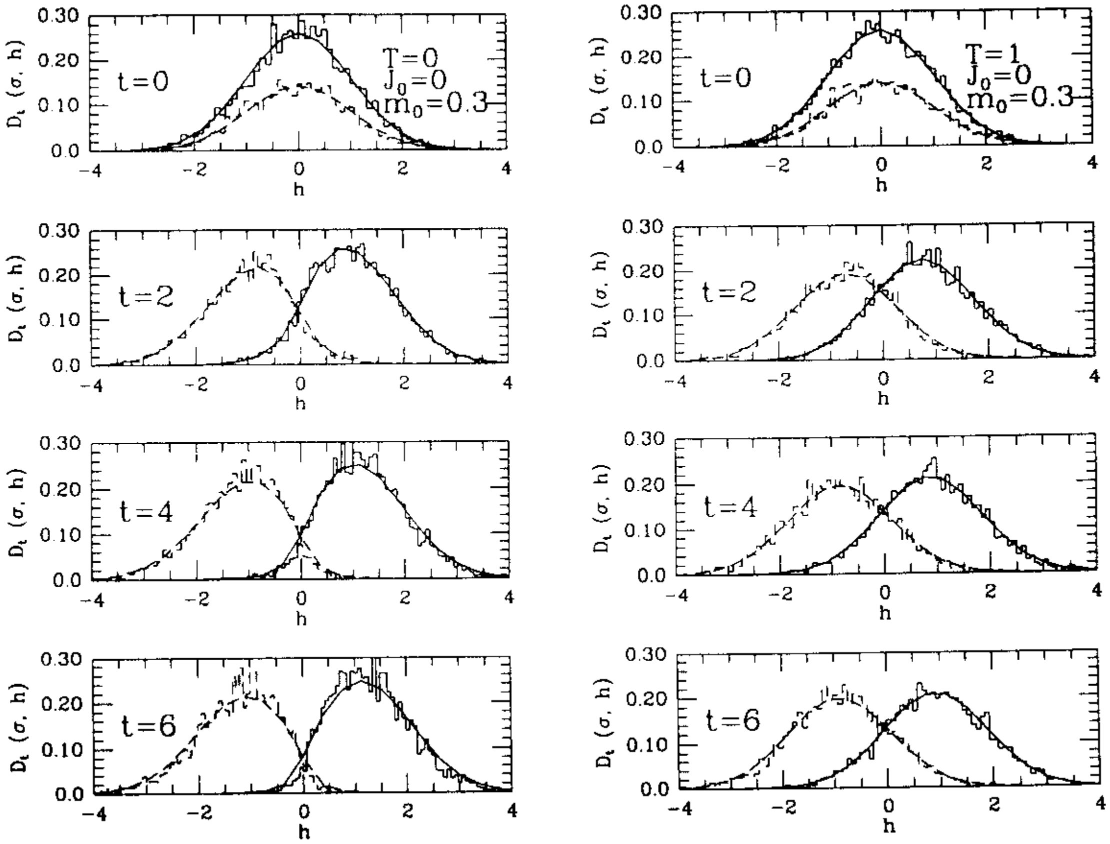


Fig. 4. Evolution of $\mathcal{D}_t(\sigma, h)$ for SK models with $J_o = 0$, $J = 1$, $T = 0, 1$. The lines are predictions of the sophisticated theory within RS ansatz, the histograms are obtained from microscopic simulations ($N = 8000$).

ensemble result which is replica-symmetry broken. To be more quantitative we note that the true RSB equilibrium binding energy at $T = 0$ is 0.7633 while RS theory gives 0.798, a difference which is small compared with that between the simple theory and the sophisticated theory/simulation over the time range shown (except for the initial times; both theories are exact at $t = 0$). On the other hand, it should be noted that it is also possible that the dynamical equations yield an additional asymptotic steady state solution different from that of Gibbsian ensemble theory, but we have not yet investigated this question.

Fig. 4 shows an example of the evolution of $D_t[\zeta, h]$ itself [5,13] for $J_o = 0, T = 0, 1$. Note that for $T = 0$ $D_t[\xi, h]$ tends to zero for $(\zeta h) < 0$ as $t \to \infty$, where as for $T > 0$ it remains non-zero over both signs of (ζh) for all times.

6. Conclusion

We have demonstrated that with an appropriate choice of order parameters and two simple ansätze one can derive closed macroscopic flow equations for range-free disordered and frustrated systems in good accord with the results of microscopic simulations. In its more sophisticated version, in terms of an order function describing the distribution of spins and fields, and including RSB effects the theory may even be exact. Discussion has been restricted to Ising spins and detailed balanced dynamics, but the sophisticated version of the theory, can be extended beyond these restrictions. In the case of neural networks both versions can also be extended to regions of phase space having finite overlaps with more than one pattern[13].

7. Acknowledgement

We would like to thank the Engineering and Physical Sciences Research Council of the United Kingdom for partial support under grant GRH 73028 (ACCC) and scholarship 923.0910 (SNL).

8. References

1. D. Sherrington and S. Kirkpatrick, *Phys. Rev. Lett.* **35** (1975) 1792
2. J.J. Hopfield, *Proc. Nat. Acad. Sci. USA* **79** (1982) 2554
3. A.C.C. Coolen and D. Sherrington, *Phys. Rev.* **E49** (1994) 1921
4. A.C.C. Coolen and D. Sherrington, *J. Phys* **A27** (1994) 7687
5. S.N. Laughton, A.C.C. Coolen and D. Sherrington, in preparation
6. M. Thomsen, M.F. Thorpe, T.C. Choy, D. Sherrington and H-J. Sommers, *Phys. Rev* **B33** (1986) 1931
7. D.J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. Lett.* **55** (1985) 1530
8. G. Parisi, *J. Phys* **A13** (1980) 1101
9. M. Mézard, G. Parisi and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore 1987).
10. D. Sherrington and A.C.C. Coolen, to be published in *Proceedings of 1994 Sitges Conference on '25 years of Non-Equilibrium Statistical Mechanics'* (Springer-Verlag, Berlin)
11. J.R.L. de Almeida and D.J. Thouless, *J. Phys* **A11** (1978) 983
12. A.C.C. Coolen and S. Franz, *J. Phys* **A27** (1994) 6947
13. S.N. Laughton, D. Phil Thesis, Oxford 1995

Landscape Paradigms in Physics and Biology: Introduction and Overview

David Sherrington

Dept. of Physics, Oxford University, Theoretical Physics, 1 Keble Rd., Oxford OX1 3NP, England

Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

July 8, 1996

Abstract

A brief introductory overview in general terms is given of concepts, issues and applications of the paradigm of rugged landscapes in the contexts of physics and biology.

In the present context, landscapes describe the structure of control functions relevant to the cooperative behaviour of systems of many interacting units. The paradigm is now ubiquitous in several branches of science, particularly for the conceptualization of behaviour which is commonly described as complex.

To biologists the landscape is typically visualized as giving a measure of fitness, to be maximized, while to physicists it is usually considered as specifying an energy, to be minimized. However, these are simply inverted representations of the of the same thing and henceforth I shall tend to use the physicists' language.

As in geography, these landscapes come in various forms, flat, smooth, discontinuous and rugged. Again as in human perception of the world around us, flat landscapes are the simplest to contemplate but rugged landscapes excite the greatest interest, in the sense of having the richest, most complex consequences. This perception of interest in ruggedness is reflected in the papers that follow, but even systems with flat landscapes can yield highly non-trivial behaviour when interactions between their inhabitants are sufficiently complicated.

What is the space of these landscapes? For some problems it is low-dimensional, as,for example, when the landscape represents the potential energy seen by an electron due to interactions with atoms, ions or other fixed objects; in this case the 'grid coordinates' (to use an analogy with a cartographic map) indicate the location of the electron and the 'elevation' measures the potential at that point [1]. A similar situation at a more coarse-grained level can apply to measures of coefficients in a Ginzburg-Landau free energy functional expansion in statistical physics. Here, however, we shall be thinking of a different situation where the landscape sits in some high-dimensional space in which each 'grid-point' specifies either a complete microstate, which describes the 'positions' of *all* the individual units which make up the many-body system, or some more coarse-grained, but still multi-dimensional, macrostate characterization. In most cases we shall think in terms of a single 'height' parameter as a function of a multi-dimensional 'location' parameter, but it is perfectly possible to have a several-dimensional height measure. Neither the 'horizontal' nor the 'vertical' coordinates of the landscape need be continuous, but, since the conventional world whose experience has molded our normal conceptualization does have this feature, for orientation I shall often use images based on such a continuous picture.

At the simplest level one might think of the dynamical behaviour of the many-body system in terms of motions on this landscape; in particular, for deterministic dynamics, in terms of gradient descent to local minima. In a non-flat landscape this leads immediately to an image of separated regions of flow and their associated attractors, corresponding to the valleys, with barriers between them, corresponding to the hills and saddles.

For many of the problems of interest the effective landscape structure is rugged in the sense that within a single closed contour of some 'height' one finds many closed contours of lower height and, further, within each of these one finds many closed contours of even lower height and so on. Equivalently, there is a hierarchy of several sub-valleys within valleys at many scales. The consequence is that deterministic microdynamics can yield many possible final stopping states, hierarchically related and often with quasidegeneracies in their 'heights'. Even for stochastic dynamics, in which uphill moves are also allowed with a probability decreasing with the height change involved, in large enough systems ruggedness of the microscopic landscape leads to the possibility of effective non-ergodicity in which non-equivalent macrostates result depending on the starting microstates, not communicating on realistic timescales. This is sometimes re-expressible as downhill moves on a still-rugged *free energy* surface in macrospace.

Let us turn now to the origin of rugged landscapes. They can arise due to competition between different microscopic few-body interactions: for example, in a magnetic context between ferromagnetic and antiferromagnetic exchange interactions; in a neural network between excitatory and inhibitory synapses or neurons. Or they can be due to conflicts between few body-forces and global constraints, as in the cost functions of graph equipartioning. They can also have their origin in competition between internal and external forces, as in the random field Ising model. Generically we refer to these conflicts as *frustration.*

On a fine scale the contours of a landscape can vary slowly and continuously or they can involve a series of quasi-steps at which the height changes rather rapidly, separating regions of slower but still hierarchical evolution; in the first case one would expect a continuous hierarchy of metastable states as in some spin-glass models [2] , while the second would suggest 'tiers' of 'conformational substates' as suggested in some studies of proteins (like myoglobin) [3].

The landscapes themselves are not immutable but can change with changes in interactions or external perturbations. Such changes can be (quasi-)continuous on some longer timescale, such as in long time potentiation (or synaptic modification) in neural network learning, or they can be sudden due to a fast perturbation, such as occurs in photoexcitation in proteins. In principle, some of these modifications can be considered within a larger space of dynamics allowing for simultaneous evolution of both the landscape and the elements it controls, typically with different timescales. Changes can be smooth or chaotic in their response to changes in global control parameters.

Another distinction we should make is between random and quasi-random ruggedness. In some systems, the landscape is controlled by truly random quantities; an example is a spin glass model in which the the exchange interactions are randomly chosen and are thought to give rise to an energy landscape with very many attractors, a high quasi-degeneracy of the energies of these attractors and slow long-time dynamics to approach them. In other systems, however, the landscape is sculpted via appropriate changes in the controlling few-body interactions. One example is in a neural network trained via a supervised learning procedure to yield desired (memory) attractors. More instances of rugged but less truly random landscapes arise in several biological, economic and ecological contexts tuned for success; for example, a biologically relevant (and realistic) protein must have dynamics leading quickly to a folded state with appropriate structure and function, suggesting that it should have one large and dominant attractor, possibly with several quasi-degenerate 'ground states' with similar functions but with large energy separations from higher states with different functions [5, 4]; while in much of nature the successful agents are those which have evolved to perform their tasks efficiently and robustly. Henceforth I shall refer to this second group as 'sculpted landscapes' [6].

Thus far we have considered only the nature of the landscapes. However, the allowed 'steps' are also very important and the attractor structure and dynamics can be different in different 'step-spaces'. Furthermore, the whole concept of motion via descents on a landscape is itself only a special case of a more general dynamic flow space for which it is not necessary to have detailed balance or a description in terms of a quantity which is always minimized in each microscopic move (a Lyapunov function); as examples of such systems one can quote some neural networks and most cellular automaton models. More fundamental is the structure of the space of dynamical flows, which is often hierarchically fractured (and therefore warrants a description as 'rugged') without having a true landscape description; it is clear that if one is to have non-fixed point attractors, such as sequences or restricted strange attractors, the naive landscape description needs such extension. We shall take the landscape paradigm to be generalized in this sense. However, for convenience we shall continue to use the language of the naive landscape paradigm in general discussion.

As noted earlier, motion on the landscapes need not not be deterministic, but can also be stochastic; the moves need not be simply locally downhill in the space of energy landscapes over the full microspace, but one can have 'probabalistic hill-climbing'; alternatively, in the description employing free energies in macrospace one has a modification of the landscape itself, typically a smoothing with increased temperature, but also with possibilities of entropically driven new attractors. A similar situation applies to the modification of flows, attractors and their basins even without detailed balance. These changes can lead to phase transitions as a function of stochasticity as well as those due to changes in global control parameters. This is also an appropriate point to emphasise the difference between thermodynamic and attractor phase diagrams; the former are concerned with systems with detailed balance dynamics, governed by the laws of equilibrium statistical mechanics and with states weighted by Bolzmann or relevant quantum statistical factors, while the latter is concerned with the occurence of dynamical attractors, even in systems without Lyapunov functions and where, even if the concept is meaningful, their energies may be so high as to exclude them from thermodynamic relevance.

There are clearly many different systems which can be considered in the terms discussed so far. Some require quite different mathematical formulation but other physically quite distinct systems can be described mathematically in very similar fashions. As an illustration let us consider a set I have referred to previously [7] under the grouping "Magnets, microchips and memories", in which the magnets are Ising spin glasses, the microchips refer to the problem of equally bipartitioning the elements of an electrical circuit between two microchips so as to minimize the number of wires between the chips, and the memories refer to recurrent neural networks of McCulloch-Pitts neurons. All these cases can be described by an energy function $H = -\sum_{(ij)} J_{ij}\sigma_i\sigma_j$ where the subscripts label the microscopic units (spins, circuit elements, neurons); respectively for the magnets, microchips and memories, $\sigma = +/-1$ refers to whether the spins are up/down, the circuit elements are on the first/second microchip, or the neurons are firing/nonfiring, while the J_{ij} measure exchange interactions between spins, the wirelength needed if two circuit elements are on different microchips, or the strength and character of the synaptic influence of one neuron on another. For the magnets and memories the J's are a mixture of positive and negative signs, while for the microchips the J's are all positive but there is a global constraint that $\sum_i \sigma_i = 0$. All these situations are frustrated, yielding rugged landscapes. The J's are typically random for the magnets, only quasi-random for the neural networks which are sculpted to yield attractors corresponding to memorized global patterns, and quasi-random, but not sculpted, for the microchips if the circuit connections are simply designed to yield an appropriate electronic operation rather than to optimize the placement and wiring problem discussed here. Furthermore, both the magnets and the memories can be considered also within a related context in which only a dynamic description is given, that the probability of updating σ_i is determined only by the instantaneous value of a 'field' $h_i = \sum_j J_{ij}\sigma_j$. In the case of the neurons there is no need for J_{ij} to be equal to J_{ji} and hence there need be no 'energy' landscape. For the microchip problem there is no *a priori* dynamics; the objective is to find a dynamics which minimizes the cost and use it to find the corresponding microstate.

As noted earlier, the dynamical behaviour of random and sculpted landscapes are typically quite different. For random systems the long time dynamics is usually slow, although there may be faster initial transients, whereas for survival in the world the kinetics of achieving a desired state are as important as the latter's structure and require an appropriate tuning of the landscape; for example, prey must respond quickly to the presence of predators, a protein must fold rapidly, and a neural network must associate or generalize quickly.

Usually one does not require knowledge of the full micro-description, but rather one wants macroscopic measures to monitor performance. Hence it is important to consider the passage from the full microdynamics to a consequential macrodynamics in terms of a few macro-observables. These macroparameters can be a set of instantaneous measures or they can involve multi-time correlation and response functions. Whatever the specific case, one is faced with the question of how many such macrovariables are needed for an adequate description; typically one cannot express the dynamics of disordered and frustrated systems adequately in terms of a single instantaneous macroparameter. Studies of the macrodynamics expose such features as (slow) glassy macrodynamics and aging in systems with very rugged landscapes.

Also relevant at the level of macrodynamics is the question of self-averaging, or its absence. Self-averaging refers to the feature of independence of macro-observables from the specific instances of microscopic randomness, with only the distributions from which the random elements are drawn being relevant in the limit of large systems. Dependent on the system, some macrovariables are self-averaging while others are not. For example, in infinite-ranged spin glasses the energy and the magnetization are self-averaging with respect to the specific choice of random exchange interactions, but the overlaps [8] between two identical but separately evolving replicas are not. There are also typically sample-to-sample fluctuations among the reaction rates of folded proteins with the same molecular sequences. For this reason care is required in specifying different types of averages. A related concept is that of ultrametricity [9, 2] of such non-selfaveraging quantities, itself an indication of the hierarchical organization of the determining landscape.

The evidence for the images presented above comes from a combination of 'real' experiments, computer simulation experiments and the analysis of theoretical models. In the papers which follow several aspects of both evidence and consequences will be discussed in many different systems with many different investigative techniques. The fundamental issues can be boiled down to two questions; (i) at the specific system level, how do the microscopics lead to macroscopic structure and function and/or what do the observed structure and function tell about the microdynamics, (ii) at the global level, to what extent is there universality in the landscape paradigm in different areas of science and compementarily what are the nuances of differences? The simplicity of these questions hides the very considerable subtlety of their answers and of the further questions they raise. They have exercised many brains for several decades before yielding the conceptual images and the experimental, analytic and computational techniques which are now in place and which have greatly enriched our understanding of the complex world around us and our toolbox of ways to probe it, yet the knowledge we have gained so far is certainly only the tip of an iceberg whose true majesty will take many more years to be exposed.

Acknowledgements My own understanding of the landscape paradigm owes much to discussion with many colleagues and friends. They are too numerous to name here but they know who they are and I do thank them sincerely.

REFERENCES

[1] The physics of such systems is encapsulated in the title of Angus Mackinnon's inaugural lecture as a Professor at Imperial College, "Bens, glens, lochs, and burns", which translates from Scottish to English as "Mountains, valleys. lakes and streams", the first two nouns characterizing the landscape and the last two its consequences.

[2] See for example M.Mézard, G.Parisi and M.Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore 1987).

[3] For a summary see H.Frauenfelder, S.G.Sligar and P.G.Wolynes, Science **254**, 1598 (1991)

[4] I might mention that flow towards a solution having unique function with respect to a particular measure but varying microdetails with respect to others is not uncommon; for example, in the neural memory task of person recognition one regularly identifies and responds emotionally to another person independently of their facial expression, clothes etc., even when the latter are noticed.

[5] Hence the picture of a 'folding funnel'; see for example P.G.Wolynes, J.N.Onucic and D.Thirumalai, Science **267**, 1619 (1995).

[6] Although some of these systems might seem random to a casual observer this 'randomness' is analagous to that of a language of which one is ignorant; like the Navajo 'code' used by the Americans in the last world war and unbroken by the Japanese, or like Japanese script to an American (Navajo or Caucasian) who does not know how to read it.

[7] D.Sherrington, Spec.Sci.Tech. **14**, 316 (1991)

[8] The overlap of two states is a measure of their microscopic similarity; for a system of binary microvariables it is given by the fraction of microvariables which are identical in the two microstates minus the fraction which are different.

[9] Given three states S, S', S'', their overlaps $q(S, S')$, $q(S', S'')$ and $q(S'', S)$ are ultrametrically related if the two smallest overlaps are equal.