

脳における情報処理の基本原理に関心が集まっている。多数のニューロンの相互作用による興奮パターンのダイナミクス、これこそが脳の並列情報処理を担うものである。さらに、このダイナミクスは学習や自己組織化によって性能を向上させていく。脳の情報原理の解明には、物理学、数学、情報科学、工学モデルなど、多くの学際的研究の協力が必要である。統計神経力学などを例として、物理学と脳の情報論のかかわりをかいまみることにしよう。

## 1. はじめに

脳の情報処理原理、これを現代の科学に残された最大の謎といってよい。脳の中で情報はどのように表現されているのか、脳はこの情報をどのようなアルゴリズムで計算するのか、記憶のメカニズムはどのようなものであるか、こういった脳の基本原則を求めて最近研究が活発になってきた。<sup>1,2)</sup> 脳はいうまでもなく、神経細胞よりなる複雑な回路網である。神経細胞の働きをマイクロに探求していくと、生体分子の精妙な分子認識作用に行き当たる。この観点から見れば、脳は巧妙な分子機械であり、ここに脳機能を支える物質的原理が見える。こうして分子生物学は脳の研究に本格的に乗り出してきた。

しかし、脳は一方では高度な情報処理機械である。そこでは多数の要素(神経細胞)の同時並列の相互作用のダイナミクスを通じて情報が処理されていく。さらに、学習や自己組織化によって神経回路のダイナミクスそのものを修飾していく。こうした情報過程における脳の基本原則は、チューリング・マシンに代表される現代のコンピュータの情報原理とは全く異ったものである。脳は神経回路網を用いてこの別種の情報原理を実現した、高度で特異な情報機械であるといえる。

脳の機能を解明するには、その物質原理だけでなく、情報原理を研究することが重要である。情報原理は、多数の要素の非線形、非平衡の相互作用とその自己組織化に支えられている。これを解明するには、比較的単純な神経回路網モデルを構成し、これをモデルとして取上げ、そこで可能な情報過程を研究する方法が有効である。<sup>3)</sup> これによりこうした回路網の原理的な可能性と限界とを明らかにしていけばよい。これまでにも、Caianiello, Cooper, Hopfield など多くの物理学者が脳の情報原理に挑戦してきた。非線

形の多体問題、さらに自己組織化という点が彼らの興味を引いたのであろう。こうした観点からは、神経回路網の動作と物理系とに共通する面も多い。しかし、既成の物理学の理論や方法をこの分野にただ持ち込むだけでは新しい発展はない。物理の原理ではなくて、情報原理の体系をこの分野に新しく打ち建てなくてはならない。<sup>3)</sup>

ここでは、統計神経力学の基礎的問題、連想記憶モデルとスピングラス、自己組織化機能を持つ神経場における情報表現の自己形成、この三つの話題を紹介しよう。物理学との方法的共通性と内容的な相違をみてほしい。

## 2. 神経細胞の数理モデル

神経細胞は、情報の観点からは、多入力1出力の情報処理素子である。一つの神経細胞は、他の  $n$  個の細胞から信号  $x_1, \dots, x_n$  を受け取り、これを総合し、答を出力信号  $z$  として他の細胞に伝える(図1)。各信号  $x_1, \dots, x_n$  を受け取る時に、それぞれに結合強度  $s_1, \dots, s_n$  がかかる。入力信号をベクトル  $x = (x_1, \dots, x_n)$ 、結合強度をベクトル  $s = (s_1, \dots, s_n)$  と書くと、信号  $x$  を受け取った時にはこれを総合した刺激の強度は内積  $s \cdot x$  となる。出力信号  $z$  はこ

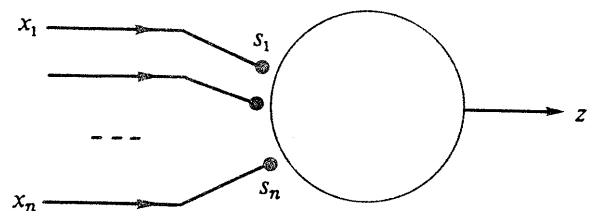


図1 神経細胞のモデル。出力  $z$  は入力信号  $x_1, \dots, x_n$  の重み付きの和、 $\sum s_i x_i$  の関数として定まる。

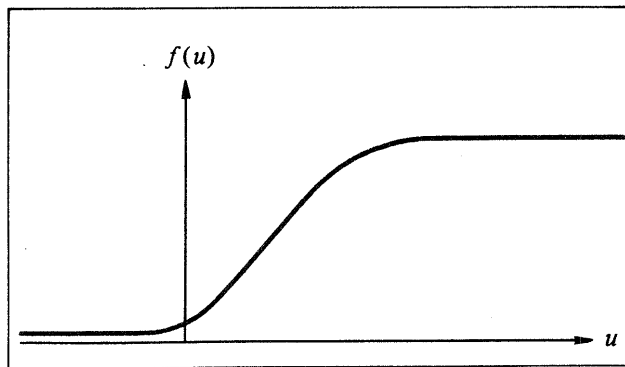


図2 出力関数の特性. 出力  $z$  は入力  $u = \sum s_i x_i - h$  の非線形関数で,  $z = f(u)$  となる.

これを非線形関数  $f$  で変換したものであって, 入出力関係は

$$z = f(u), \quad u = s \cdot x - h \quad (1)$$

のように書ける.  $f$  は図2に示すような単調関数,  $h$  は閾値で,  $s \cdot x$  が  $h$  より小さいときは出力はほとんど出ない.

関数  $f$  を次の単位階段関数

$$f(u) = \begin{cases} 1, & u > 0 \text{ のとき} \\ 0, & u \leq 0 \text{ のとき} \end{cases}$$

で近似すれば, 出力信号は 0, 1 の 2 値を取る. ( $f$  として階段関数でなくて符号関数  $\text{sgn}$  を用いることもある.  $\text{sgn}(u)$  は  $u$  の正負に応じて +1, -1 の値を取る.) 一般には, 入出力信号はアナログ値を取る. 現実の神経では, これは興奮パルスのパルス頻度という形で表される.

一方, 神経細胞の入出力動作に確率的な性質を入れようというモデルもある. 入出力信号は 0, 1 の 2 値をとるものとし, 入力の重みつき総和が  $u$  のときは, 神経細胞は確率  $f(u)$  で興奮して出力  $z=1$  を出し, 確率  $1-f(u)$  で  $z=0$  となるとするものである. いま  $f(u)$  がパラメータ  $T$  に依存して

$$f_T(u) = \frac{1}{1 + e^{-u/T}} \quad (2)$$

のような形をしていたとしよう.  $T \rightarrow 0$  の極限では  $f_T(u)$  は 0, 1 の 2 値を取る階段関数になるので確率的なゆらぎは消える.  $T$  が大きいと  $f_T(u)$  の傾きはゆるやかになり,  $T \rightarrow \infty$  では  $f_T(u) = 0.5$  となって, 入力の強さに関係せず出力は 0 と 1 を確率 0.5 で出すことになる.  $T$  は確率的なゆらぎの強さを表すパラメータで, これを物理系になぞらえて“温度”と呼ぶ.

今までは, 時間の要素を無視して入出力関係を記述した. 時間を離散的に刻んで  $t=1, 2, 3, \dots$  として, 各整数時点て信号が伝えられると考える場合にはこれでよい. しかし, 連続時間で考えるときは, 入力刺激の総和効果を表す  $u$

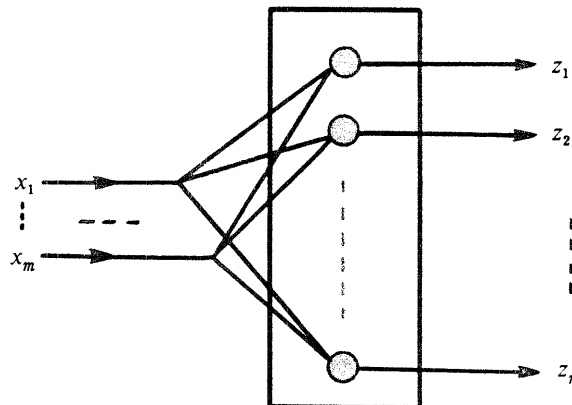


図3 神経回路網による情報変換. 最も単純な層状の神経回路網は, 一束の入力信号  $x=(x_1, \dots, x_m)$  を一束の出力信号  $z=(z_1, \dots, z_n)$  に非線形に変換する.

と入出力信号を連続時間  $t$  の関数とし,  $\dot{u} = du/dt$  として,  $u(t)$  が

$$\begin{aligned} \tau \dot{u} &= -u(t) + s \cdot x - h, \\ z(t) &= f[u(t)] \end{aligned}$$

のような微分方程式に従うものとし, 出力  $z$  は  $u$  の関数と考えるモデルが多く用いられている.

神経回路網の特性は学習により変わっていく. 特に生体の発達初期には, 回路網の構造を整備し外界の情報構造に適合するように調整する自己組織化機能が重要である. 神経細胞一つを取り上げれば, 入力信号  $x$  に対する結合効率  $s$  を変えることで, 自己の入出力関係を調整できる.  $s$  の変化の法則は生理学的にはまだ明確にはなっていないが, モデルは次の法則を仮定する.

$$\tau' \dot{s} = -s + crx. \quad (3)$$

つまり,  $s$  は放っておけば時定数  $\tau'$  ( $\tau$  よりはずっと大きい) でゆっくりと減衰するが, 学習用の信号  $r$  が 0 でないときはその時の入力信号  $x$  に比例して増加する.  $c$  は定数で, 学習信号  $r$  はいろいろな状況や信号に依存して決まる. この中で有名なものは Hebb の学習仮説で, この神経細胞が興奮したとき, このときに限って学習が進むというものである. これは  $r=z$ , すなわち自分の出力  $z$  自体を学習信号とする学習方式である.

### 3. 神経回路網のダイナミクス

はじめに, 図3に示すように, 神経細胞を  $n$  個並べた集団を考えよう. 各細胞からの出力信号をそれぞれ  $z_1, \dots, z_n$  とし, これをまとめた  $z=(z_1, \dots, z_n)$  をこの集団(回路網)全体の出力とする. この回路網へは  $m$  本の入力信号が入って来るとし, それをまとめてベクトル信号  $x=(x_1, \dots, x_m)$  で表す. 第  $i$  番目の神経細胞がこの  $m$  本の入

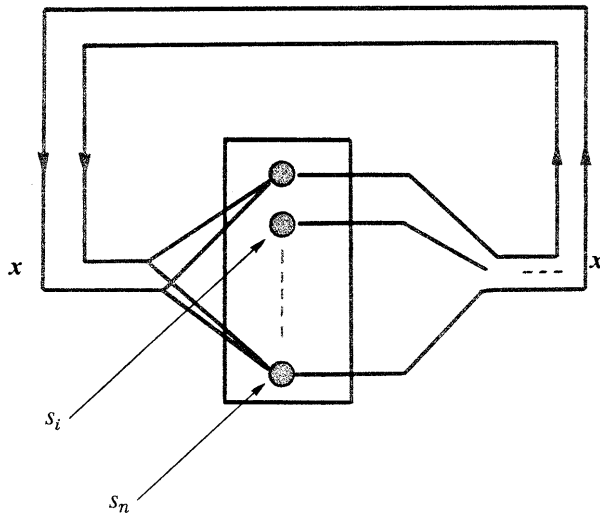


図4 神経回路網のダイナミクス. 相互結合のある神経回路網は、内部状態  $x=(x_1, \dots, x_n)$  がフィードバックして次の状態を決める.  $s_i$  は外部から直接に各ニューロンに入る入力で、 $a_i$  は  $a_i = s_i - h$  におく.

力信号を受け取るときの結合効率を  $s_{i1}, \dots, s_{im}$  としよう. このとき、回路網の入出力動作は

$$z_i = f\left(\sum_{j=1}^m s_{ij}x_j - h\right) \quad (4)$$

と書ける. この神経回路網は入力情報  $x$  を出力情報  $z$  に変換する装置とみなせる. この変換は結合効率の行列  $S=(s_{ij})$  に依存するから、これを

$$z = T_S x \quad (5)$$

と書こう.  $T_S$  は (4) 式で示される非線形の変換である.

次に神経細胞どうしが相互に結合しているもっと本格的な回路網を考えよう.  $i$  番目の神経細胞の出力が  $j$  番目の神経細胞に入力として結合するときの結合効率を  $s_{ji}$  としよう. このような回路網は、図3の回路の出力の全体をぐるっとフィードバックして入力側にまわしたものと考えるとよいから、図4のように表せる. 時間  $t$  が離散的であるとして、時間  $t$  での神経細胞の出力を  $x_1(t), \dots, x_n(t)$  とし、これをまとめてベクトル  $x(t)$  で表して、回路の状態と呼ぼう. (前は出力を  $z$  と書いたが、どうせこれが入力例にまわって次の時間の入力になるから、状態  $x$  だけを用いればよい.) また、各神経細胞にはフィードバックで入ってくる  $x$  の信号だけでなく、回路網の外部からの入力もあるものとしてよい. 第  $i$  番目のニューロンに外部から直接入る刺激の総和  $s_i$  から閾値  $h$  を引いたものをまとめて  $a_i$  と書くことにする. このとき、回路網で、時間  $t+1$  の各細胞の状態は

$$x_i(t+1) = f\left(\sum_{j=1}^n s_{ij}x_j(t) + a_i\right) \quad (6)$$

のように書ける. これをベクトルで書いて

$$x(t+1) = T_S x(t) \quad (7)$$

のように表そう.  $T_S$  は非線形の状態遷移演算子で、内部

の結合行列  $S$  と共に外部入力  $a_i$  にも依存している.

連続時間でダイナミクスを書く時は

$$\tau \dot{u}_i = -u_i + \sum s_{ij}x_j(t) + a_i, \quad (8)$$

$$x_j(t) = f[u_j(t)]$$

のようによければよい. しかし、しばらくの間は離散時間で話を進めよう.

仮に、各  $x_i$  が 0 から 1 の 2 値をとるものとしよう. このとき、可能な状態  $x$  の数は全部で  $2^n$  個ある. すなわち、状態空間  $X$  は  $2^n$  個の要素からなる有限集合である. 状態遷移演算子  $T_S$  は、各  $x$  に対してその次の状態  $x' = T_S x$  を一つ定める. (ここで外部入力  $a$  は与えられた定数として固定して考えている. この外部入力  $a$  は状態をリセットしたり、初期状態を設定するのにも使う.) 回路網の状態遷移のダイナミクス (7) は、各  $x$  から次の状態  $x'$  に線を引きついで次の状態を示した状態遷移図を見ればよくわかる. いま

$$T_S x_0 = x_0 \quad (9)$$

であるような状態、つまり次の状態も自分と同じ状態があったとしよう. これは平衡状態であって、一度この状態に入るとここから抜け出せない. ある状態  $y$  から出発して、何回かの状態遷移の後に状態が  $x_0$  になったとしよう. つまりある  $k$  に対して

$$T_S^k y = x_0$$

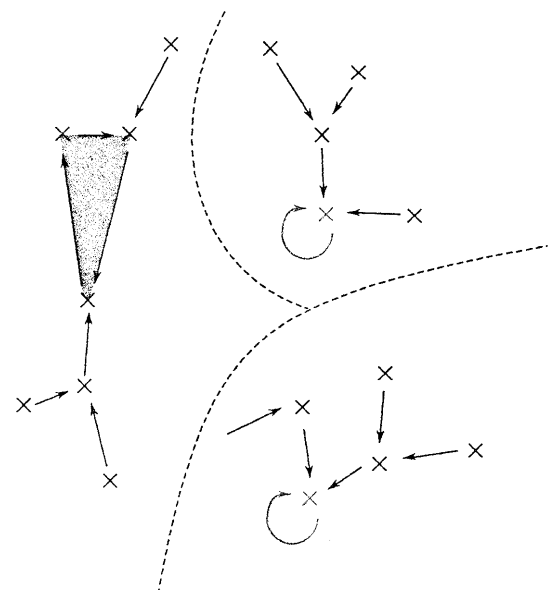


図5 状態遷移図. 各状態  $x$  から次の状態  $x'$  へ矢印を引いたもので、図の赤が平衡状態、紫の三角形に相当するものが周期状態になっている.

となるとする。このような状態  $y$  をすべて集めた集合  $B(x_0)$  を、平衡状態  $x_0$  の“流域”という。回路網の状態がここから出発すれば、最後の状態は  $x_0$  に落ち着く。平衡状態は一般に多数個ある。

一方、

$$x \rightarrow T_S x \rightarrow T_S^2 x \rightarrow \dots \rightarrow T_S^k x = x$$

のように、 $k$  回の状態遷移で元に戻る繰り返しの状態を、周期  $k$  の周期状態という。

いま、結合効率が対称 ( $s_{ij}=s_{ji}$ ) で、しかも自己結合がない ( $s_{ii}=0$ ) 回路網を考えてみよう。ここで状態  $x$  の“エネルギー”という仮想的な状態量を

$$E(x) = -\frac{1}{2} \sum_{i,j} s_{ij} x_i x_j - \sum_i a_i x_i \quad (10)$$

という二次形式で定義する。ここで、 $x$  の第  $i$  成分  $x_i$  が 0 であったとして、これが 1 に変わるとエネルギーはどう変わるか、これを計算してみると

$$\begin{aligned} \Delta_i E &= E(x; \text{ただし } x_i=1) - E(x; \text{ただし } x_i=0) \\ &= -\sum_j s_{ij} x_j - a_i \end{aligned}$$

となる。これは  $\Delta_i E < 0$  ならば  $\sum_j s_{ij} x_j + a_i > 0$  で状態遷移によって  $x_i$  は 0 から 1 に変わるし、逆に  $\Delta_i E > 0$  ならば  $x_i$  が 1 のときは次には  $x_i=0$  になることを意味している。つまり、神経回路網のダイナミクスは、常にエネルギー  $E$  を減少させる方向に状態を動かし、 $E$  の極小値が平衡状態に対応していて、ここに状態が落ちついてしまうことがわかる。<sup>4)</sup> 正確にいうと、これは各細胞が非同期的に勝手に状態遷移するときにいえることである。

さて、これにより対称結合の場合はエネルギーは常に減少して極小状態に落ち着くこと、したがって周期状態は存在しないことがわかった。それゆえ回路網の動作を知るにはエネルギー関数  $E(x)$  の形を調べればよい。二次式の最大最小は簡単と思うかもしれないが、 $E$  の値を  $2^n$  個の離散的な  $x$  (つまり  $n$  次元超立方体の格子点  $X = \{0, 1\}^n$ ) の上で評価するから、2次式といっても  $E$  の形は複雑である。また逆に、周期解などの動的な様相を論ずるには対称結合の回路網ではだめということでもある。

後で述べる連想記憶モデルでは、記憶したベクトル  $z_1, \dots, z_k$  を  $k$  個与えたときに、これらが回路網の平衡状態になるように学習によって結合効率をうまく調節する問題を考える。記憶したパターンの一つ、例えば  $z_1$  を“思い出す”には、初期状態として  $z_1$  の流域  $B(z_1)$  に属する状態を外部から入力すれば、あとはダイナミクスで  $z_1$  が出てくる。だから流域が大きければ大きいほどよい。また認

知科学では、いろいろな制約条件のもとで外部の状況を解釈する過程として、このような対称結合の回路のダイナミクスを想定している。<sup>1)</sup>  $E$  が小さいほど、矛盾の少ない解釈を与えるという想定で、回路網の多安定性 (平衡状態がたくさんあること) は解釈の多義性に対応する。また、Hopfield と Tank<sup>5)</sup> は、“巡回セールスマン問題”のような組合せ最適化問題をエネルギー最小化問題として定式化し、これをアナログ型の連続時間の神経回路モデルに埋め込んで、回路網のダイナミクスでこの問題を並列に短時間で解く方法を示して注目を浴びた。もっとも、神経回路では厳密な最小解は求まらず、極小解が得られるだけである。Hopfield らは埋め込み方をうまく工夫すれば、“かなりよい”解、つまり実用上は十分な近似解が得られると主張している。

神経細胞の動作が確率的な場合には、状態遷移  $x \rightarrow T_S x$  は確率的に決まる。これは数学の言葉でいえばマルコフ連鎖になっていて、時間がたつにつれ定常的な確率分布  $p(x)$  に収束していく。定常確率分布を陽に求めるのは一般には困難である。しかし、状態遷移の確率が (2) 式で決まる場合には、定常分布、つまり十分時間がたった後に、状態が  $x$  にいる確率は

$$p(x) = \frac{1}{Z} \exp\left\{-\frac{E(x)}{T}\right\} \quad (11)$$

で与えられる。 $Z$  は規格化定数 (物理の言葉でいえば分配関数)、 $T$  は確率的ゆらぎの強さを示す“温度”である。これはまさしく“ボルツマン分布”に他ならない。しかも、 $x$  はエネルギーの小さなところに大きな確率が集中している。温度  $T \rightarrow 0$  の極限では、エネルギー最小値のところに確率が集中し、 $T \rightarrow$  大となるとゆらぎが大きくなり  $p(x)$  は一様分布に近づいてしまう。こうした特性を利用して、“ボルツマン・マシン”と呼ぶ神経回路を作り、そこに外界の情報確率構造を学習によって獲得させるアイデアがある。<sup>6)</sup> また、極小値に落ち込まないで  $E(x)$  の最小値を得る方法として、温度  $T$  を“準静的”に下げていくというアイデアがある。これをシミュレーションによる焼きなまし (シミュレーテッド・アニーリング) 法という。神経回路網のダイナミクスの特性について、まだまだ理論的に研究すべき課題は多い。以下にその一端を示す。

#### 4. 統計神経力学とスピングラス

結合効率  $s_{ij}$  がある確率法則に基づいて、ランダムに決められた回路網を考えてみよう。各  $s_{ij}$  を一度ランダムに

決めると、あとは固定する。このようにランダムに作られた回路網を多数集めた神経回路網のアンサンプルを $\varepsilon$ とし、細胞数 $n$ は十分大きいとして、 $\varepsilon$ の中のほとんどすべての回路網に成立するような“巨視的”な性質を探求していく。これを統計神経力学と呼ぼう。確率論の言葉でいえば $\varepsilon$ の中に $s_{ij}$ の確率に基づいて確率測度が導入され、この測度で議論することになる。<sup>7)</sup>

もちろん、現実の神経回路網はランダム結合ではないだろう。しかし、ランダム回路網について成立する巨視的な法則は、 $\varepsilon$ のほとんどすべての回路網について成立し、またランダムでなくても複雑な形の結合をしている回路網の動作についても示唆するところが多いと考えられる。

例として、 $s_{ij}$ がすべて同一の確率分布からそれぞれ独立に選ばれる場合を考えてみよう。現在の回路の状態を $x$ とすると、次の時間での状態 $x' = T_S x$ は $S$ によるから、 $x'$ は $\varepsilon$ の中の各回路毎に異なっている。ここで、状態 $x$ そのものを微視的に見るのをやめて、巨視的な状態量として、回路網の中で何個の細胞が興奮しているかその割合を示す量

$$A(x) = \frac{1}{n} \sum x_i \quad (12)$$

を採用してみよう。 $A$ を回路網の活動度という。現在の状態 $x$ の活動度が $A$ であるとき、次の時間での活動度 $A' = A(T_S x)$ は結合効率の行列 $S$ に依存して決まるから、 $\varepsilon$ の中の回路毎に異なっているように見える。しかし、 $T_S x$ の形と $A$ の形を見ればわかるように、 $A'$ には非常に多数の $s_{ij}$ が和の形で関与し、結果として大数の法則が働いてしまい、 $A'$ はほとんどすべての回路でほとんど同じ値になってしまう。そこには $\varepsilon$ の回路網に共通に

$$A' = \varphi(A) \quad (13)$$

という関係が成立する。巨視的な状態遷移関数 $\varphi$ は $s_{ij}$ の確率法則にはよるが、個々の $s_{ij}$ の値にはよらない。(関数 $\varphi$ は $s_{ij}$ の平均と分散のみで表せる。中心極限定理を用いてこれを導いてみよう。)

神経回路網のダイナミクス $x(t+1) = T_S x(t)$ に戻ると、微視的な状態 $x(t)$ 自体は回路毎に異なっているだろう。しかし、巨視的な状態

$$A_t = A\{x(t)\}$$

はどの回路にも共通に、

$$A_{t+1} = \varphi(A_t) \quad (14)$$

という方程式に従うだろうか。(13)が成立するからといって(14)が成立すると即断するわけにはいかない。 $x(t)$

は $S$ に依存しており、これより $T_S x(t)$ を計算するとき、 $T_S$ の $S$ と $x(t)$ に含まれる $S$ との微妙な相関がきいてくるかもしれないからである。この辺の事情はボルツマンが気体の速度分布の発展方程式(ボルツマン方程式)をたて、そこからエントロピー増大の $H$ 定理を導いたときと全く同じである。すなわち、“分子混沌の仮定(衝突数の仮定)”に対応する仮定( $T_S$ と $x(t)$ の相関を断ち切る仮定)が神経回路網の場合に成立するかどうかにかかっている。

力学系の場合にはKacらの努力にもかかわらず、この問題は解決をみていない。しかし、統計神経力学ではある程度の理論が作られていて、<sup>8)</sup>(14)は“弱い意味”で成立することが証明されている。また、二つの状態 $x$ と $y$ の間に距離

$$D(x, y) = (1/n) \sum |x_i - y_i|$$

を導入したときに、 $T_S x$ と $T_S y$ の距離

$$D' = D(T_S x, T_S y)$$

は $x$ と $y$ との距離 $D = D(x, y)$ にどのように依存しているかを論じて、

$$D' = \varphi(D)$$

の形の法則を求めることもできる。ここから、 $\varepsilon$ の典型的な回路網における状態遷移図の様相を調べ、さらに神経回路網のダイナミクスで収束がいかに速いかについて、ある程度の手がかりが得られる。<sup>7)</sup>

結合はランダムに決まるけれども、対称性 $s_{ij} = s_{ji}$ が常に成立している場合にはどうなるだろう。この場合のポテンシャル $E(x)$ の形を良く見ると、スピン系との類似に気がつく。神経細胞の興奮非興奮を $x_i = 1, 0$ で表すのではなく、 $-1$ で表し、これをスピンの上向き下向きにたとえてみよう。すると、 $E$ は $n$ 個スピン系のエネルギー、 $s_{ij}$ は第 $i$ 番目のスピンと第 $j$ 番目のスピンの間の相互作用エネルギー(Heisenbergの交換エネルギー)、それに $a_i$ は $i$ 番目の素子の外部磁場の強さになっている。こうした系は相互作用のダイナミクスをしながら、 $E$ の極小値へ落ちつくだろう。熱的なゆらぎがあれば(温度 $T$ が有限)、確率的な動作をしながら平衡分布であるボルツマン分布に近づくが、これはボルツマン・マシンそのものである。

強磁性では $s_{ij}$ は正である。 $s_{ij}$ が隣接したスピンの間だけで正、あとは0で近似すると、イジング・モデルになる。反磁性体というものもある。ところでスピングラスというのは、 $s_{ij}$ の値が正負あたまもランダム値をとる。遠隔のスピンの間まで影響が及んでいいので、いっそ $s_{ij}$ が $s_{ij} = s_{ji}$ の対称性を保ちながら各成分が独立にランダム

ムに定まったとして解析してみようということになる。Sherrington-Kirkpatrick<sup>9)</sup>のモデルがこれで、対称結合のランダム神経回路網と全く同じモデルになっている。対称性を入れただけで、前に述べた  $s_{ij}$  と  $s_{ji}$  も別々にランダムに決める非対称の場合とは話が大部ちがってくる。非対称の場合は周期解があり、平衡状態の数の期待値は1であるが、対称の場合は周期解がなく平衡状態の数の期待値は約  $2^{0.29n}$  である。しかし両者に共通点も多く、両者の性質を考えることは統計神経力学として興味がある。

対称の場合に、平衡状態を  $k$  個  $z_1, \dots, z_k$  のように外部から指定し、これらすべてが平衡状態

$$z_\mu = T_S z_\mu, \quad \mu = 1, 2, \dots, k \quad (15)$$

になるように結合  $S$  を指定できないものだろうか。これは相互作用のダイナミクス  $T_S$  の設計の問題である。各  $z_\mu$  を“記憶しておきたい信号”とするならば、回路網のダイナミクスの中にこれらを記憶したことになり、各  $z_\mu$  の流域を初期状態として与えれば  $z_\mu$  を再生できる。人間の連想記憶の奥にひそむ仕掛けも案外こんなものかもしれない。流域はなるべく大きくしたい。一つの回路網に平衡状態として記憶できる信号の数  $k$  はどの位であろうか。

いま、信号  $z_\mu$  の各成分は確率0.5で1か-1の値を取るように独立に選ばれたとしよう。このとき、回路網の結合効率  $s_{ij}$  を

$$s_{ij} = \frac{1}{n} \sum_{\mu=1}^k z_{\mu i} z_{\mu j} \quad (16)$$

のように選んでみる。  $z_1, \dots, z_k$  が互いに厳密に直交していれば

$$S z_\mu = z_\mu$$

がすべての  $\mu$  について成立するから、(15)も当然成立する。  $z_\mu$  がランダムに選ばれたものならこれらは近似的に直交に近いから、(15)が成立しそうな気がする。こういうアイデアで出されたのが、相関形連想記憶のモデルであり、<sup>10)</sup> 物理系とのアナロジーを強調したのが Hopfield<sup>9)</sup> である。このモデルでは、 $S$  の要素  $s_{ij}$  はランダムではあるが互に独立ではなく、 $z_{\mu i}$  を通じて互に相関がある。しかし統計神経力学の考えは、こうした場合にも適用できる。

Hopfield は計算機シミュレーションによって、記憶容量(記憶できる信号の数)は  $k=0.15n$  であろうと推察した。Amit ら<sup>11)</sup> はスピングラスを解析する手法を發展させ、確率的に動作する神経回路網の平衡状態の分布や、相転移の様子を調べ、温度  $T$  が0の極限である意味で Hopfield の予想が正しいことを示した。流域の大きさを議論するには

平衡状態だけではなくて非平衡のダイナミクスを論じなくてはいけない。我々は統計神経力学を用いてこれを解析し、収束すべき状態と現在の状態の距離を巨視的状态と考えて、

$$D_i = \xi(D_{i-1})$$

の形の方程式を作ろうと試みた。興味あることに、 $D$  に関しては衝突数の仮定は成立せず、以前の状態との相関が無視できない。このため、相関を示す状態量を新しく導入し、巨視的状态量を拡大することで、連想記憶モデルの想起過程の非平衡の過渡的ダイナミクスの興味ある性質を明らかにしている。<sup>12)</sup>

## 5. 外部世界の内部表現と自己組織化

これまで、神経回路のダイナミクスを調べて来た。神経回路はダイナミクスによって並列の情報処理を協同して行うだけでなく、学習や自己組織化によって自己のダイナミクスを外界の情報構造に合わせて改良していく。その一つとして、神経場が外界情報の表現を自己組織化により作っていく様子を示すモデルを考えてみよう。

外界の情報は  $n$  次元ベクトル  $x$  で与えられるものとし、信号空間  $X$  上で確率分布  $p(x)$  が与えられているとする。つまり  $p(x)$  の大きい信号  $x$  は頻度高く出現するし、 $p(x)$  が0であったり小さい信号はめったに出現しない。極端な場合は、有限個の信号  $x_1, \dots, x_k$  のみが発生するという確率構造(状況)を考えてもよい。分布  $p(x)$  に従って次々に発生するこうした信号を2次元の神経場が受取るものとして、神経場とは、神経細胞が2次元にぎっしりと並んだもので、これを空間的に連続化して“場”と考える。神経場にその位置を示す2次元の座標系  $\xi = (\xi_1, \xi_2)$  を導入し、場所  $\xi$  の付近にある神経細胞の時間  $t$  のポテンシャルは平均的に  $u(\xi, t)$  であるとして、神経場内部で神経細胞は相互にフィードバック結合をしているものとし、場所  $\xi'$  にある神経細胞の出力  $z(\xi', t) = f\{u(\xi', t)\}$  は、場所  $\xi$  にある神経細胞に結合効率  $w(\xi, \xi')$  の強さで入力するものとする。すると、神経場の興奮パターンのダイナミクスを記述する方程式は

$$\tau \frac{\partial u(\xi, t)}{\partial t} = -u(\xi, t) + w \circ f[u] + S(\xi, t) \quad (17)$$

のように書ける。第一項はポテンシャルの減衰項、

$$w \circ f[u] = \int w(\xi, \xi') f[u(\xi', t)] d\xi' \quad (18)$$

は他の場所  $\xi'$  にある神経細胞の出力が  $\xi$  にある神経細胞

に与える影響の総和である。通常  $w(\xi, \xi')$  は  $w(\xi - \xi')$  のみに依存し (場の一様性の仮定), さらに  $|\xi - \xi'|$  が小さいところでは正の値を取り,  $|\xi - \xi'|$  が大きいところでは負という相互抑制形のもを仮定することが多い。  $S(\xi, t)$  は外部からの入力信号  $x$  が場所  $\xi$  にある神経細胞に直接与える影響の総和である。  $x$  は場のあらゆる場所に共通に入力するものとし,  $\xi$  にある神経細胞はこれを結合効率  $s(\xi)$  で受けとるものとする。さらに場には共通の抑制性的入力  $x_0$  が入るものとし, この結合効率は負で強さは  $s_0(\xi)$  であるとする。すると, 場所  $\xi$  にあるニューロンが受取る外部からの入力の総和は, 入力が  $x$  であるときは

$$S(\xi; x) = s(\xi) \cdot x - s_0(\xi) x_0 \quad (19)$$

のように書ける。  $x$  が時間と共に変われば  $S$  も変わる。

入力が  $x$  であるときの式 (17) の平衡状態を  $U(\xi; x)$  としよう。  $U(\xi; x) > 0$  であるような場所  $\xi$  は, 入力  $x$  が来たときに興奮するし, そうでない場所は興奮しない。だから場の興奮パターン  $U(\xi; x)$  は, 入力信号  $x$  の神経場における表現であるとみることができる。とくに, 相互抑制結合  $w(\xi, \xi')$  の抑制が強力なときは,  $U$  はどこか 1 カ所の小さな領域の興奮しか起こらないため, 信号  $x$  は神経場で  $U(\xi, x)$  が正となる小さな場所  $\xi$  の興奮で表現される。こうして,  $X$  から神経場への写像が情報の内部表現として定まる。

この表現は結合効率  $s(\xi)$ ,  $s_0(\xi)$  に依存している。いま内部結合  $w(\xi, \xi')$  は固定で,  $s, s_0$  の二つは自己組織的に変化していった, 外界の情報構造 ( $X$  上の確率分布  $p(x)$ ) に適合するように  $X$  の情報の内部表現を作っていくとしよう。  $s, s_0$  の変化の法則は Hebb の考えたようなもの, つまり神経細胞が興奮したときはその時の入力  $x$  の強さに比例して結合効率が上がるとする。このとき

$$\tau' \frac{\partial s(\xi, t)}{\partial t} = -s(\xi, t) + c x f[u(\xi, t)], \quad (20)$$

$$\tau' \frac{\partial s_0(\xi, t)}{\partial t} = -s_0(\xi, t) + c' x_0 f[u(\xi, t)] \quad (21)$$

という自己組織化の方程式が得られる。<sup>13)</sup>

この方程式を解くのは一般には容易でない。しかし, 1)  $\tau$  と  $\tau'$  では  $\tau' \gg \tau$  が成立し, 準静的近似が使えること, 2)  $x(t)$  は確率  $p(x)$  に従って発生する信号系列であるから,

方程式 (20), (21) の  $x$  を含む項 ( $u$  も  $x$  に依存する) に対してアンサンブル平均を用いる近似を行うこと, この二つを使ってある程度の数学的な解析ができる。

この結果, 信号  $x$  の構造を反映した表現が神経場にできる様子がわかる。特に, 信号間の相関構造が場における表現の形成に関係する。これは神経場トポグラフィの形成に関係していて, その場合の数学的理論が提出されている。<sup>13, 14)</sup> また神経場における表現の分解能や, 表現が離散的なブロック構造に自動的に分かれる場合があること<sup>14)</sup> などが, 場の方程式の平衡状態の安定性を解析することでわかる。つまり, ある場合には連続的な解が不安定になって, 場に非一様構造の自己組織解が現れる。

## 6. おわりに

脳の情報処理様式をめぐって, 生物学的研究, 情報科学的研究, ニューロンコンピュータを目指す応用研究のすべてが活性化してきている。これらは互に交流し刺激しあいながら, 新しい総合的な脳科学を形成していくであろう。現在の動きを単なる人為的ブームに終わらせないためには, 基礎のしっかりした理論体系を一步步築いていく必要がある。本文はそうした試みの一端をのべた。物理学を志す人達の中からこの分野で取り組む人々が出てくることを期待したい。

## 参考文献

- 1) D. E. Rumelhart, et al.: *Parallel Distributed Processing* (MIT Press, 1986).
- 2) T. Kohonen: *Self-Organization and Associative Memory*, 2nd ed. (Springer, 1988).
- 3) 甘利俊一: 神経回路網の数理 (産業図書, 1978).
- 4) J. J. Hopfield: Proc. Nat. Acad. Sci. (USA) 79 (1982) 2445.
- 5) J. J. Hopfield and D. W. Tank: Science 233 (1986) 625.
- 6) D. H. Akley, et al.: Cognitive Sci. 9 (1985) 147.
- 7) S. Amari: Kybernetik 14 (1974) 201.
- 8) S. Amari, et al.: SIAM J. Appl. Math. 33 (1977) 95.
- 9) D. Sherrington and S. Kirkpatrick: Phys. Rev. Lett. 32 (1975) 1792.
- 10) K. Nakano: IEEE Trans. Syst. Man & Cybern. SMC-2 (1972) 381.
- 11) D. J. Amit, et al.: Phys. Rev. A 2 (1985) 1007.
- 12) S. Amari and K. Maginu: Neural Networks 1 (1988) 63.
- 13) S. Amari: Bull. Math. Biol. 42 (1980) 339.
- 14) A. Takeuchi and S. Amari: Biol. Cybern. 35 (1979) 63.