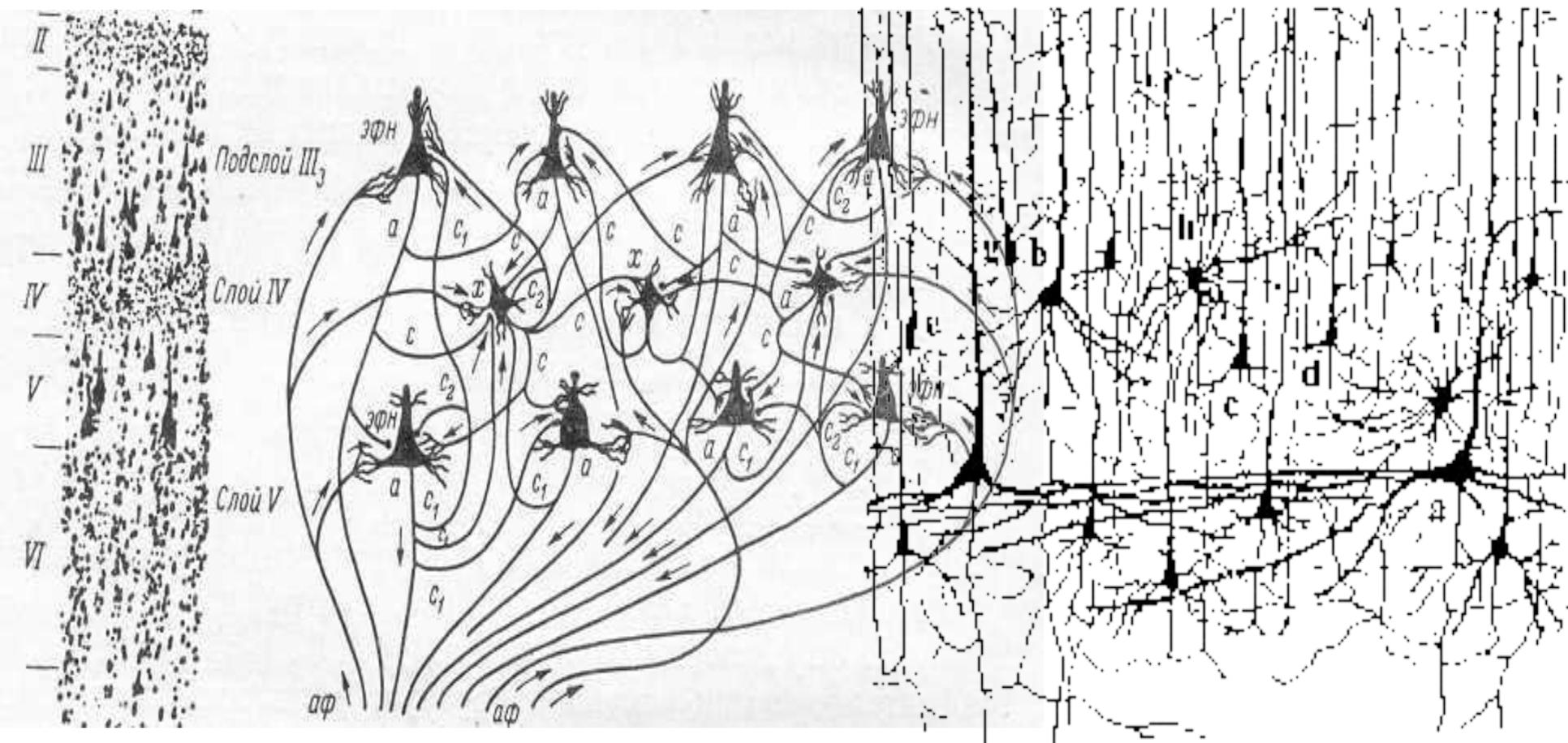


Динамические нейронные сети и их аттракторы

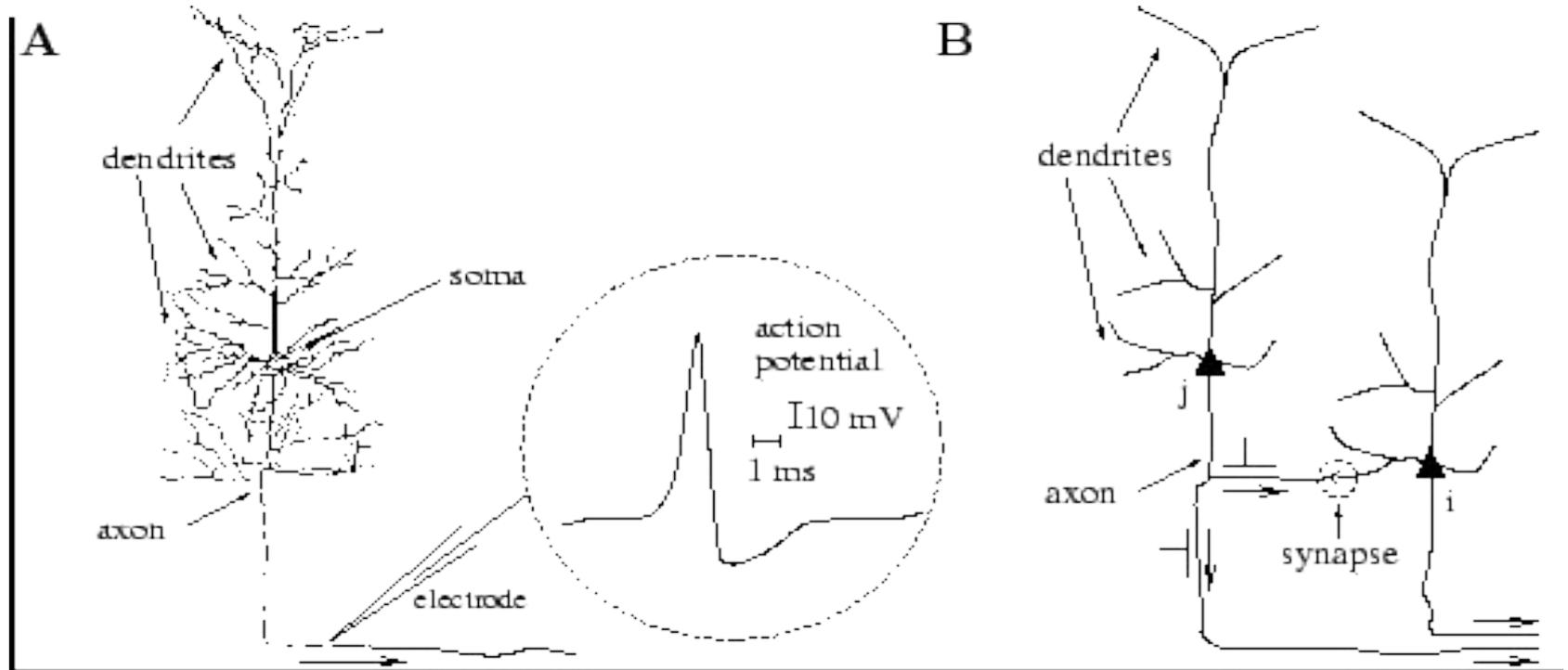
Резник Александр Михайлович

Институт проблем математических
машин и систем НАН Украины
отдел Нейротехнологий
neuro@immsp.kiev.ua

Фрагмент нервной ткани коры мозга



Нервная клетка

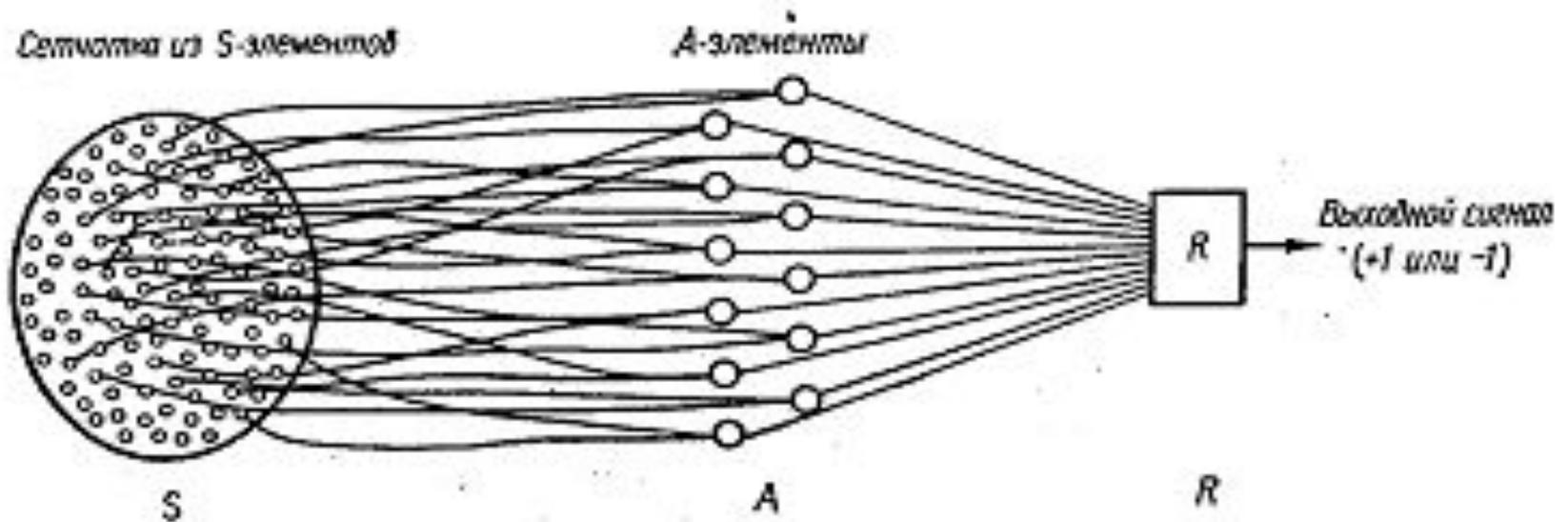


$$y = f\left(\sum_{i=1}^n w_i x_i - b\right)$$

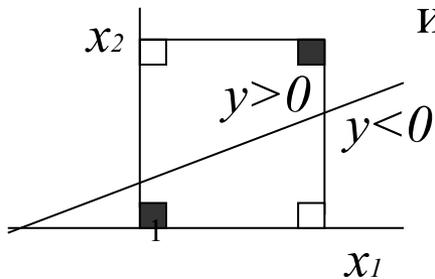
$$f(v) = 1 / (1 + e^{-\alpha v})$$

$$f'(v) = \alpha \cdot f(v) [1 - f(v)]$$

Перцептрон Розенблатта (1957г.)

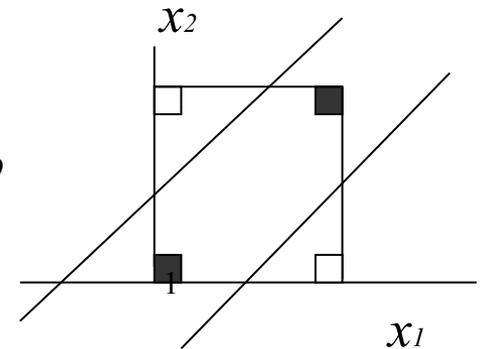
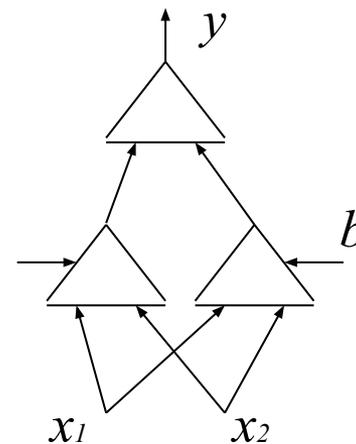


Проблема “исключающее ИЛИ”



и ее решение

$$y = \text{sign}(a_1 x_1 + a_2 x_2 - b)$$



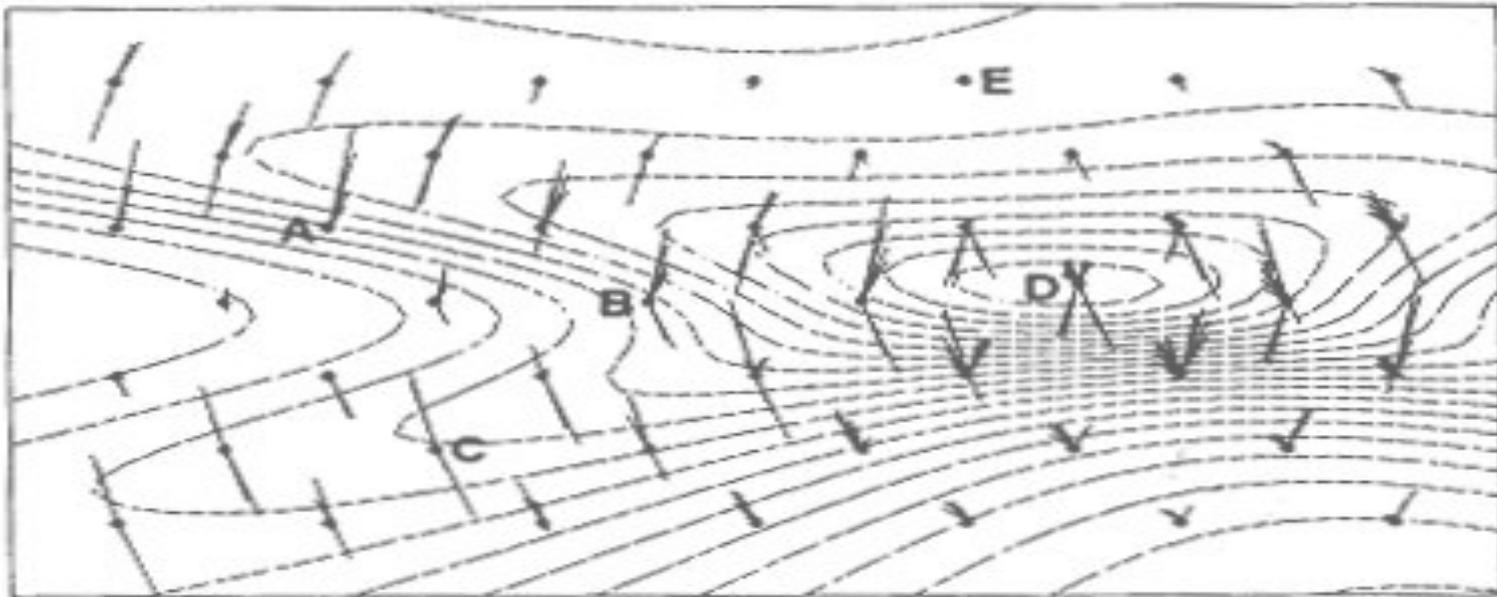
Градиентный подход к обучению

Обучающая выборка $\{X^m, T^m\}_{m=1}^M$

Средняя ошибка $E^M = \frac{1}{2} \sum_{m=1}^M (e^m)^T e^m$ $e^m = T^m - Y^m$

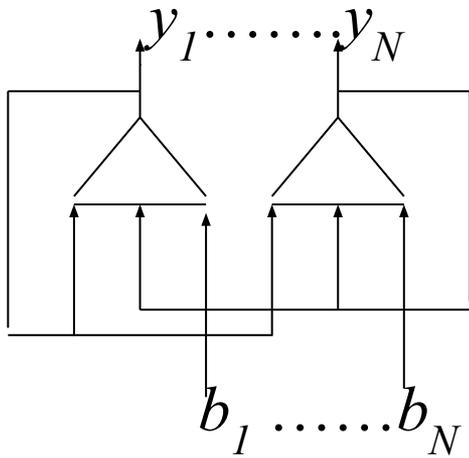
Градиент ошибки $g(W) = \frac{\partial E^M}{\partial W} = \sum_{m=1}^M (e^m)^T J^m$ $J^m = \frac{\partial e^m}{\partial W}$

$$E^M(W_0 + \Delta W) = E^M(W_0) + g^T(W_0)\Delta W + \frac{1}{2}\Delta W^T H(W_0)\Delta W + o(\|\Delta W\|^2)$$



Сеть Хопфилда

Сеть Хопфилда

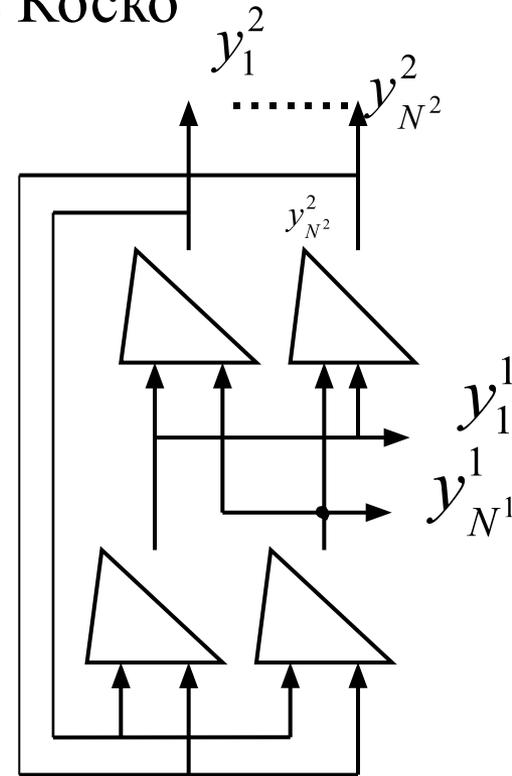


$$y_i(t) = f[s_i(t) - b_i(t)]$$

$$s_i(t) = \sum_{i=1}^N w_{ij} y_i^2(t-1),$$

$$w_{ij} = N^{-1} \sum_{m=1}^M y_i^m y_j^m,$$

Двусторонняя ассоциативная память Коско

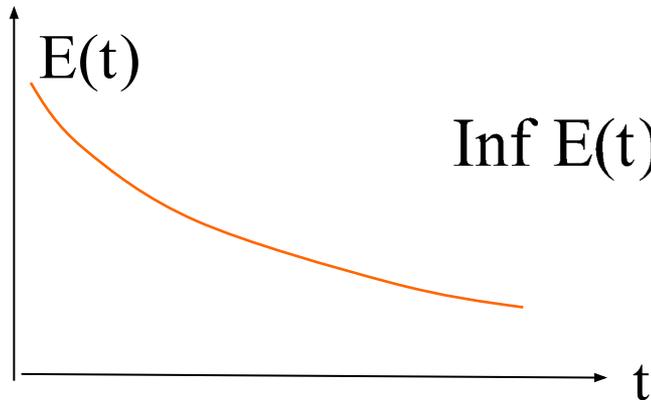


$$y_i^2(t) = \text{sign} \left[\sum_{j=1}^{N^1} w_{ij} y_j^1(t-1) \right]$$

Уравнение энергии сети

$$\begin{aligned} E(t) &= -\frac{1}{2} Y^T(t) S(t) = -\frac{1}{2} Y^T(t) W Y(t-1) = \\ &= -\frac{1}{2} Y^T(t) W^T Y(t+1) + \frac{1}{2} Y^T(t) [W^T Y(t+1) - W Y(t-1)] \end{aligned}$$

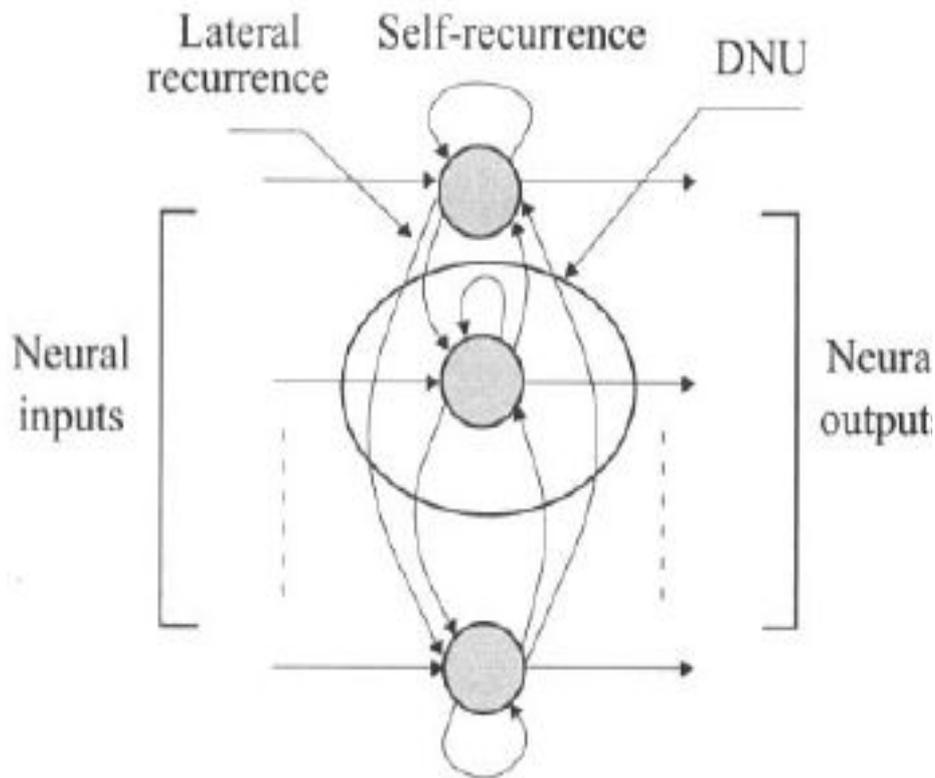
$$E(t) = E(t+1) + \sum_{h=1}^{H(t-1,t+1)} |s_h(t+1)| + \frac{1}{2} Y^T(t+1) (W^T - W) Y(t-1)$$



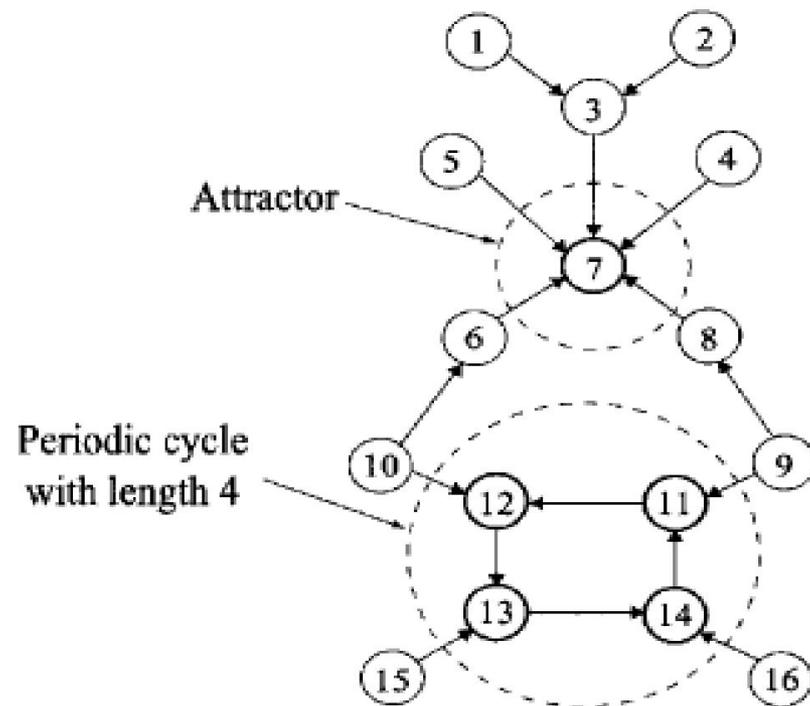
$$\text{Inf } E(t) = -\frac{1}{2} \sum_{i=1}^N y_i(t) \sum_{j=1}^N y_j(t) w_{ij}^{E(t)} = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N |w_{ij}|$$

Обратные связи и аттракторы нейронной сети Хопфилда

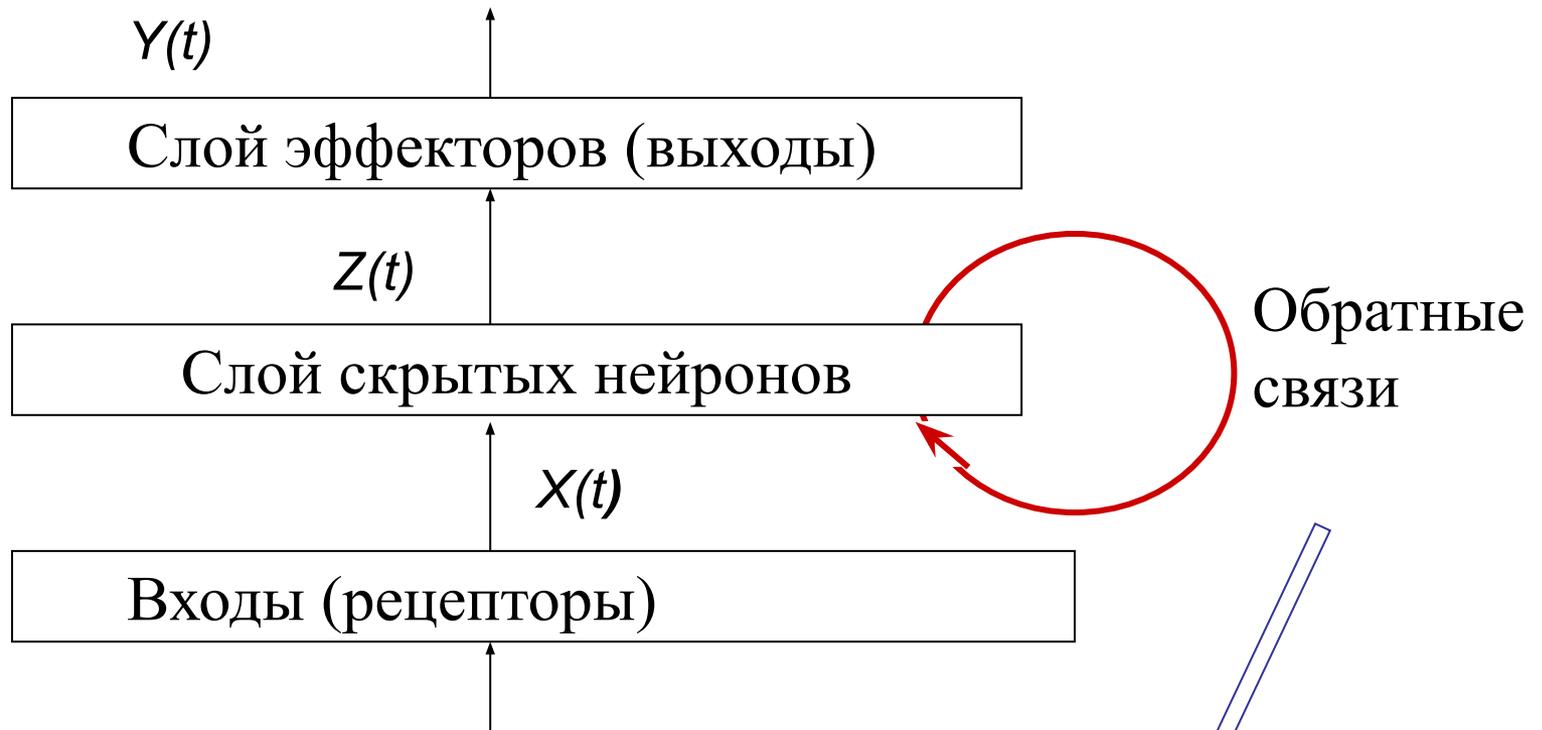
Рекуррентные связи
динамических нейронов



Статический и динамический
аттракторы сети из 4
нейронов



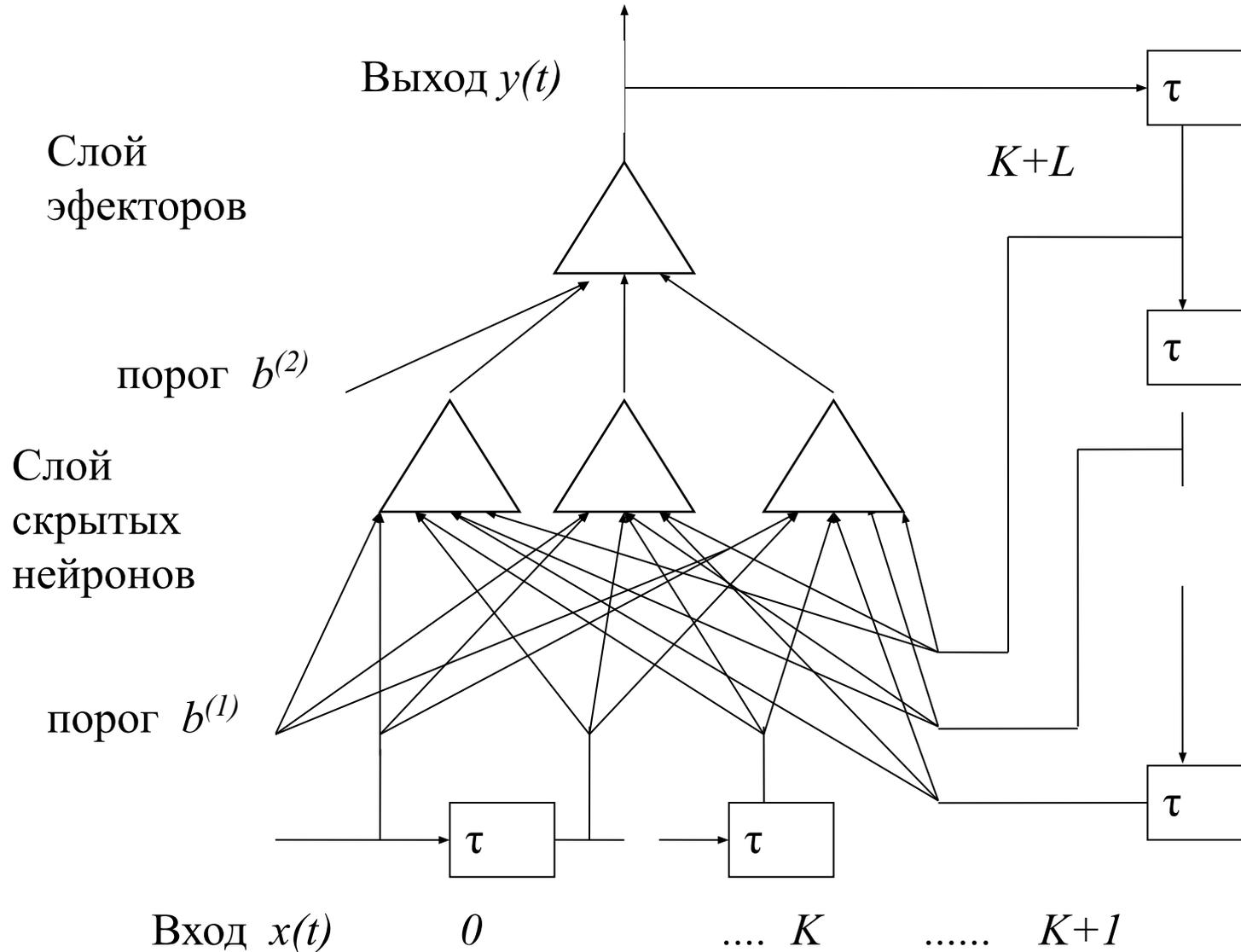
Рекуррентная нейронная сеть



$$Y(t) : \quad y_k(t) = g\left[\sum_j w_{kj} z_j(t) + \beta_i\right]$$

$$Z(t) : \quad z_j(t) = f\left[\sum_i w_{ji} x_i(t) + \sum_h w_{jh} z_h(t-1) + \beta_j\right]$$

Рекуррентный многослойный перцептрон (RMLP).



Нейроны рекуррентной сети

$$z_n^l(t) = f_n^l[s_n^l(t)]$$

$s_n^l(t)$ - постсинаптический потенциал:

$$s_n^l(t) = \sum_{k=l}^L \sum_{j=1}^{N^k} \sum_{\tau \in \Theta_{n,j}} w_{n,j}^l z_j^k(t - \tau) + \sum_{k=0}^{N^{l-1}} w_{n,k}^l z_k^{l-1}(t)$$

$f_n^l(\cdot)$ - нелинейная активационная функция нейрона;

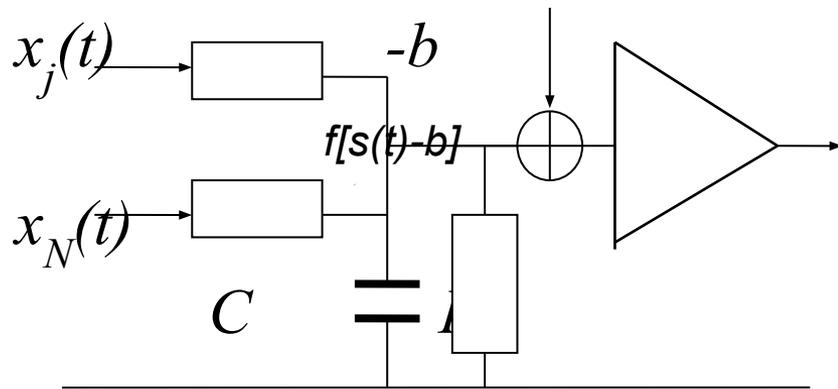
$w_{n,j}^l$ - вес связи для j -входа нейрона l -го слоя

$\Theta_{n,j}$ - множество значений задержки связи τ

Условие устойчивого состояния

$$\partial z_n^l(t) / \partial t = (\partial f_n^l / \partial s_n^l) \partial s_n^l(t) / \partial t = 0$$

Динамический нейрон



$$s(t) + RC \partial s(t) / \partial t = \sum_{j=1}^N (R/R_j) x_j(t)$$

$$z_j(t) = f_j[s_j(t) - b]$$

$$\partial s_i(t) / \partial t = -\alpha_i s_i(t) + h_i(t)$$

$$h_i(t) = \sum_{r=1}^{N_{ir}} w_{i,r} z_r(t - \theta_r) + \sum_{j=0}^{N_i} w_{i,j} z_j(t)$$

настраиваемые параметры:

$w_{i,j}$ веса межнейронных связей

α_i коэффициенты инерции

Обучение рекуррентной нейросети

Функция
ошибки

$$E = \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^{N^L} (d_n^m - z_n^{L,m})^2$$

Обучающая последовательность

$$\{X^m, D^m\}_{m=1}^M; \quad X^m = \{x_i^m\}_{i=1}^{N^0}; \quad D^m = \{d_n^m\}_{n=1}^{N^L}$$

Дельта-
правило

$$(w_{p,q}^l)_{new} = (w_{p,q}^l)_{old} - \mu \frac{\partial E}{\partial w_{p,q}^l}$$

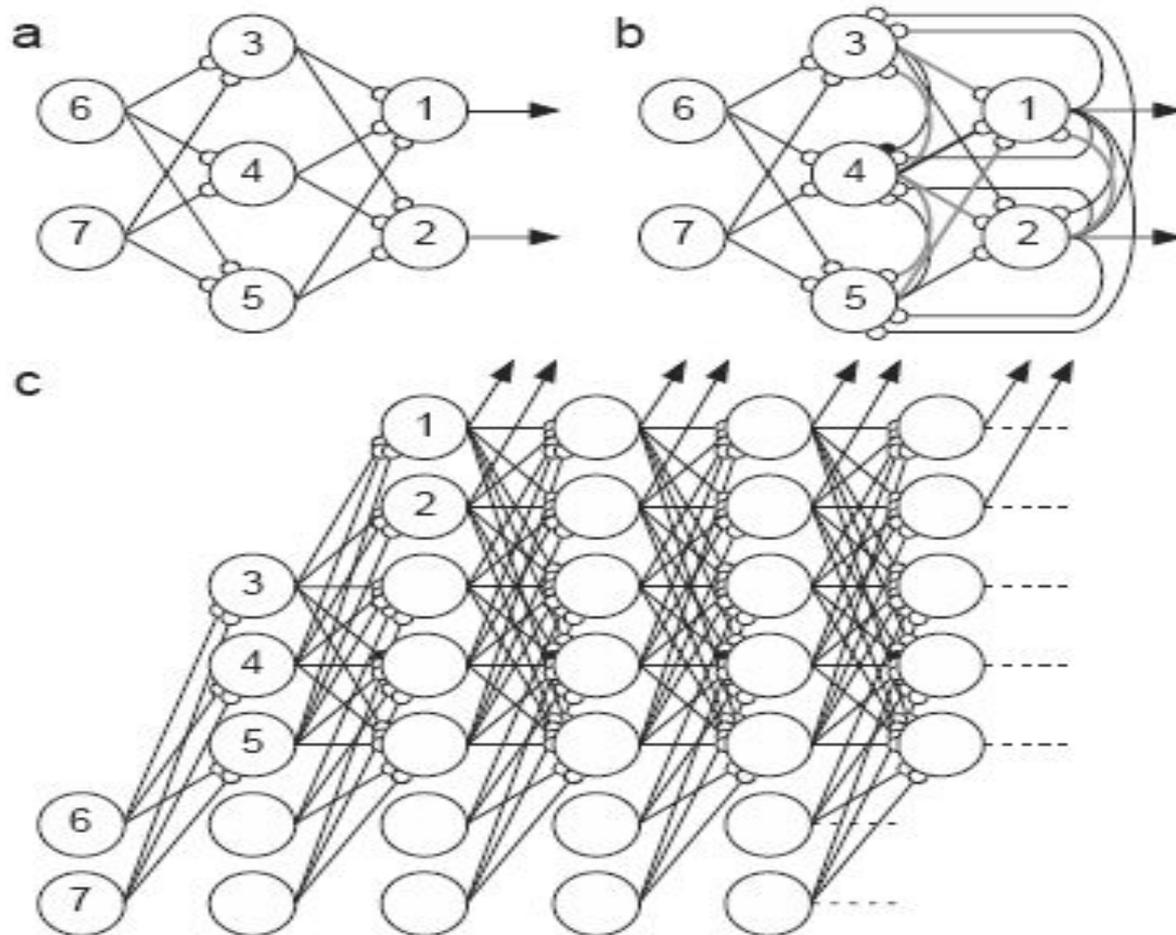
$$\frac{\partial E}{\partial w_{p,q}^l} = - \sum_{m=1}^M \sum_{n=1}^{N^L} (d_n^m - z_n^{L,m}) \frac{\partial z_n^{L,m}}{\partial z_p^{l,m}} \frac{\partial z_p^{l,m}}{\partial w_{p,q}^l}$$

$$z_p^{l,m} = \begin{cases} z_p^{l,m} & \text{для прямых связей} \\ z_p^{l,(m-r)} & \text{для обратных связей} \end{cases}$$

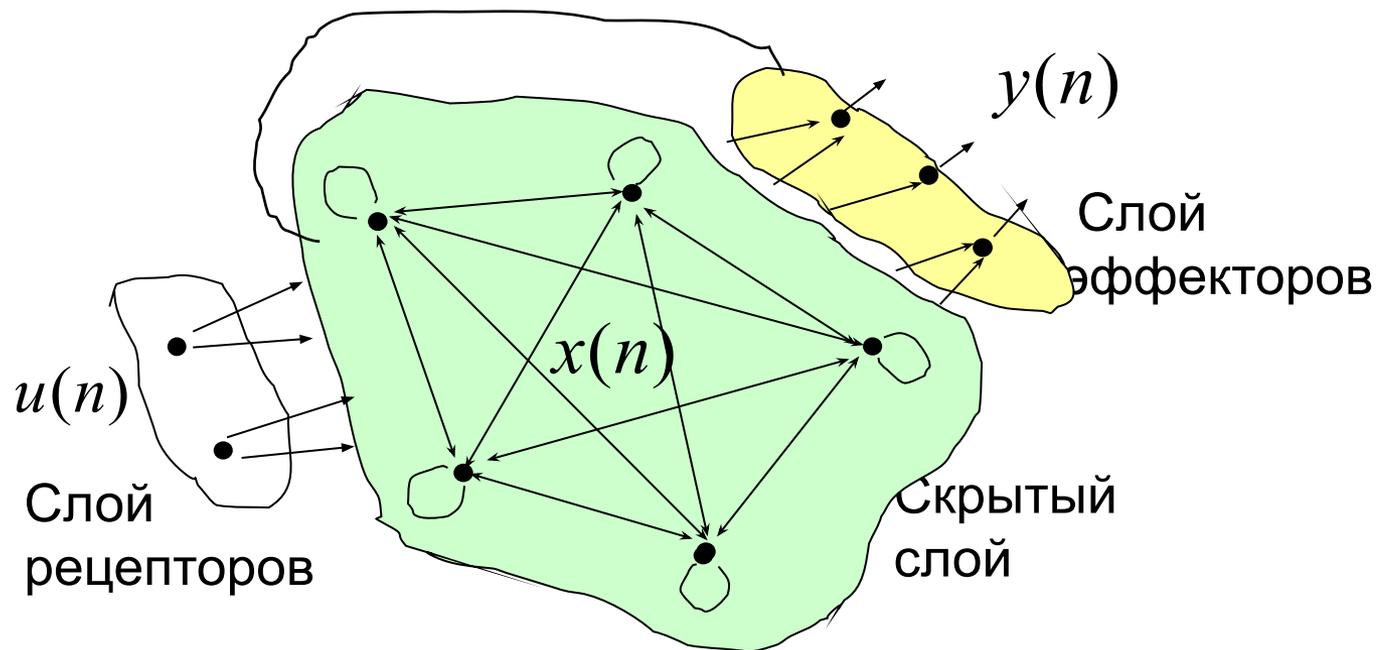
Схема коррекции веса прямых (W_{\square}) и обратных (W_{\square}) связей



Принцип обратного распространения во времени



Echo State Network

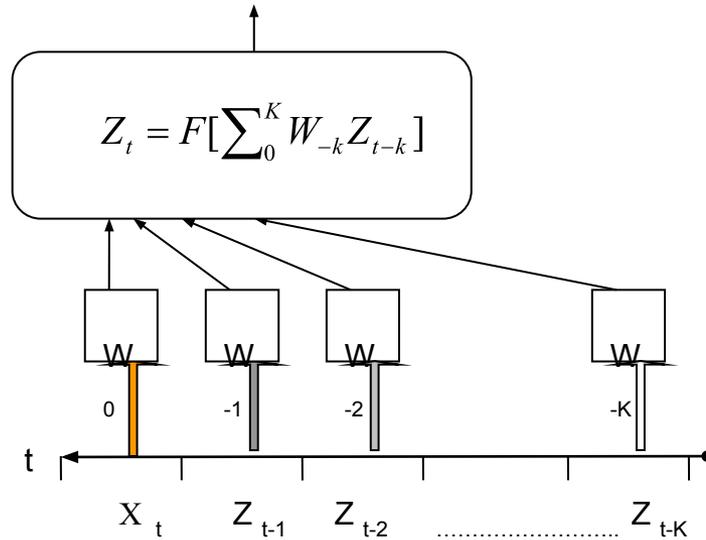


$$u(n+1) = f[Wu(n) + W^{in}x(n+1) + W^{fb}y(n)]$$

$$y(n) = g[W^{out}u(n)]$$

$$W^{out} = (U^+ D)^T$$

Виртуальная нейросеть



$$Z_t = F[\sum_1^K W_{-k} Z_{t-k} + W_0 X_t]$$

$$= L_0 W_0 \{ \sum_{k=1}^K \sum_{i=1}^K W_{-i} Z_{t-k-i} L_1 [\sum_{i=1}^K W_{-i} Z_{t-k-i} + W_0 X_{t-k}] + W_0 X_t \} =$$

$$= L_0 W_0 \{ X_t + \sum_{k=1}^K L_k W_k [X_{t-k} + \sum_{j=1}^K L_j W_j (\dots)] \}$$

Уравнение состояния сети

Нейронная сеть как динамическая система (X,Y,U) $X(t) \in \mathfrak{R}^{Inp}$

$$\frac{\partial U(t)}{\partial t} = F[U(t), X(t)] \quad Y(t) \in \mathfrak{R}^{Out}$$

$$Y(t) = \Psi[U(t)] \quad U(t) \in \mathfrak{R}^N$$

Траектория
состояния $\mathfrak{J}^f = \{U^f(t)\}_{t_0}^T \in \mathfrak{R}^{N \times T}$

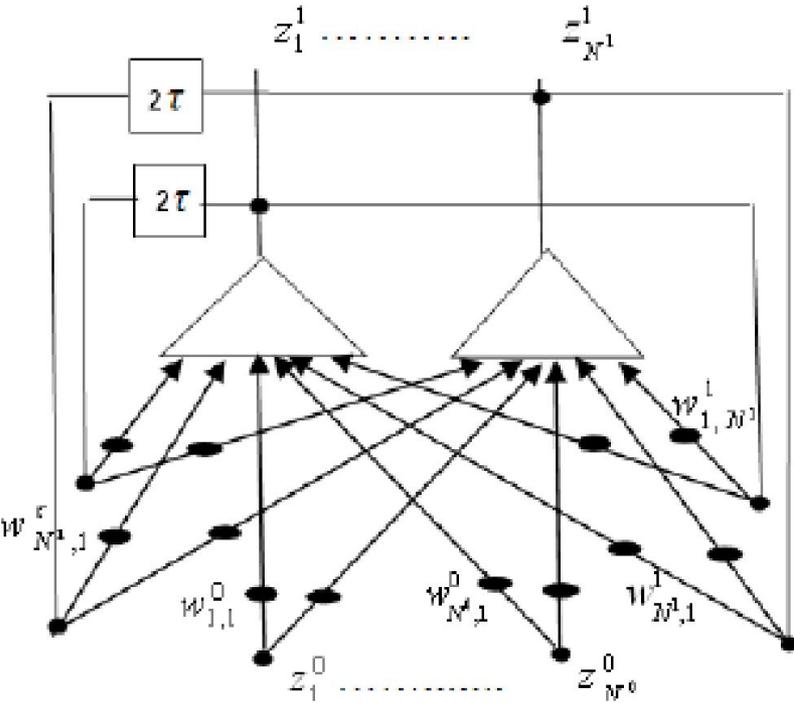
Условие
аттрактора $\frac{\partial U^f(t)}{\partial t} = F[U^f(t), X^f(t)] = 0$

Аттракторы
динамической
нейросети $\frac{\partial U_{\Sigma}(t)}{\partial t} = -AU_{\Sigma}(t) + W_{\Sigma}Z_{\Sigma}(t) = 0$

$$Z_{\Sigma}(t) = F[U_{\Sigma}(t)] = \Lambda U_{\Sigma}(t), \quad \lambda_i \geq 0,$$

$$W_{\Sigma} = A\Lambda^{-1}U^f(t)[U^f(t)]^+$$

Открытая динамическая нейронная сеть



Набор динамических аттракторов:

$$\mathfrak{R}^* = \begin{vmatrix} R^{1,1} & \dots & R^{1,m} & \dots & R^{1,M} \\ R^{1,0} & \dots & R^{1,m-1} & \dots & R^{1,M-1} \\ R^{0,1} & \dots & R^{0,m} & \dots & R^{0,M} \end{vmatrix}$$

$$Z^*(t) = \{z_n^*\}_{n=1}^{N^*}; \quad N^* = N^1 + N^\tau + N^0$$

Условие аттрактора:

$$Z^*(t) = F(S^*(t)) = \Lambda W^* Z^*(t); \quad \lambda_{ii} \geq 0$$

$$S^*(t) = \begin{vmatrix} W^{11} Z^1(t) + W^{1\tau} Z^1(t-\tau) + W^{1\tau} Z^0(t) \\ W^{\tau 1} Z^1(t) + W^{\tau\tau} Z^1(t-\tau) + W^{\tau 0} Z^0(t) \\ W^{01} Z^1(t) + W^{0\tau} Z^1(t-\tau) + W^{0\tau} Z^0(t) \end{vmatrix}$$

$$\mathfrak{R}^* = \Lambda W^* \mathfrak{R}^*$$

$$W^* = \lambda^{-1} \mathfrak{R}^* (\mathfrak{R}^*)^+$$

Псевдоинверсный алгоритм

$$w_{i,j}^{*m+1} = w_{i,j}^{*m} + (r_i^{m+1} - s_i^{m+1})(r_j^{m+1} - s_j^{m+1}) / d^{m+1}$$

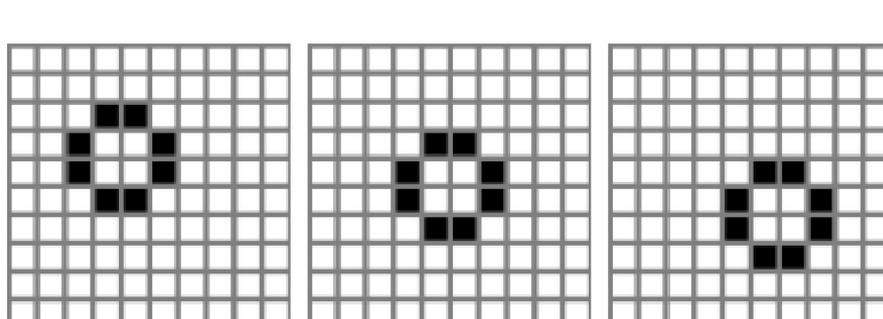
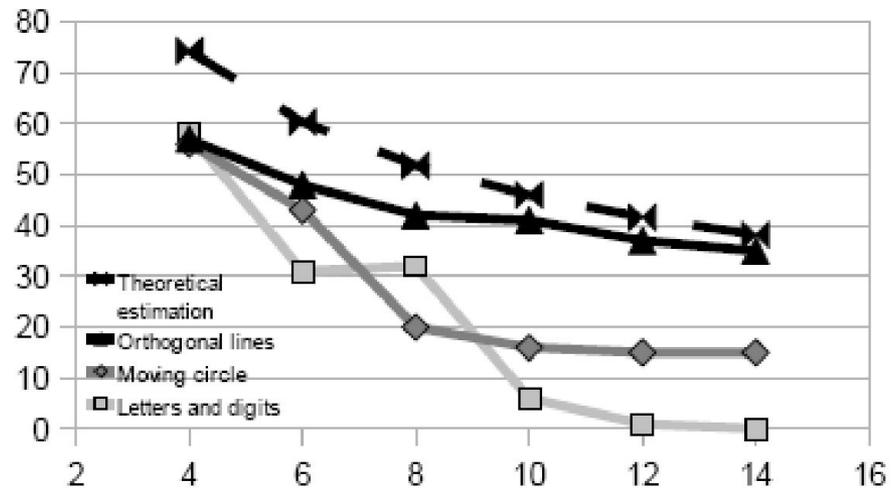
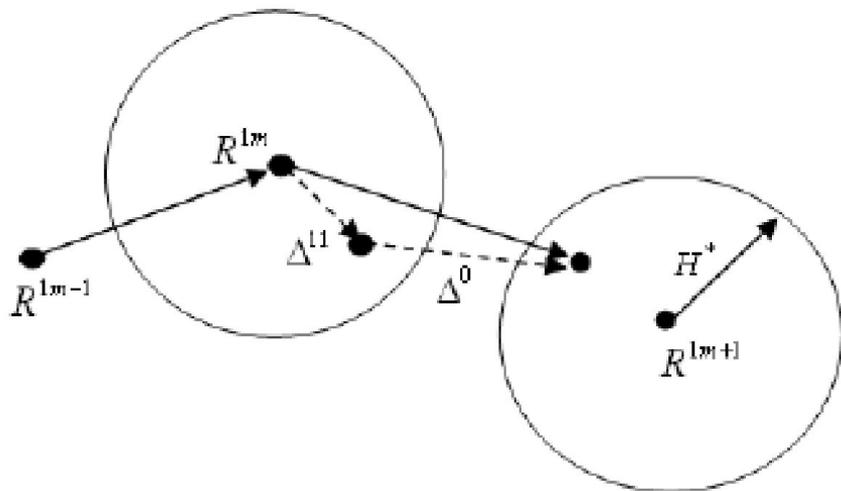
$$s_i^{m+1} = \sum_{k=1}^{N^*} w_{ik}^{*m} r_k^{m+1} \quad d^{m+1} = \sum_{k=1}^{N^*} r_{ik}^{m+1} (r_k^{m+1} - s_k^{m+1})$$

Свойства матрицы $W^* = \mathfrak{R}^* (\mathfrak{R}^*)^+$

$$W^* = (W^*)^2; \quad \text{Tr} W^* = \sum_{i=1}^{N^*} w_{i,i}^* = M; \quad w_{i,i}^* = \sum_{j=1}^{N^*} (w_{i,j}^*)^2$$

$$\overline{w_{i,i}^*} = \frac{M}{N^*}; \quad \frac{M}{N^*} \left[1 - \frac{M}{N^*} \right] \geq (w_{i,j}^*)^2 \geq \frac{M}{N^* (N^* - 1)} \left[1 - \frac{M}{N^*} \right], \quad i \neq j$$

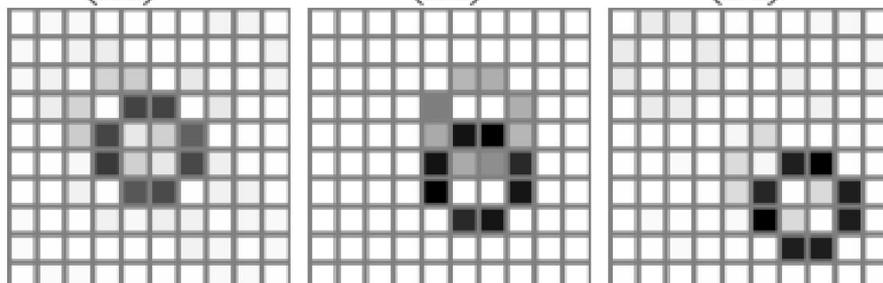
Динамические аттракторы



(i1a)

(i2a)

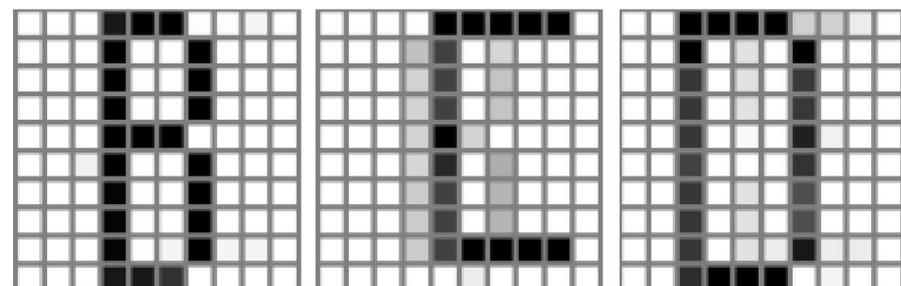
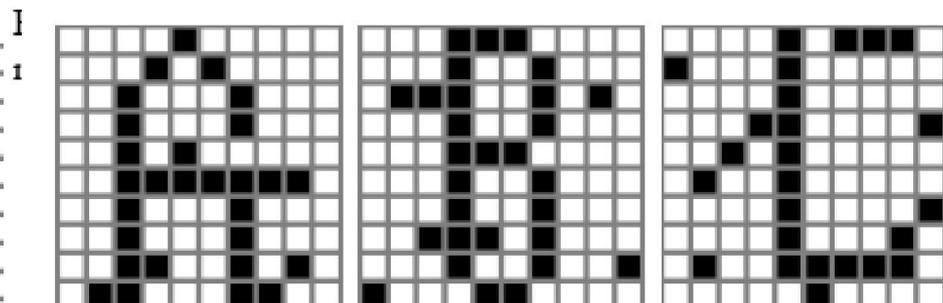
(i3a)



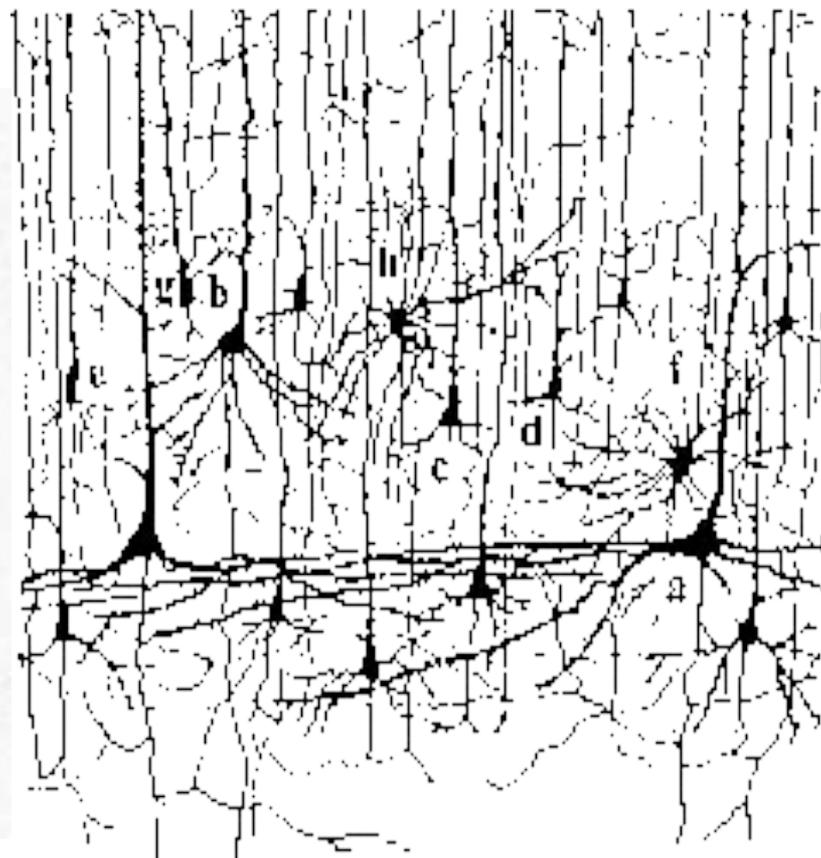
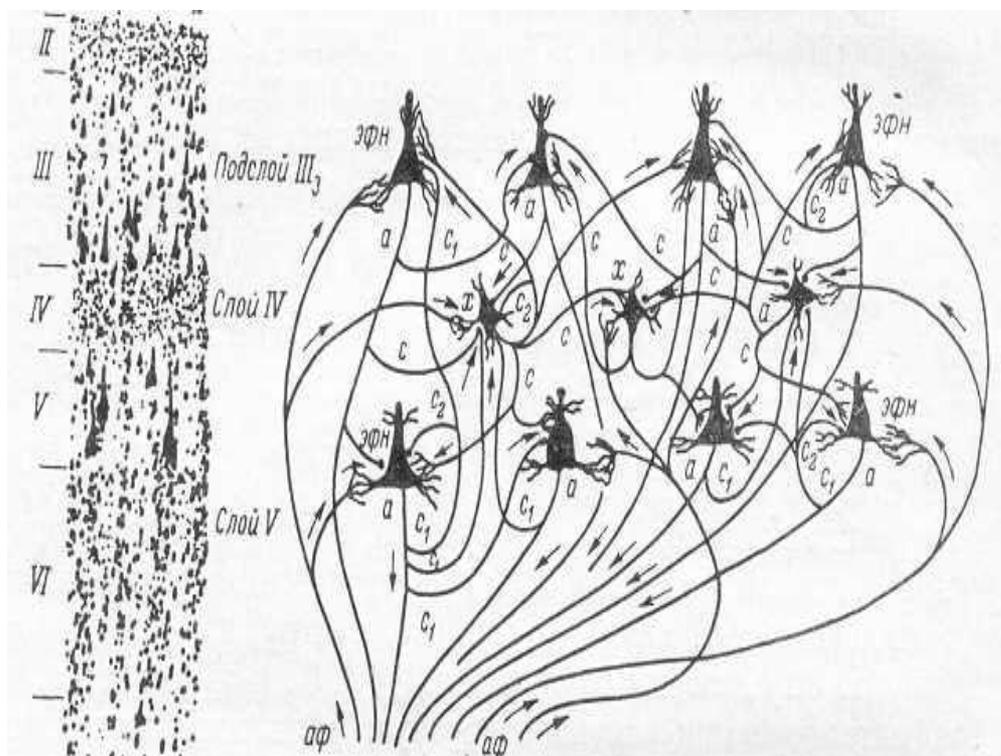
(p1a)

(p2a)

(p3a)



Фрагмент нервной ткани коры мозга



Структура синаптической матрицы слоя нейронов

