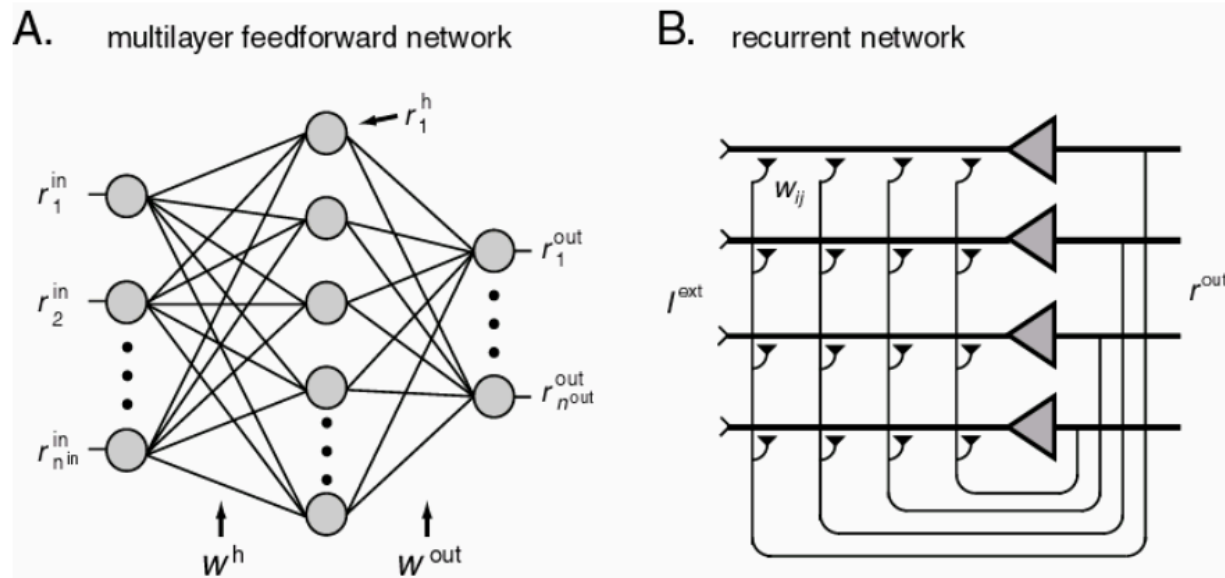


Neuronale Netzwerke: „Feed-forward“ versus „recurrent“ (d.h. „feed-back“)



A: Schrittweise vorwärts-gerichtete Abbildung:

Eingangssignal (Input) $r^{in} \rightarrow$ Ausgangssignal (Output) r^{out}

„Überwachtes Lernen“ (wie z.B. bei „deep learning“): Minimiere Differenzen zwischen r_i^{out} und erwünschten Outputs durch Veränderung der „Gewichte“ W^h and W^{out} (synaptische Kopplungen). Dazu wird das Fehlersignal vom Output zum Input zurückgerechnet („error back-propagation“).

B: Rückgekoppeltes („feed-back“) System: Output r^{out} wird in das System rückgekoppelt und mit dem externen Input I^{ext} kombiniert.

\rightarrow Dynamisches System; wird im Allgemeinen eine komplexe Dynamik haben

Erwünschte Zustände werden mit Hilfe geeignet gewählter Gewichte W_{ij} als Attraktoren der Netzwerk-Dynamik realisiert („Assoziativ-Speicher“)

Netzwerk-Dynamik mit „graded-response“ Neuronen

Die dynamischen Gleichungen des Hopfield-Modells (1984) lauten:

$$C_i \frac{du_i(t)}{dt} = -\frac{1}{R_i} u_i(t) + \sum_{j=1}^N T_{ij} V_j(t) + I_i \quad (I)$$

Annahmen:

$$C_i > 0, R_i > 0, T_{ij} = T_{ji}. V_j(t) = g_j(u_j(t)) \text{ streng monoton, } \lim_{x \rightarrow -\infty} g(x) = 0, \lim_{x \rightarrow +\infty} g(x) = 1$$

Die Zustandsfunktion $E(V_1, V_2, \dots, V_N)$,

$$E(V_1, V_2, \dots, V_N) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N T_{ij} V_i V_j + \sum_{i=1}^N \frac{1}{R_i} \int_{\frac{1}{2}}^{V_i} g_i^{-1}(V) dV - \sum_{i=1}^N I_i V_i \quad (II)$$

ist nach unten beschränkt und nimmt längs jeder Lösung monoton ab. (Beweis: nächste Folie)

E ist also eine „Lyapunov-Funktion“ (oft auch etwas salopp „Energie-Funktion“ genannt).

Das System ist also *dissipativ* und strebt für lange Zeiten gegen (lokale) Minima von E .

Dynamische Systeme, bei denen die Energie erhalten ist, werden auch *konservativ* genannt.

Ein Beispiel: Ungedämpfter harmonischer Oszillator (mit Dämpfung: dissipativ).

Netzwerk-Dynamik mit „graded-response“ Neuronen

Beweis

Allgemein gilt:

$$\frac{dE(t)}{dt} = \sum_{k=1}^N \frac{\partial}{\partial V_k} E(V_1, V_2, \dots, V_N) \frac{dV_k(t)}{dt}$$

Für (II) zudem:

$$\frac{\partial}{\partial V_k} E(V_1, V_2, \dots, V_N) = -\frac{1}{2} \sum_{j=1}^N T_{kj} V_j - \frac{1}{2} \sum_{i=1}^N T_{ik} V_i + \frac{1}{R_k} g_k^{-1}(V) - I_k$$

Mit (I) eingesetzt ergibt dies:

$$\frac{\partial}{\partial V_k} E(V_1, V_2, \dots, V_N) = -C_k \frac{du_k(t)}{dt}$$

Insgesamt also:

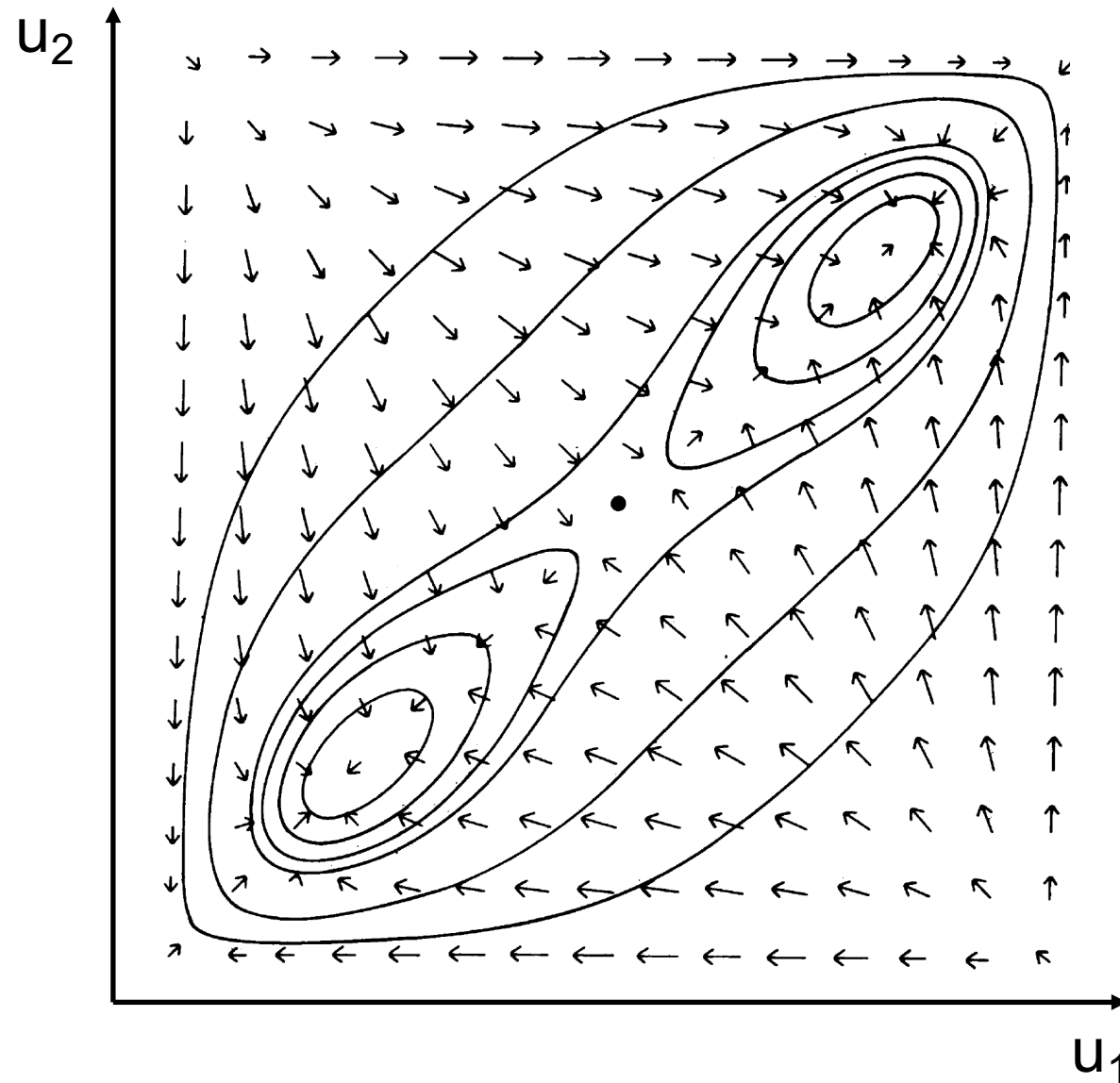
$$\frac{dE(t)}{dt} = -\sum_{k=1}^N C_k \frac{du_k(t)}{dt} \frac{dV_k(t)}{dt} \leq 0$$

die Ungleichung gilt, da $\frac{du_k(t)}{dt}$ und $\frac{dV_k(t)}{dt}$ wegen des monotonen Wachstums der g_k gleiches Vorzeichen haben (beide positiv, negativ oder null), ihr Produkt also immer positiv ist oder null.

Dass E nach unten beschränkt ist, kann man auf zwei Weisen zeigen. (a) Da $0 \leq V_k(t) \leq 1$ (siehe die Anforderungen an g_k), ist der erste und dritte Term von E im Bereich der durch die Dynamik des Systems erreichten V-Bereiche endlich und damit nach unten beschränkt, der zweite Term ist sowieso nicht-negativ. (b) Selbst wenn man die Forderung an die g_k aufweicht und erlaubt, dass g_k nicht nach oben beschränkt ist, dann dominiert der zweite Term der E-Funktion den gesamten Ausdruck für großes V, solange die g_k sublinear anwachsen. Dann nämlich wachsen die g_k^{-1} supralinear, ihr Integral also stärker als quadratisch. Der erste Term fällt jedoch maximal quadratisch, der dritte Term maximal linear, so dass der zweite Term die Beschränktheit sichert.

Netzwerk-Dynamik mit „graded-response“ Neuronen

System mit zwei Neuronen:



Phase-plane analysis

Speicherkapazität des Hopfield-Modells:

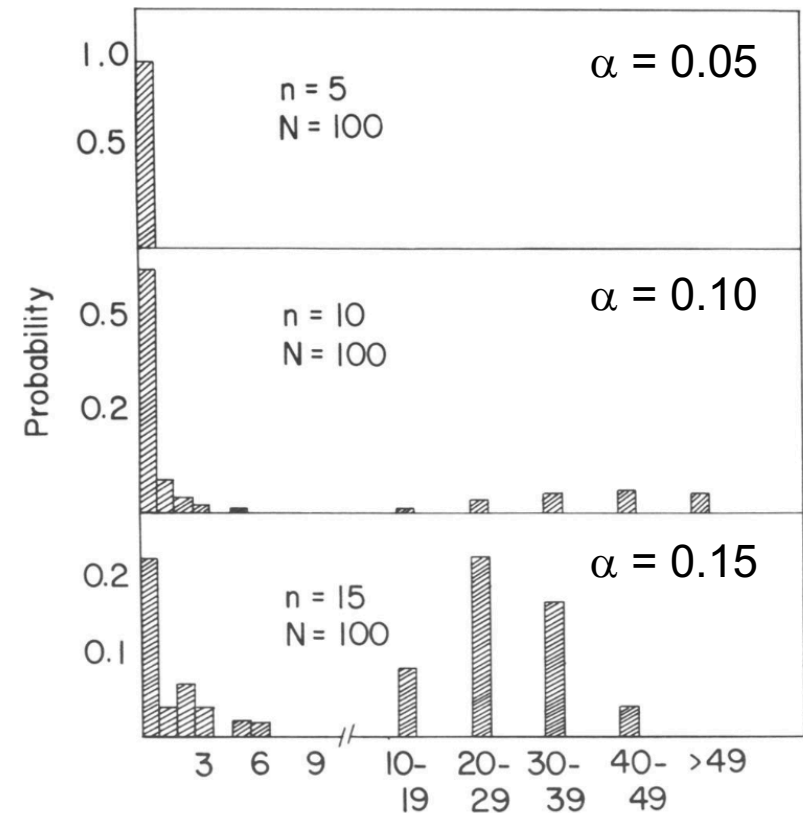
Hebb'sches Lernen (DO Hebb, 1904-85)
„fire together, wire together“:

Synaptische Kopplungen (p Muster):

$$T_{ij} = N^{-1} \sum_{k=1}^p \xi_i^k \xi_j^k$$

ξ_i^k : Aktivität von Neuron i in Muster k

Speicherkapazität (Hopfield 1982)
Computer-Simulation
für „two-state neurons“ ($S_i = +/-1$)



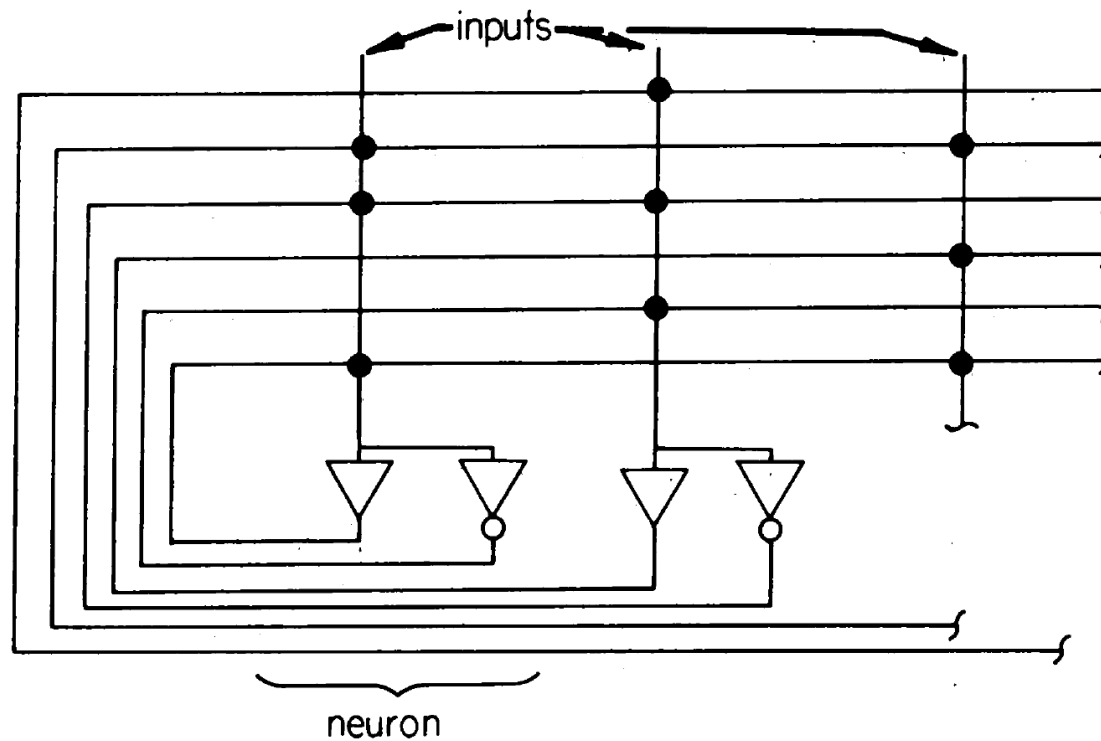
Amit et al. 1985: exakte Lösung mit Methoden der theoretischen Festkörperphysik:
p (# Muster) skaliert wie N (# Neuronen): $p = \alpha N$; kritisches $\alpha_c \sim 0.14$

$\alpha < \alpha_c$ Trajektorien der Netzwerk-Dynamik, die in der Nähe eines gespeicherten Musters ξ starten, konvergieren zu ξ : **assoziative Mustererkennung**
 $\alpha > \alpha_c$ die meisten Trajektorien erreichen keines der gespeicherten Muster:
Gedächtniskatastrophe

Elektrische Analogie

$$C_i(du_i/dt) = \sum_j T_{ij}V_j - u_i/R_i + I_i$$

$$u_i = g_i^{-1}(V_i).$$

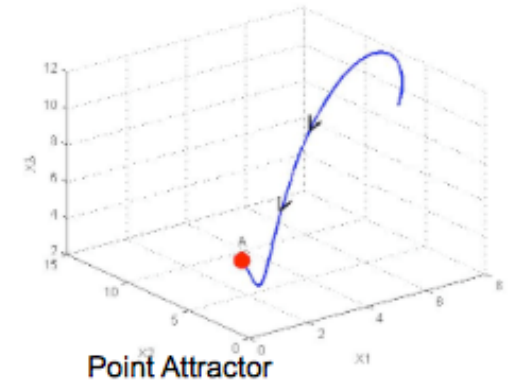


-  amplifier
-  inverting amplifier
-  resistor in T_{ij} network

Vergleich verschiedener Attraktor-Netzwerk-Typen

a) „Point attractors“

zB Assoziativ-Speicher



b) „Continuous attractors“

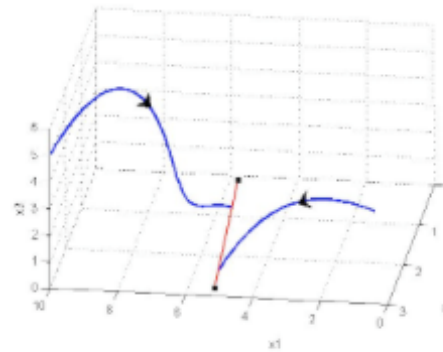
1D: Linien-Attraktoren:

zB Arbeitsgedächtnis

speziell: Ring-Attraktoren:

zB Orientierungs-Tuning im
visuellen System

Kopf-Richtungssystem

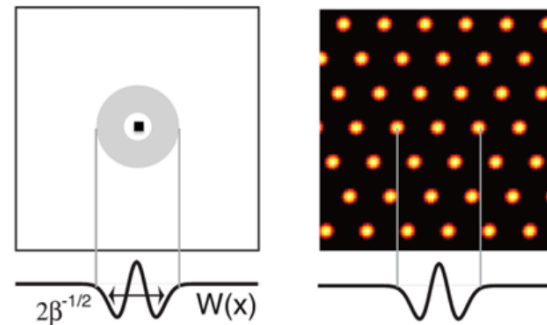
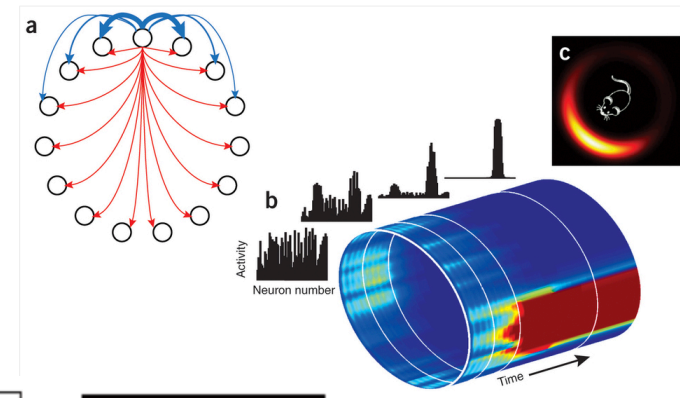


Line Attractor

2D: zB Torus-artige Attraktoren:

Gitterzell-System im

entorhinalen Cortex



c) „Time-dependent attractors“

(periodisch oder chaotisch):

z.B. Mustersequenzen

(sensorisch, motorisch)