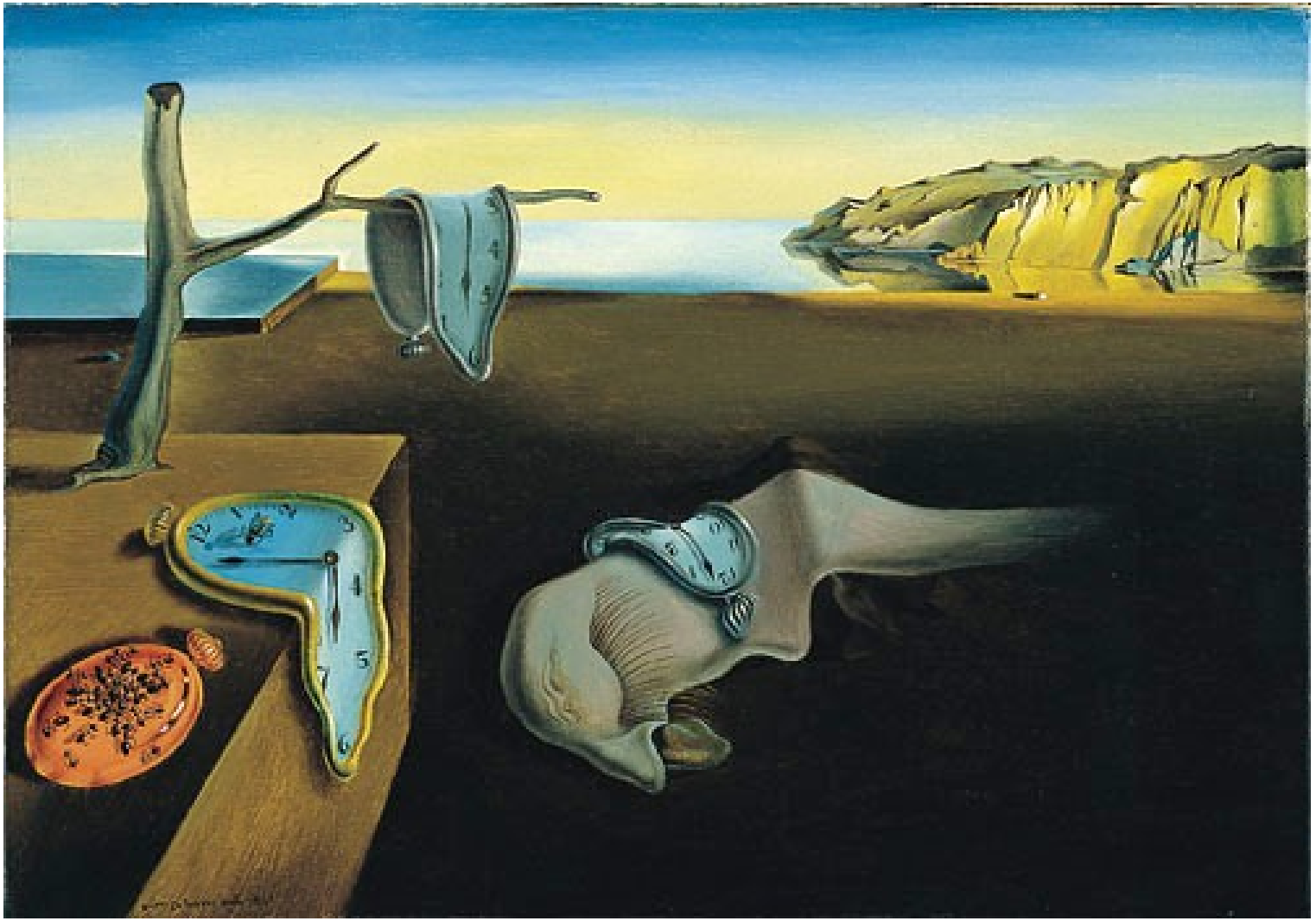
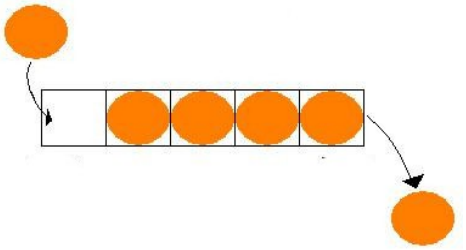
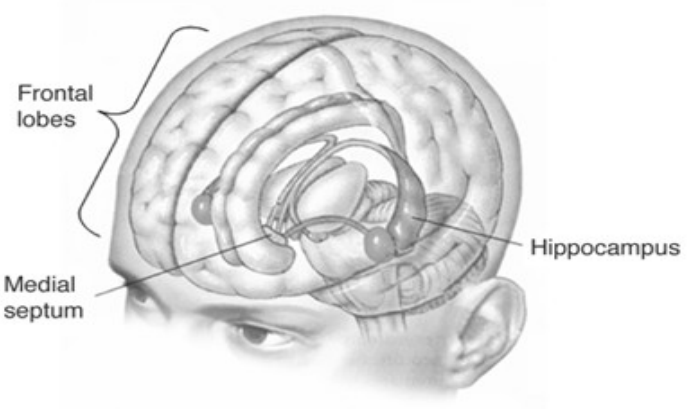
Az idegrendszeri memória modelljei

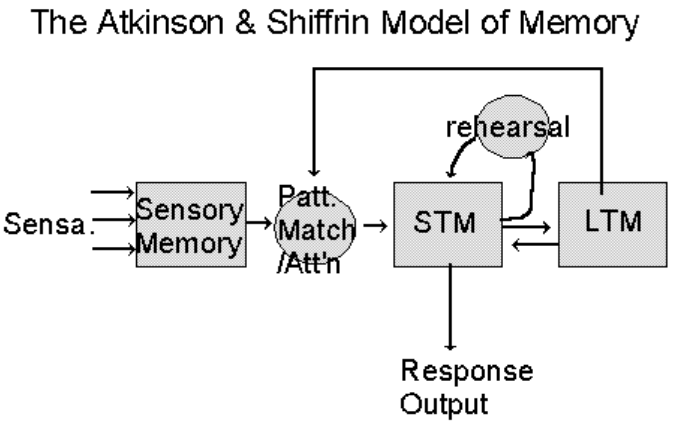


A memória típusai

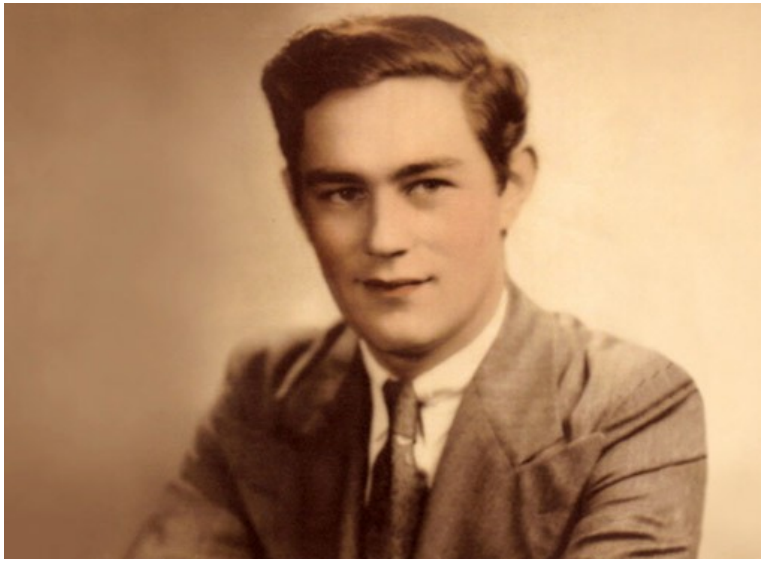
- Rövidtávú
  - Working memory - az aktuális feladat
  - Vizuális, auditórikus,...
  - Prefrontális cortex, szenzorikus területek
  - Kapacitás: 7 +-2 minta
- Hosszútávú
  - Epizodikus, szemantikus
  - Technikailag: asszociatív
  - Temporális lebeny, hippokampusz
- Interakció a rendszerek között







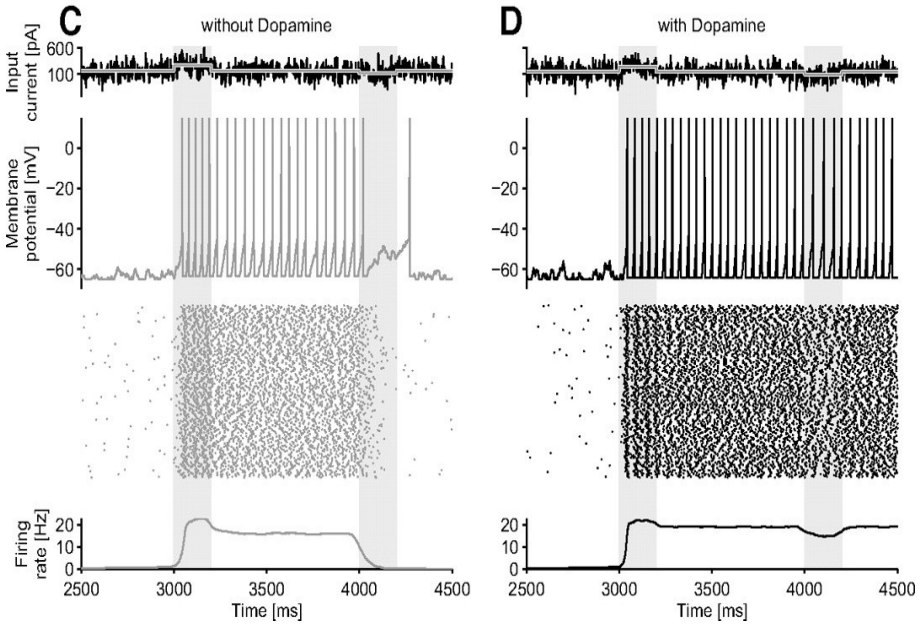
H. M.

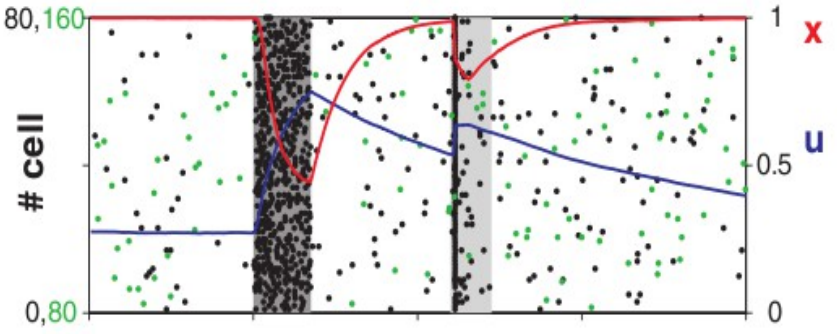


- Súlyos epilepsziája volt, amit a hippokampusz egy részének eltávolításával orvosoltak 1953-ban.
- Ettől kezdve elvesztette az epizodikus memóriaformáció képességét – a korábbi emlékei megmaradtak.
- A rövidtávú memóriája ép maradt, valamint a motoros tanulási képessége is. Megtanult pl. tükörben rajzolni.
- A térbeli memóriája erősen sérült.
- Bizonyítékot szolgáltatott a különböző memóriarendszerek létezésére.

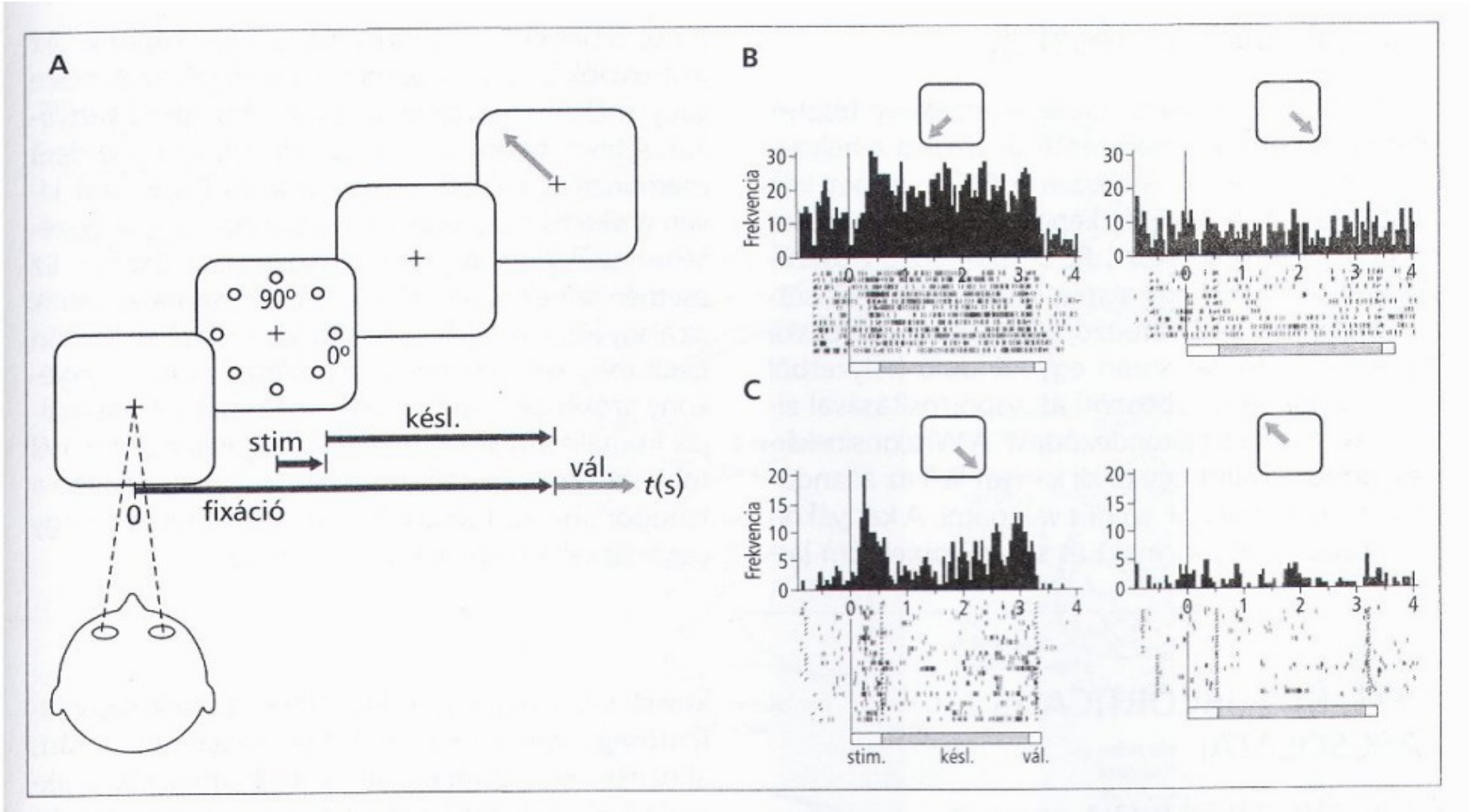
Munkamemória modelljei

- Rekurrens hálózati modellek, serkentő-gátló populációkkal
- Perzisztens aktivitás
  A kódoló populáció a jel beírása után magasabb rátával tüzel
- Előfeszített állapot
  A sejtek a beíráskor facilitált állapotba kerülnek, a kiolvasáskor szinkron tüzelés valósul meg
- Oszcillációs modell (később)
- Disztrakció: kis zavaró jelet ignorálni szeretnénk, nagyra viszont elromlik a memória





Perzisztens aktivitás



A majom prefrontális kérgében egyes sejtek megnövekedett aktivitást mutatnak bizonos stimulusok után a késleltetési szakaszban, ami meghatározza az adott választ is.

Szinaptikus modell

- Szinaptikus facilitáció és depresszió dinamikája integrate and fire neuronokban

$$\dot{u}_j(t) = \frac{U - u_j(t)}{\tau_F} + U\left[1 - u_j(t)\right]\sum_k \delta\left(t - t_k^{(j)}\right)$$

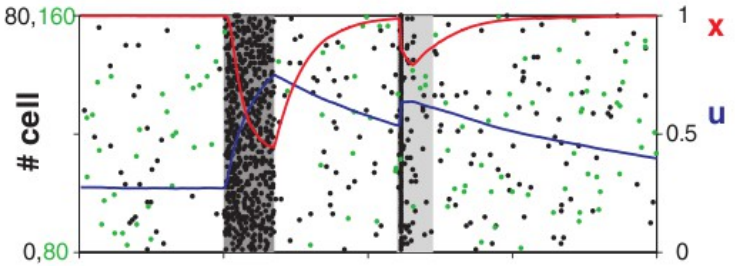
$$\dot{x}_j(t) = \frac{1 - x_j(t)}{\tau_D} + u_j(t)\,x_j(t)\sum_k \delta\left(t - t_k^{(j)}\right)$$

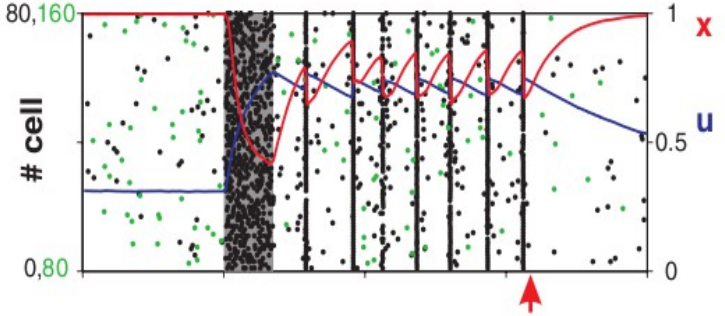
$$\tau_m \dot{V}_i = -V_i + I_i^{(rec)}(t) + I_i^{(ext)}(t)$$

$$I_i^{(rec)}(t) = \sum_j \hat{J}_{ij}(t) \sum_k \delta\left(t - t_k^{(j)} - D_{ij}\right)$$

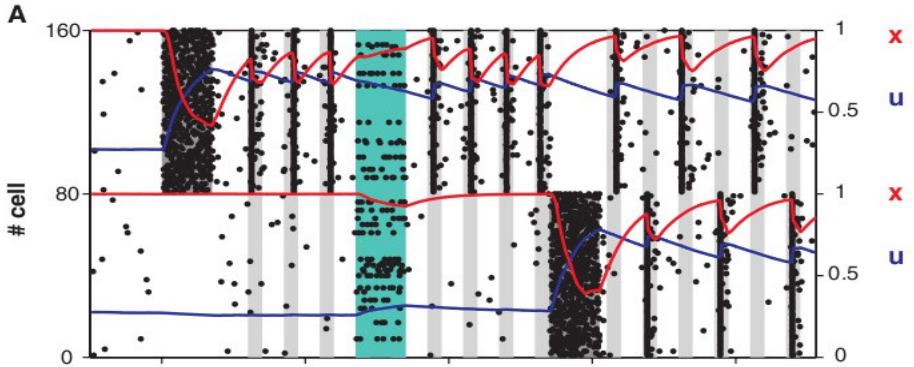
$$\hat{J}_{ij}(t) = J_{ij} \cdot u_j\left(t - D_{ij}\right) \cdot x_j\left(t - D_{ij}\right)$$

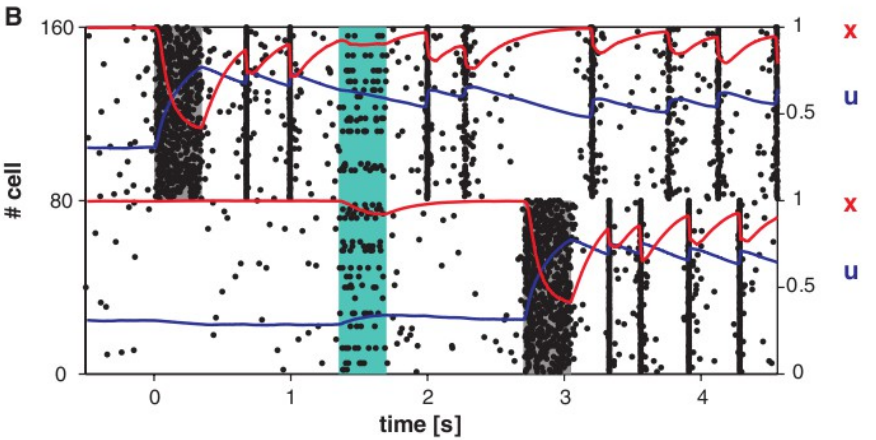
- Fixpont vagy oszcillációs dinamika





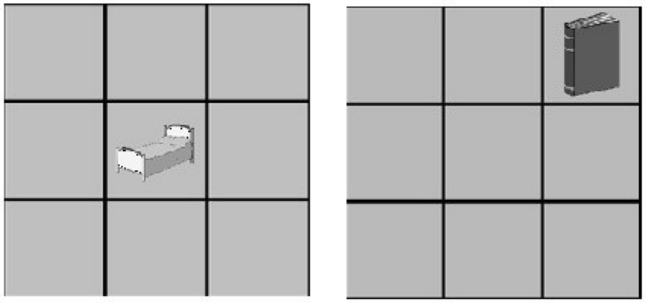
- Több elem tárolása

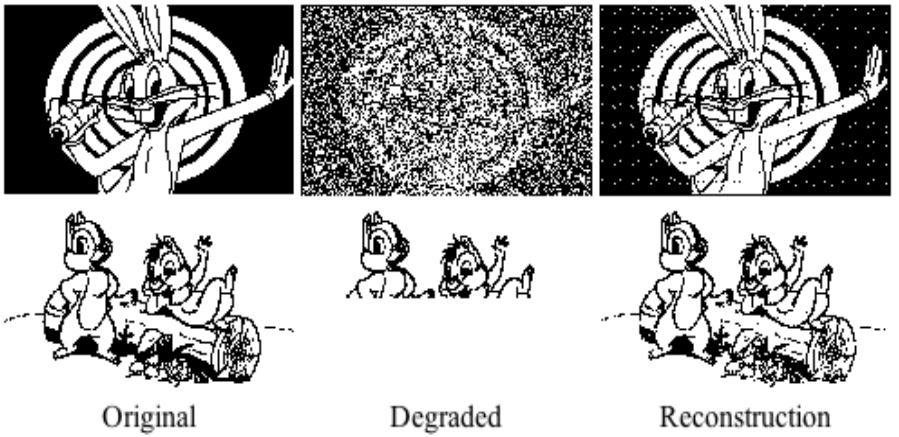




Asszociatív memória

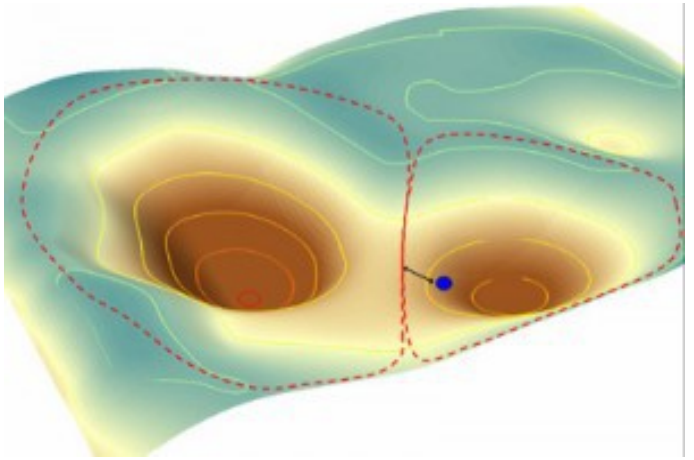
- Heteroasszociatív
  - pl. hely-objektum
- Autoasszociatív
  - Töredékes jelből az eredetit
- Különbség a számítógép memóriája és az AM között: címzés módja





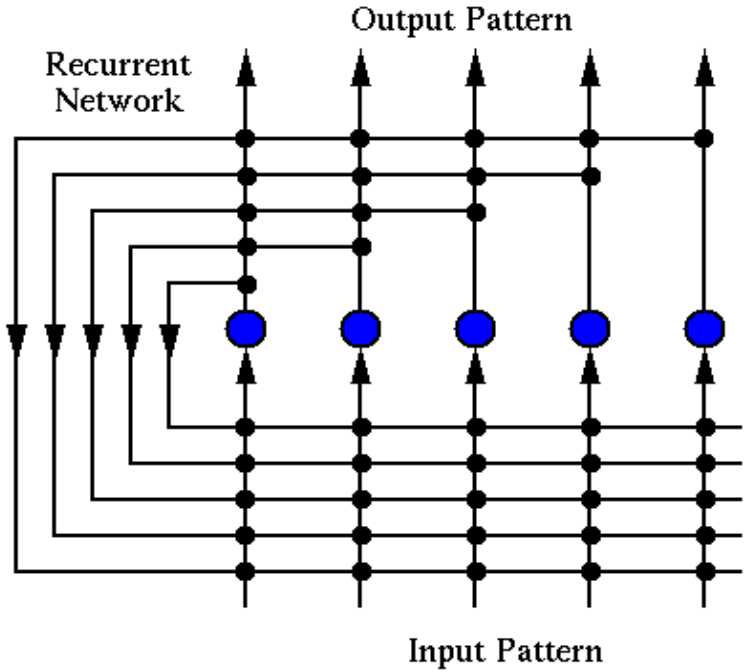
- Kapacitás: hány mintát tudunk eltárolni úgy, hogy azok visszahívhatók legyenek (többféle definíció)
- Stabilitás: minden mintára a legközelebbi tárolt mintát szeretnénk visszakapni

Attraktorhálózatok



- Attraktorok típusai
  - Pont
  - Periodikus
  - Kaotikus
- Vonzási tartományok
- Realizáció: rekurrens neurális hálózatok
- Attraktorok tárolása: szinaptikus súlyokon
  - Offline tanulás
  - Online tanulás
  - One-shot learning
- Előhívás: konvergencia tetszőleges pontból egy fix pontba

Hopfield-hálózat



- Asszociatív memória
- Bináris MCP-neuronok
- Minták tárolása: bináris vektorok
- Szimmetrikus súlymátrix
- Dale's law: egy sejt nem lehet egyszerre serkentő és gátló – ezt most megsértjük
- Rekurrens (dominánsan) hálózatok az agyban: hippokampusz CA3 régió, ...

- Offline learning tanulandó minták: $\{s^1 \dots s^N\}$ szabály

$$W_{ij} = \frac{1}{N} \sum_n^N s_i^n s_j^n \quad \leftarrow$$

Hebbi

- Léptetési szabályok: szinkron és szekvenciális

$$\mathbf{x}^{t+1} = sgn(\mathbf{W}\mathbf{x}^t - \theta) \qquad x_k^{t+1} = sgn\left(\sum_i^K W_{ik} x_i^t - \theta_k\right)$$

A HN dinamikája

- Nemlineáris rendszerek stabilitás-analízise: Lyapunov-függvény segítségével definiáljuk az állapotokhoz rendelhető energiát. Ha a függvény:
  - Korlátos
  - Belátható, hogy a léptetési dinamika mindig csökkenti (növeli)

  Akkor a rendszer minden bemenetre stabil fix pontba konvergál.
- Hopfield-hálózat Lyapunov-függvénye:

$$E = -\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} - \boldsymbol{\theta}\,\mathbf{x}$$

- Attraktorok az eltárolt mintáknál, de más helyeken is
- A HN használható kvadratikus alakra hozható problémák optimalizációjára is

A HN kapacitása

- Információelméleti kapacitás
  - A tárolandó mintákat tekintsük Bernoulli-eloszlású változók halmazának

$$P(s_i^n = 1) = P(s_i^n = 0) = 0.5$$

  - Követeljük meg az egy valószínűségű konvergenciát

$$\lim_{n\to\infty} P(s^a = sgn(\mathbf{W}\mathbf{s}^\mathbf{a})) = 1 \quad \forall\, a = 1 \ldots M$$

  - Ekkor (sok közelítő lépéssel) megmutatható, hogy

$$M \approx \frac{N}{2\log_2 N}$$

- Összehasonlítás a CA3-mal
  - Kb. 200000 sejt, kb. 6000 minta tárolható
- Más becslések
  - figyelembevéve a minták ritkaságát

$$P(s_i^n = 1) = \alpha$$

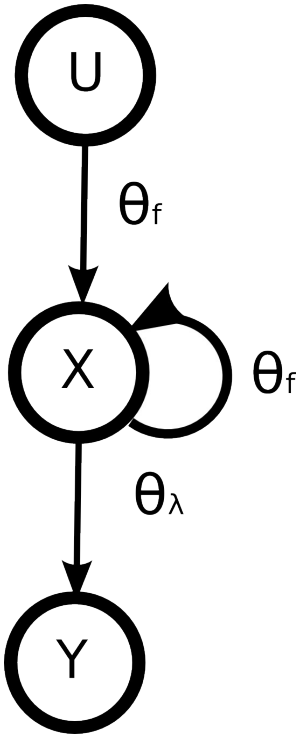
$$M \approx N \frac{1}{\alpha \log_2 \frac{1}{\alpha}}$$

Reprezentációs tanulás

- Valószínűségi leírás
- 3féle dolgot tanulhatunk: csak predikció, kimenetek valószínűsége, underlying rejtett változók/dinamika
- Explicit rejtett változós modellek
- Implicit rejtett változós modellek
- Modellösszehasonlítás
- Becslési algoritmusok: EM

Rejtett változós modellek

$$\dot{x} = f(x, u, \theta_u) + \epsilon \quad \epsilon = P(0, \Sigma_\epsilon)$$
$$y = g(x, \theta_x) + \nu \quad \nu = P(0, \Sigma_\nu)$$



posterior

likelihood

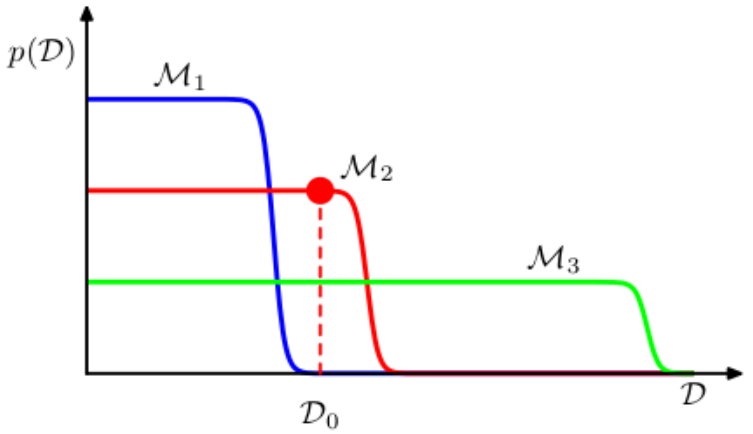
prior

$$p(\theta|y, M) = \frac{p(y|\theta, M)\, p(\theta|M)}{p(y|M)}$$

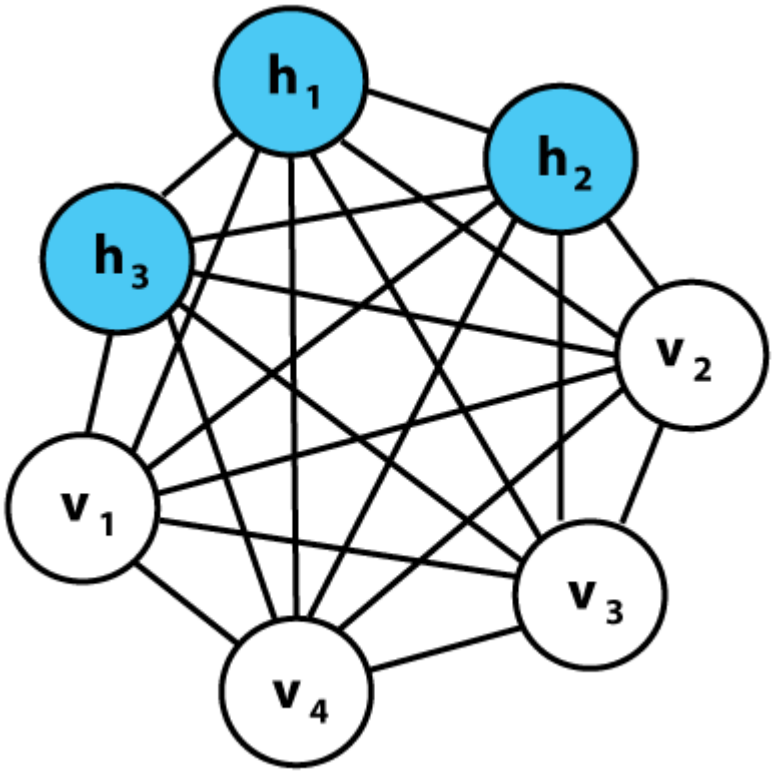
Evidence
(marginal likelihood)

Predictive distribution:

$$p(y'|\theta, y, M)$$



A Boltzmann-gép architektúrája



$$P(s_i = 1) = \frac{1}{1 + e^{-b_i - \sum_j s_j w_{ij}}}$$

$$P(\boldsymbol{v}, \boldsymbol{h}) = \frac{e^{-E(\boldsymbol{v}, \boldsymbol{h})}}{\sum_{\boldsymbol{v}', \boldsymbol{h}'} e^{-E(\boldsymbol{v}', \boldsymbol{h}')}}$$

$$P(\boldsymbol{v}) = \frac{\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v}, \boldsymbol{h})}}{\sum_{\boldsymbol{v}', \boldsymbol{h}'} e^{-E(\boldsymbol{v}', \boldsymbol{h}')}}$$

$$-E(\boldsymbol{v}, \boldsymbol{h}) = \sum_i s_i^v b_i + \sum_{j>i} s_i^v s_j^v w_{ij} + \sum_k s_k^h b_k + \sum_{l>k} s_k^h s_l^h w_{kl} + \sum_{m>n} s_m^v s_n^h w_{mn}$$

Geoffrey Hinton, Terry Sejnowski

Mintavételezés

- A normalizációs tagok kiszámolása exponenciális komplexitású → Markov Chain Monte Carlo mintavételezés
- Elindítjuk a gépet véletlenszerű állapotból, és megvárjuk, hogy beálljon a hőmérsékleti egyensúly
- Mintavételezés csak a rejtett egységekből: a látható egységeket az adatvektorhoz rögzítjük
  - A rejtett egységek az adatvektor “magyarázatát” adják, a jobb magyarázatokhoz alacsonyabb energia tartozik

Tanulás Bolzmann-géppel

- Maximum Likelihood tanulás

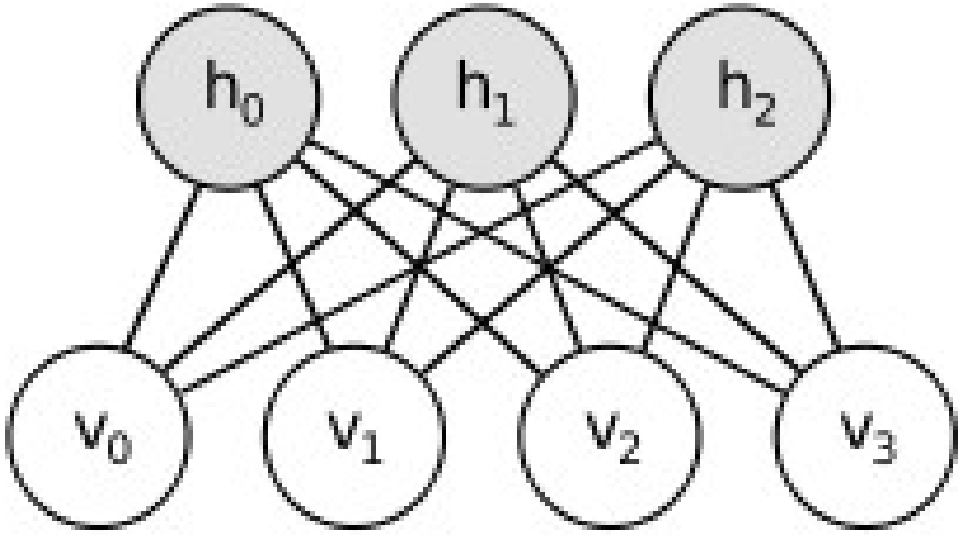
$$argmax_W \; p(V|W) = \prod_{\boldsymbol{v} \in V} p(\boldsymbol{v}|W) = \sum_{\boldsymbol{v} \in V} \log p(\boldsymbol{v}|W)$$

- Gradiens-módszer

$$\frac{\partial \, p(\boldsymbol{v})}{\partial \, w_{ij}} = \langle s_i \, s_j \rangle_{\boldsymbol{v}} - \langle s_i \, s_j \rangle_{rand}$$

- Pozitív fázis
  - A látható egységeket fixálva várjuk meg az egyensúlyt, minden tanulóvektorra átlagoljuk a statisztikát
  - A Boltzmann-valószínűség számlálóját növeli
- Negatív fázis
  - Véletlen kiindulópontból várjuk meg az egyensúlyt jó sokszor, aztán átlagoljunk
  - A Boltzmann-valószínűség nevezőjét csökkenti

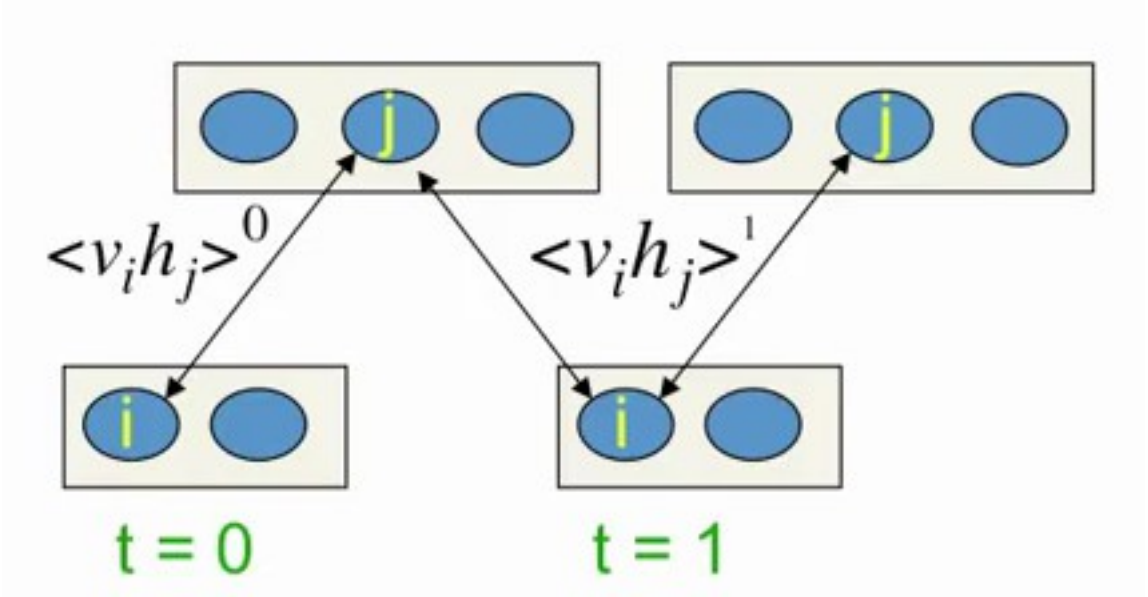
Restricted Boltzmann Machine

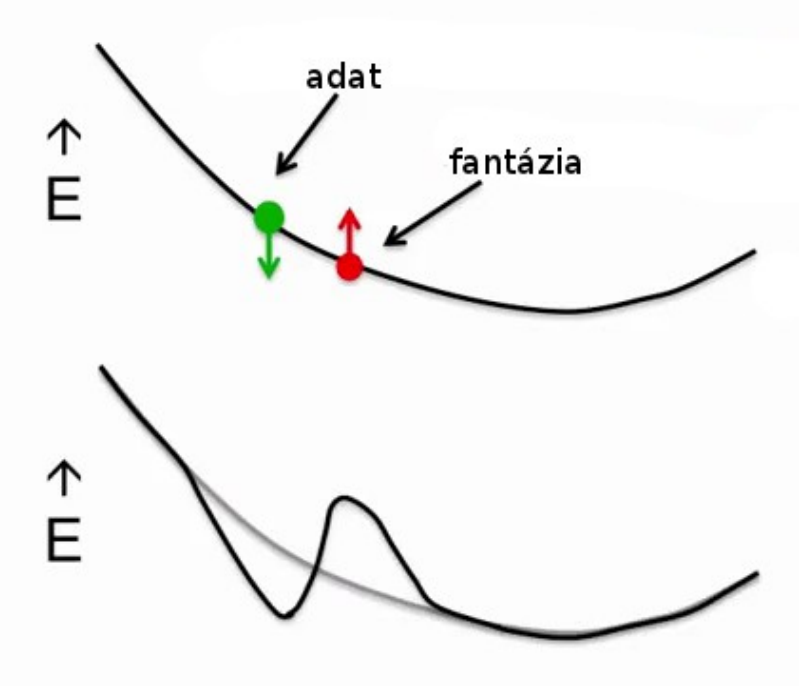


- Egy látható és egy rejtett réteg
- Rétegen belül nincsenek kacsolatok → független rejtett egységek

- A rejtett rétegben egylépéssel elérjük az egyensúlyt
- A negatív statisztikához indítsuk a hálózatot ebből az állapotból

Contrastive divergence



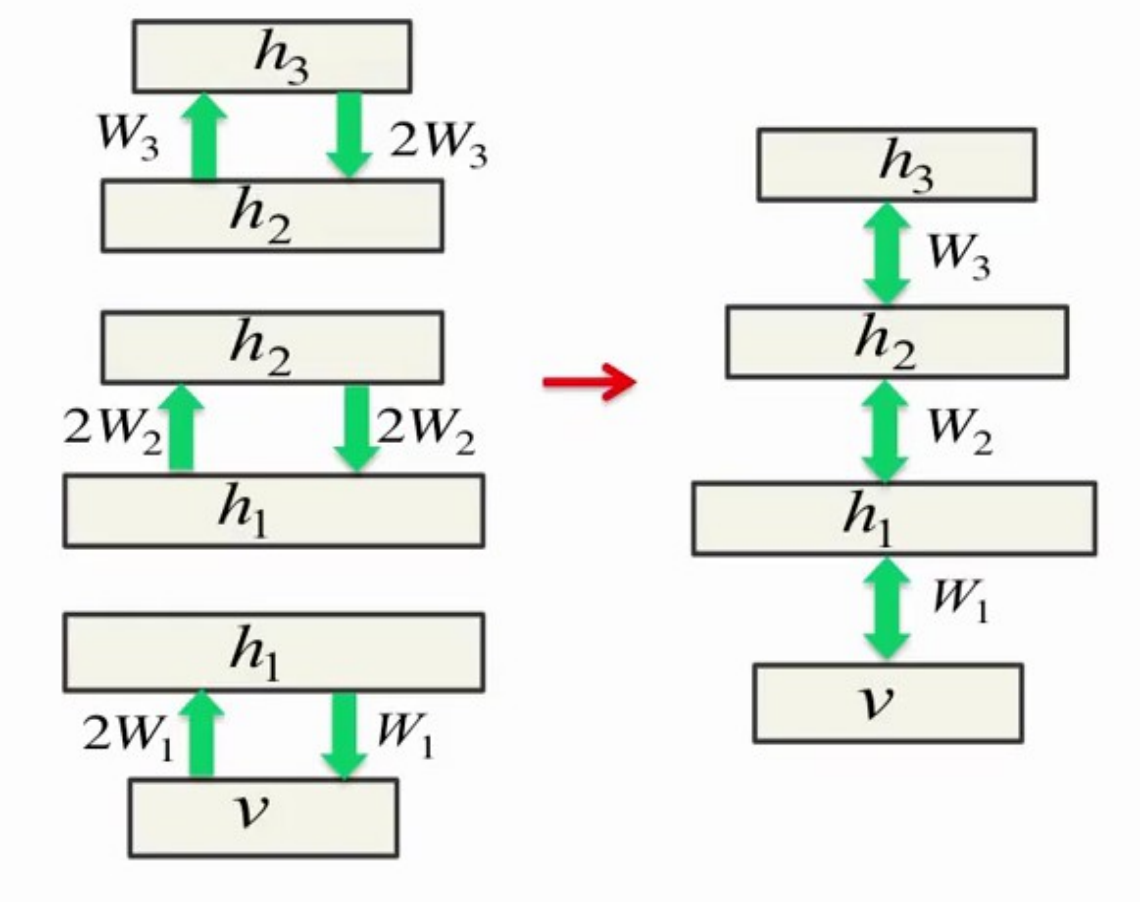


$$\Delta w_{ij} \sim \langle s_i s_j \rangle_0 - \langle s_i s_j \rangle_1$$

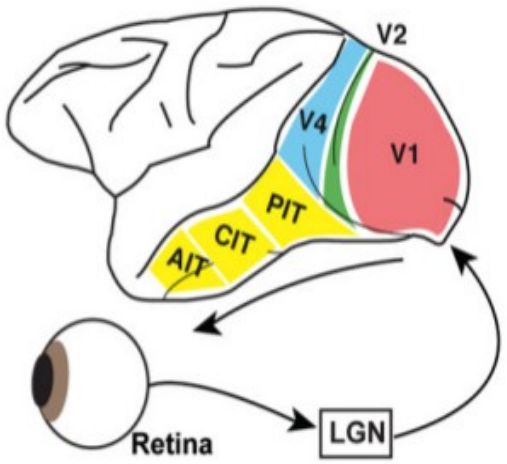
- Az adattól távolabbi minimumokat nem látja
- Miután valamennyire rátanult a hálózat az adatra, többlépéses CD-re térünk át: CD3, CD6, ...

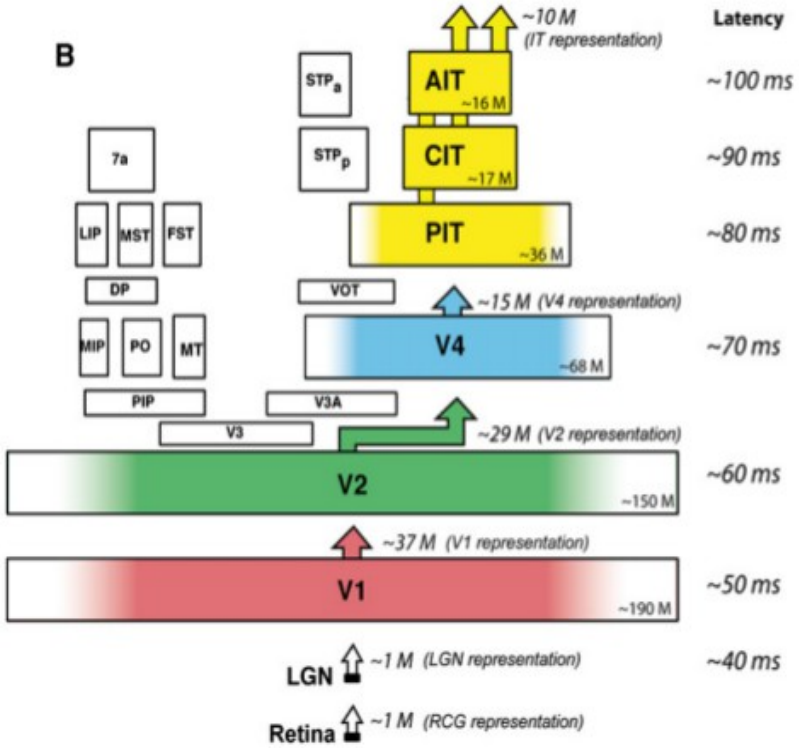
DBM létrehozása előre tanított RBM-ekből

- Az első RBM a bemenet feature-eit tanulja
- A második a feature-ök feature-eit
- Öszzeillesztjük a rétegeket, és együtt finomítjuk a tanult reprezentációt

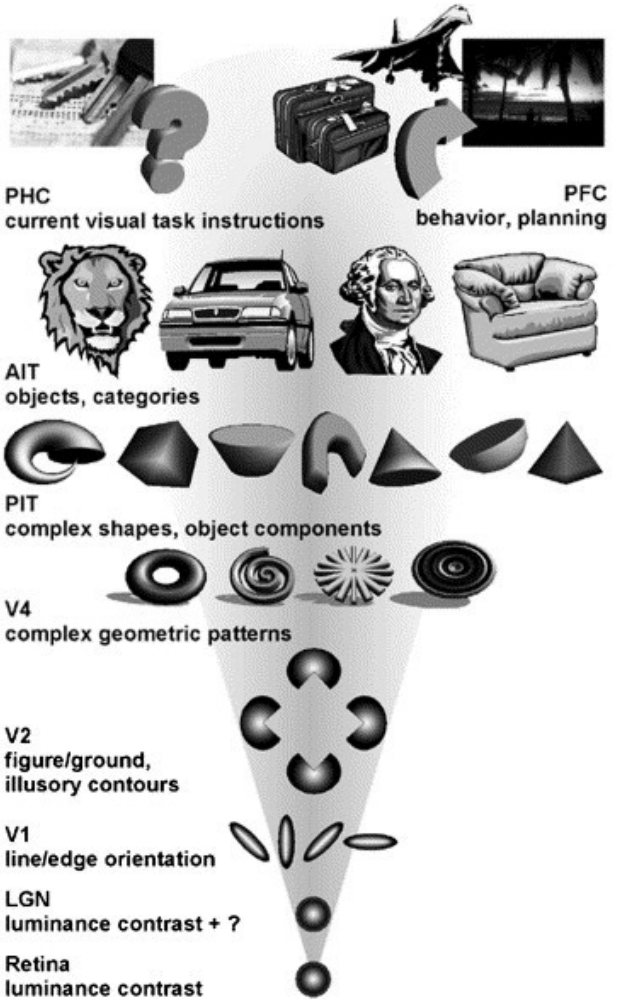


Absztrakciós hierarchia a látórendszerben





James DiCarlo



Rufin VanRullen