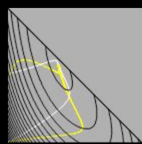


Helmholtz Machines

Memory • Fluctuation • Dreams

Academy of Media Arts, Cologne
June 28, 2006

Kevin G. Kirby
Evan and Lindsay Stein Professor of Biocomputing
Department of Computer Science, Northern Kentucky University

DEPARTMENT OF COMPUTER SCIENCE

Outline

- Machines
  Going minimalist in computer science
- Memory
  From associations to neural connections
- Fluctuation
  Surprise, energy, temperature, entropy
- Dreams
  Minimizing surprise by dreaming a world

Helmholtz I

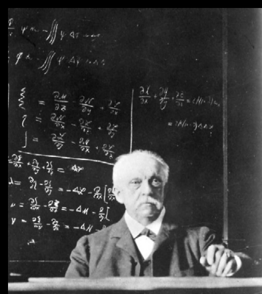
"Moreover, the visual image for Helmholtz is a sign, no passive copy of external things such as a photographic image, but rather a symbolic representation constructed by the mind to facilitate our physical interaction with things.

[...] Helmholtz sought to establish a psychological principle capable of guiding eye movements consistent with an empiricist approach to vision as learned behavior constantly in need of some self-corrective learning procedures to adapt it to the exigencies of visual practice."

- **"The Eye as Mathematician**: Clinical Practice, Instrumentation, and Helmholtz's Construction of an Empiricist Theory of Vision."

T. Lenoir, in *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science*, D. Cahan (Ed.) University of California Press, 1994.

Helmholtz II

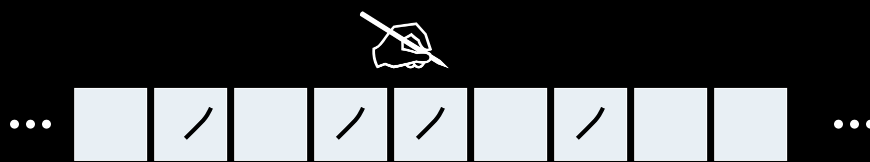


Average Energy
− Temperature × Entropy

Free Energy

Machines

Turing Machine 1936

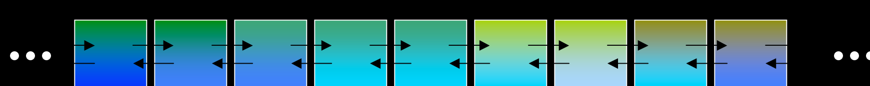


STATE

0 1 2 3 4 5 6 . . . 362 363 364

.
.
.
state 23, blank → state 42, erase, left
state 23, mark → state 53, mark, left
state 24, blank → state 18, mark, right
state 24, mark → state 93, mark, left
.
.
.

Turing Machine 1952



rate of local change of M1 = combination of M1 and M2 + flux from boundaries
rate of local change of M2 = combination of M1 and M2 + flux from boundaries

DIFFUSION CAN TRIGGER MORPHOGENESIS

Machines and Minimal Models

1936 Turing Machine model of computation
- Used to characterize effective algorithmic processing
- Allows a universal machine (one that can simulate any other TM)
- Allows simple proofs of the limits of computing (Rice's Theorem)

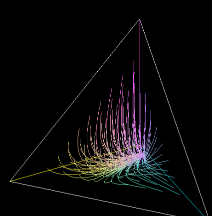
Tonight,
for this and other Machines-Named-After-People:
Let us regard them as minimal models to understand kinds of computation, not as engineered algorithms.

"The Boltzmann Machine is ~~painfully~~ slow."
^
exquisitely

(machines as objects to catalyze new thinking)

Themes / Current Interests

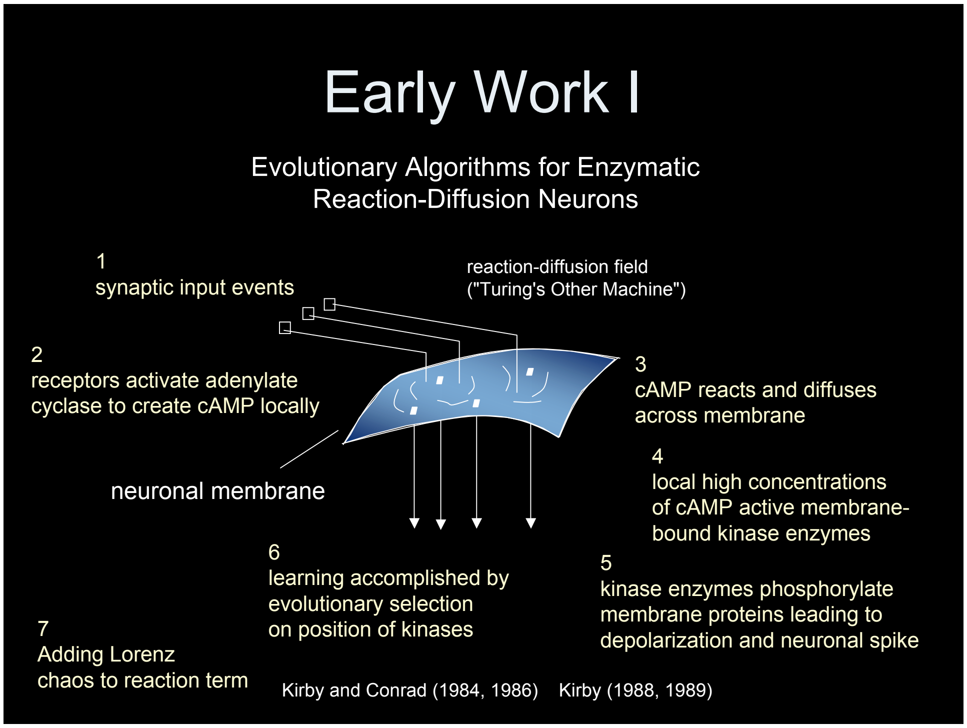
Limits and hermeneutics of natural and artificial computing

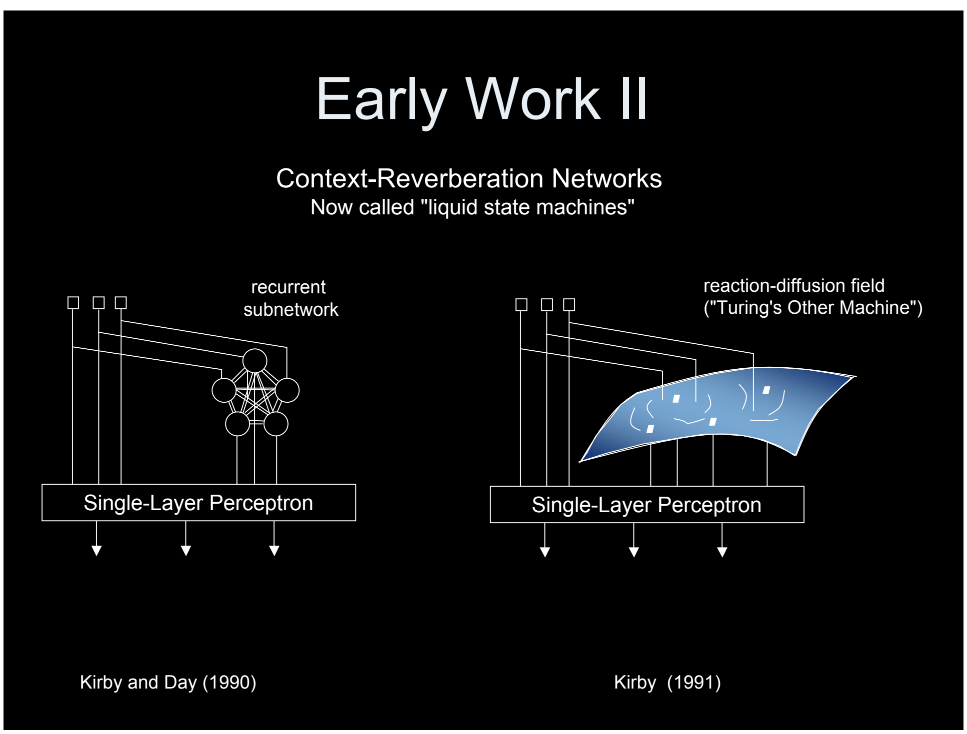


Applied information geometry and belief dynamics

λ

ALT computer science interdisciplinary pedagogy


Early Work I
Evolutionary Algorithms for Enzymatic
Reaction-Diffusion Neurons
1
synaptic input events
reaction-diffusion field
("Turing's Other Machine")
2
receptors activate adenylate
cyclase to create cAMP locally
3
cAMP reacts and diffuses
across membrane
neuronal membrane
4
local high concentrations
of cAMP active membrane-
bound kinase enzymes
6
learning accomplished by
evolutionary selection
on position of kinases
5
kinase enzymes phosphorylate
membrane proteins leading to
depolarization and neuronal spike
7
Adding Lorenz
chaos to reaction term
Kirby and Conrad (1984, 1986)
Kirby (1988, 1989)


Early Work II
Context-Reverberation Networks
Now called "liquid state machines"
recurrent
subnetwork
reaction-diffusion field
("Turing's Other Machine")
Single-Layer Perceptron
Single-Layer Perceptron
Kirby and Day (1990)
Kirby (1991)

Early Work III

Hermeneutics, the breakdown of simulation, and the "Putnam Theorem*"

* A rock can simulate any finite automaton (Hilary Putnam, 1988).

$$\begin{tikzcd} Q' \arrow[r, "\delta'"'] \arrow[d, "\phi"'] & Q' \arrow[d, "\phi"] \\ Q \arrow[r, "\delta"'] & Q \end{tikzcd}$$

"the meaning bath"

Kirby (1991, 1995)

Memory

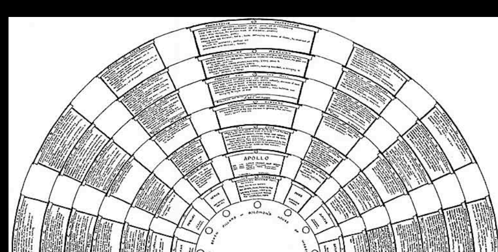
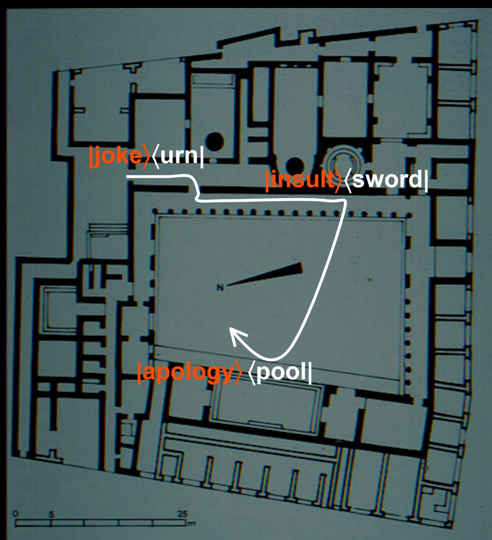


Image ↔ Place



Memory
as <u>accumulation</u> of associations:

$$M = |\text{joke}\rangle\langle\text{urn}| + |\text{insult}\rangle\langle\text{sword}| + |\text{apology}\rangle\langle\text{pool}| + \ldots$$

Quintilian: Memory buildings should be as "spacious and varied as possible"

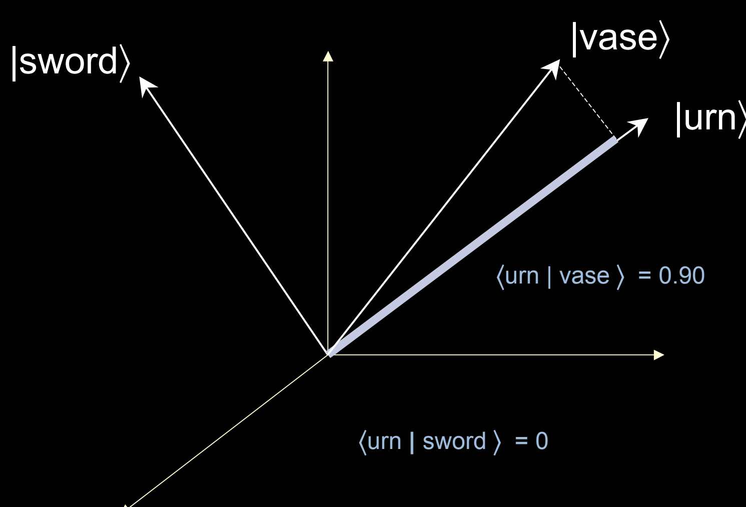
Similarity

$\langle \text{urn} | \text{urn} \rangle = 1$

$\langle \text{urn} | \text{vase} \rangle = 0.9$

$\langle \text{urn} | \text{sword} \rangle = 0$

De umbris idearum



Recall

Memory as <u>accumulation</u> of associations:

$M = |\text{joke}\rangle\langle\text{urn}| + |\text{insult}\rangle\langle\text{sword}| + |\text{apology}\rangle\langle\text{pool}| + \ldots$

The shadow cast on memory by the object reveals the image to be recalled:

$$\begin{aligned}
M|\text{urn}\rangle &= \boxed{(|\text{joke}\rangle\langle\text{urn}| + |\text{insult}\rangle\langle\text{sword}| + |\text{apology}\rangle\langle\text{pool}| + \ldots)}\,|\text{urn}\rangle \\
&= |\text{joke}\rangle\langle\text{urn}\,|\,\text{urn}\rangle + |\text{insult}\rangle\langle\text{sword}\,|\,\text{urn}\rangle + |\text{apology}\rangle\langle\text{pool}\,|\,\text{urn}\rangle + \ldots \\
&= |\text{joke}\rangle\ 1 + |\text{insult}\rangle\ 0 + |\text{apology}\rangle\ 0 + \ldots \\
&= |\text{joke}\rangle.
\end{aligned}$$

Atomize the Pattern

$|\text{urn}\rangle$

$\langle 1\,|\,\text{urn}\rangle = 2.3$

$\langle 2\,|\,\text{urn}\rangle = -0.4$

$\langle 3\,|\,\text{urn}\rangle = 8.3$

$\langle 4\,|\,\text{urn}\rangle = 1.9$

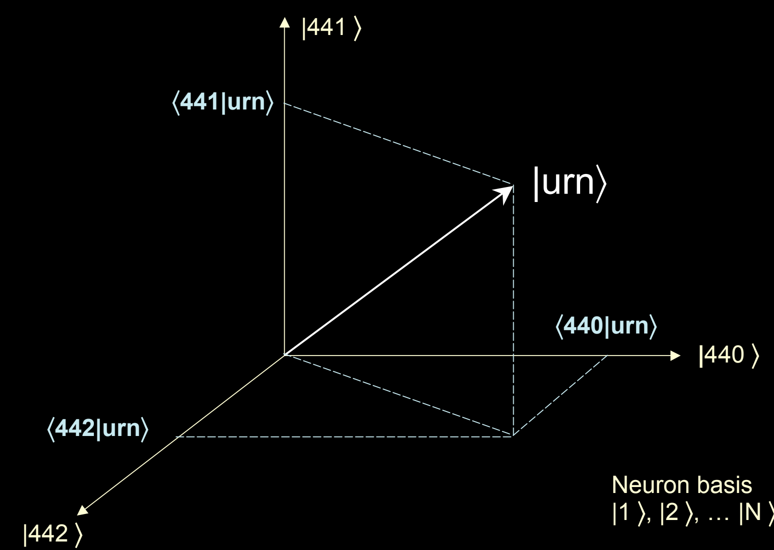
.
.
.

$\langle 365\,|\,\text{urn}\rangle = 5.3$

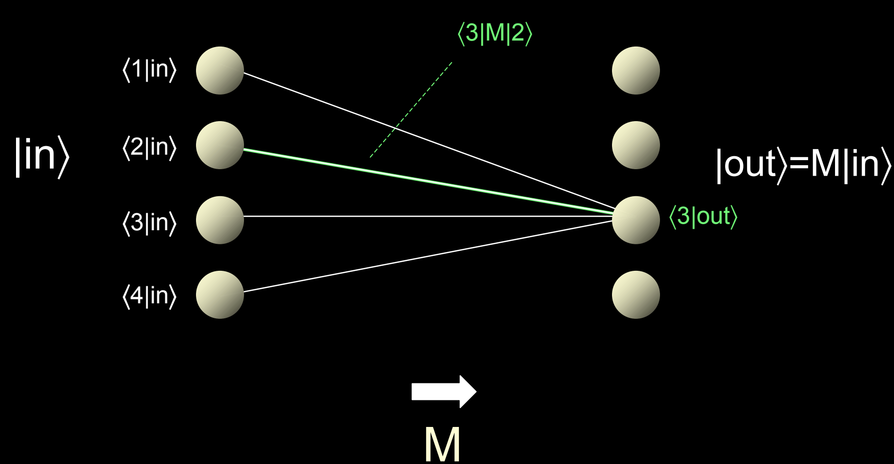
Neurons!

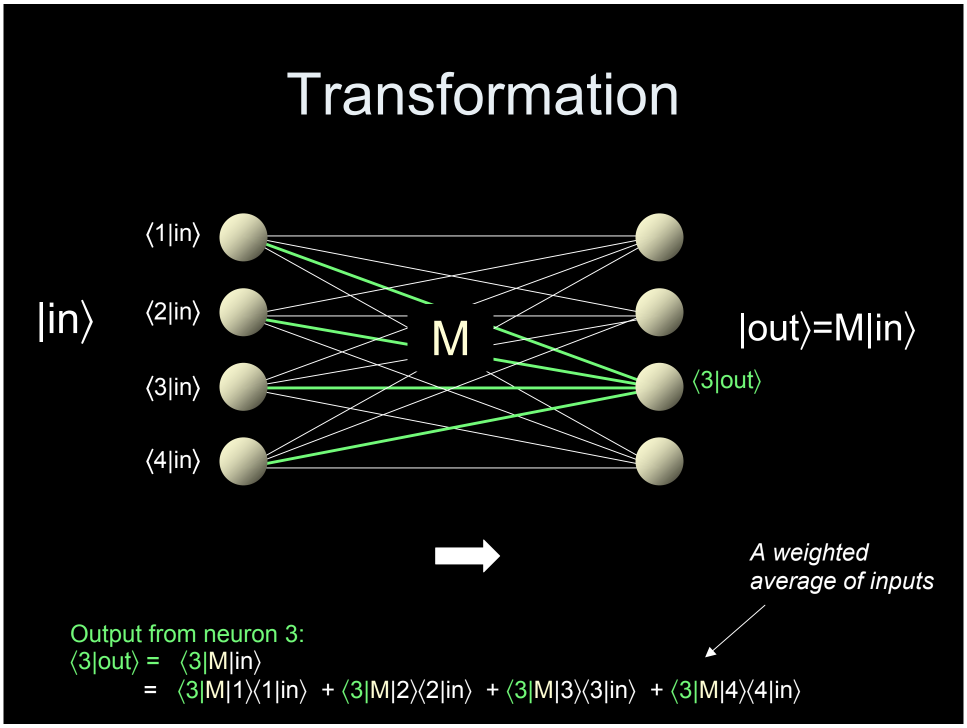
Patterns as vectors

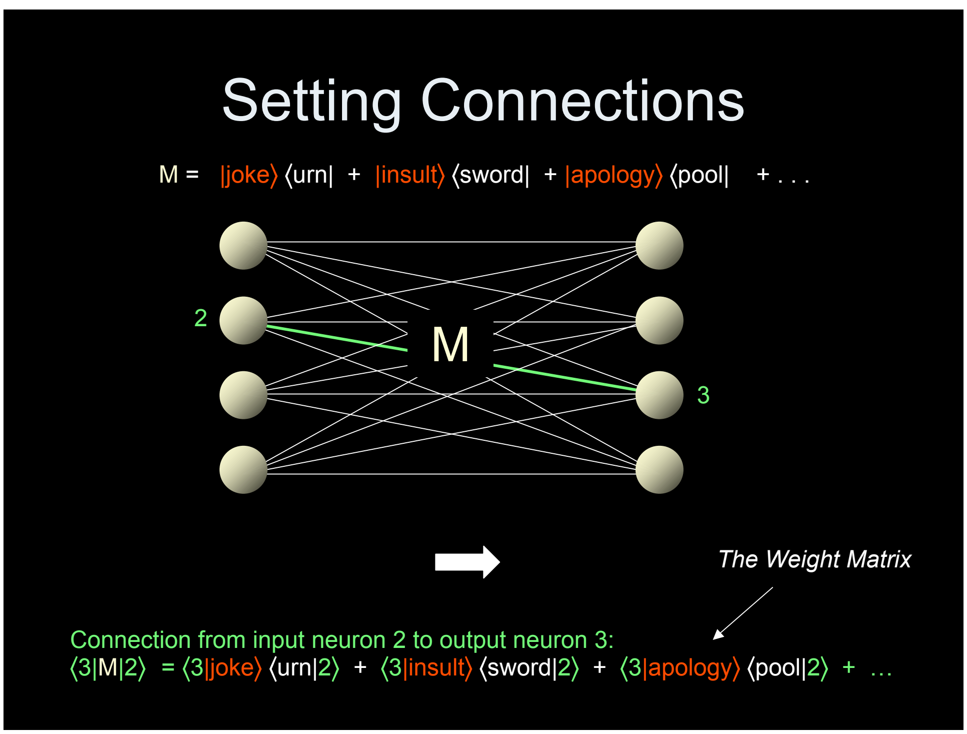
The Neuron Basis

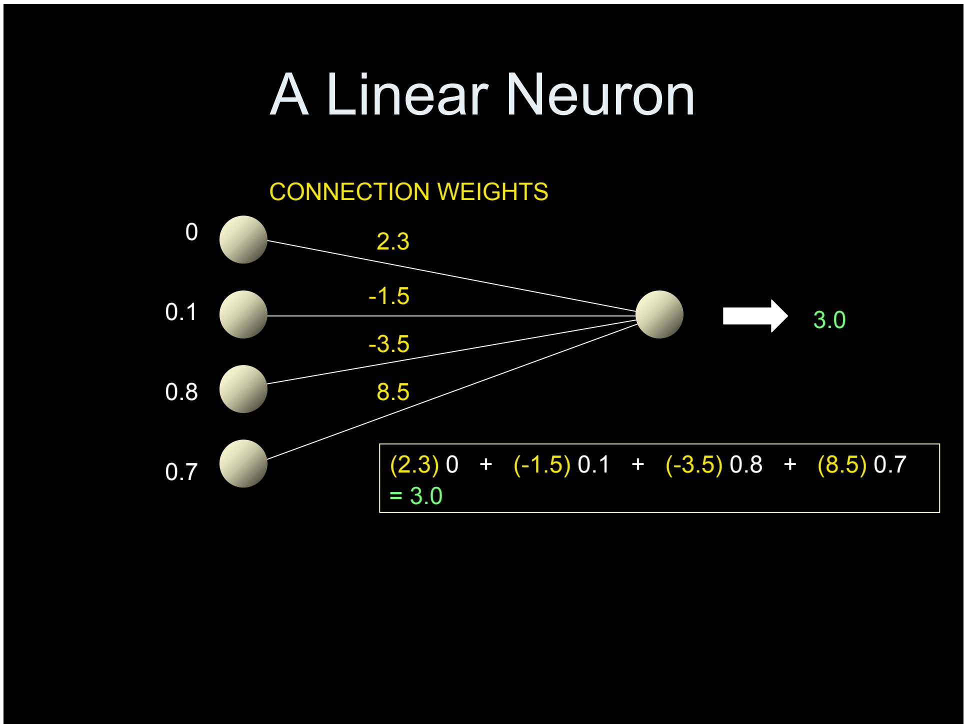


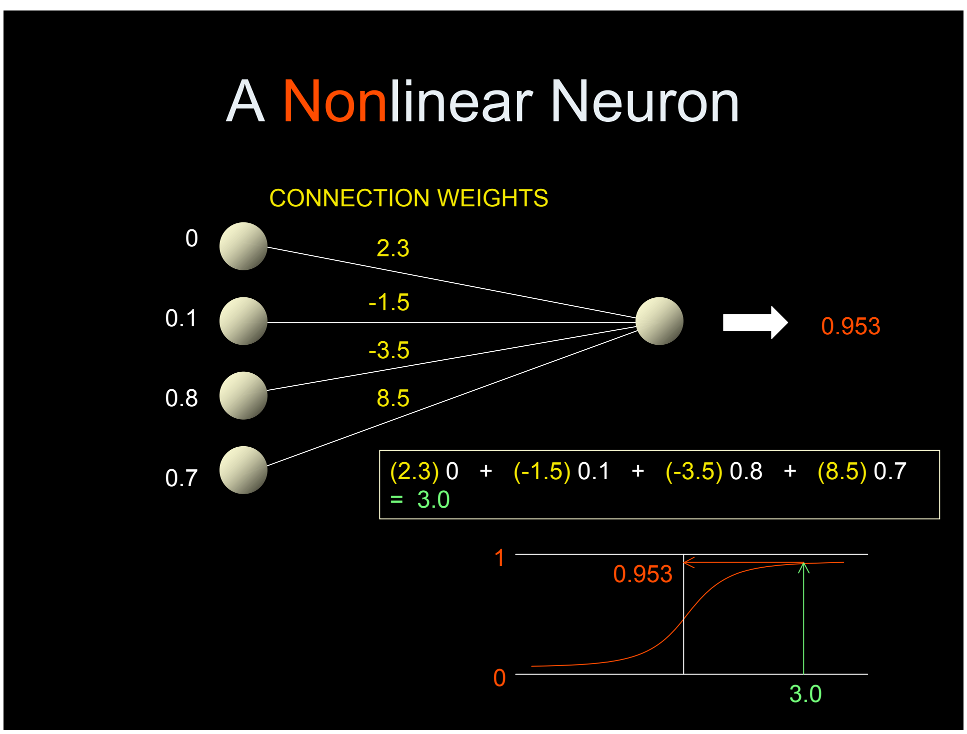
Connections



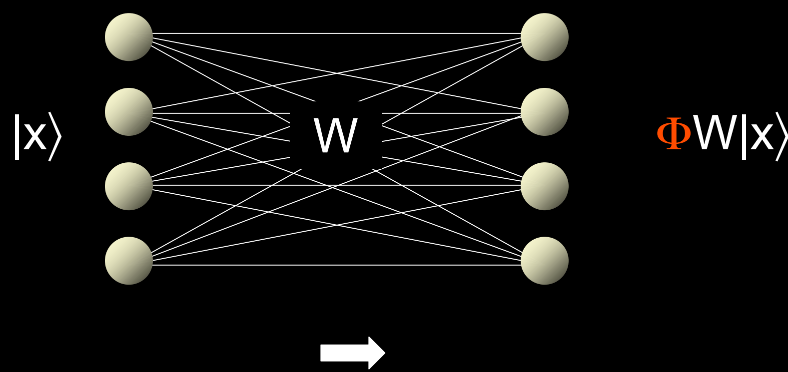

Transformation
⟨1|in⟩
⟨2|in⟩
⟨3|in⟩
⟨4|in⟩
|in⟩
M
|out⟩=M|in⟩
⟨3|out⟩
A weighted average of inputs
Output from neuron 3:
⟨3|out⟩ = ⟨3|M|in⟩
= ⟨3|M|1⟩⟨1|in⟩ + ⟨3|M|2⟩⟨2|in⟩ + ⟨3|M|3⟩⟨3|in⟩ + ⟨3|M|4⟩⟨4|in⟩


Setting Connections
M = |joke⟩ ⟨urn| + |insult⟩ ⟨sword| + |apology⟩ ⟨pool| + . . .
2
M
3
The Weight Matrix
Connection from input neuron 2 to output neuron 3:
⟨3|M|2⟩ = ⟨3|joke⟩ ⟨urn|2⟩ + ⟨3|insult⟩ ⟨sword|2⟩ + ⟨3|apology⟩ ⟨pool|2⟩ + …

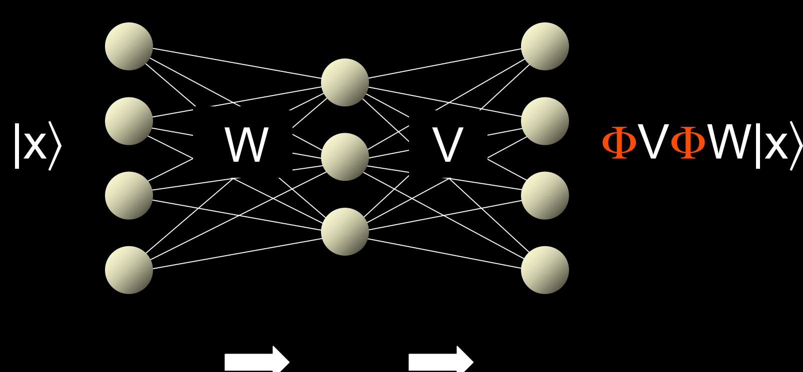

A Linear Neuron
CONNECTION WEIGHTS
0
0.1
0.8
0.7
2.3
-1.5
-3.5
8.5
3.0
(2.3) 0 + (-1.5) 0.1 + (-3.5) 0.8 + (8.5) 0.7
= 3.0


A Nonlinear Neuron
CONNECTION WEIGHTS
0
0.1
0.8
0.7
2.3
-1.5
-3.5
8.5
0.953
(2.3) 0 + (-1.5) 0.1 + (-3.5) 0.8 + (8.5) 0.7
= 3.0
1
0.953
0
3.0

Layered Nonlinear Network

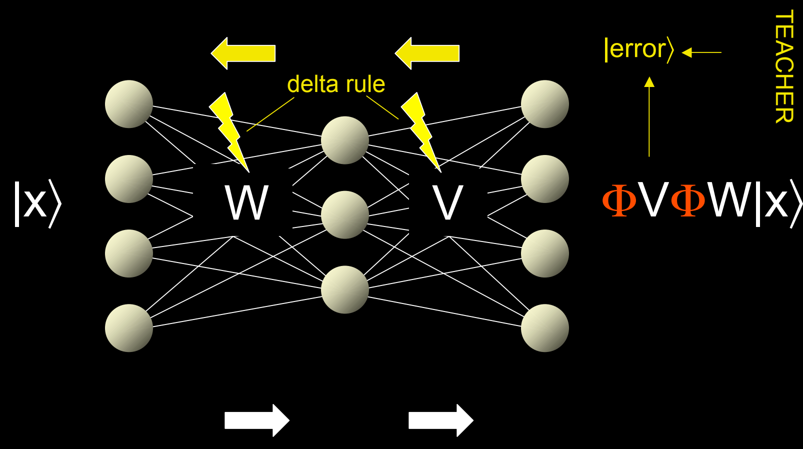


Adding a Hidden Layer



Almost universal in its ability to represent arbitrary transformations of patterns

Supervised Learning



Mathematical optimization problem: Adjust W,V to minimize error.

Memory, Learning

The "Quintilian rule" for memory

repeat:

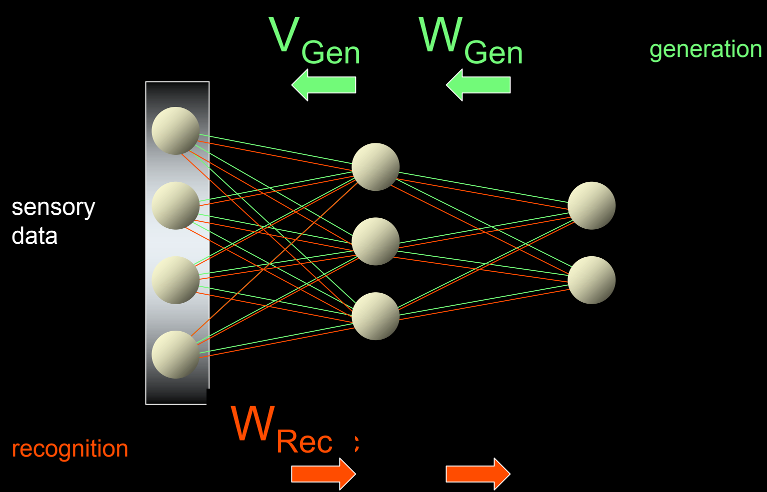
Memory += **| response ⟩ ⟨ trigger |**

The delta rule for learning

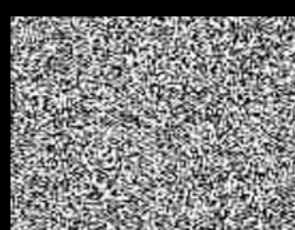
repeat:

Memory += ε **| error in response ⟩ ⟨ trigger |**

Helmholtz Machine (almost)



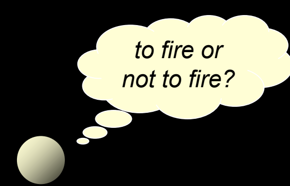
Fluctuation



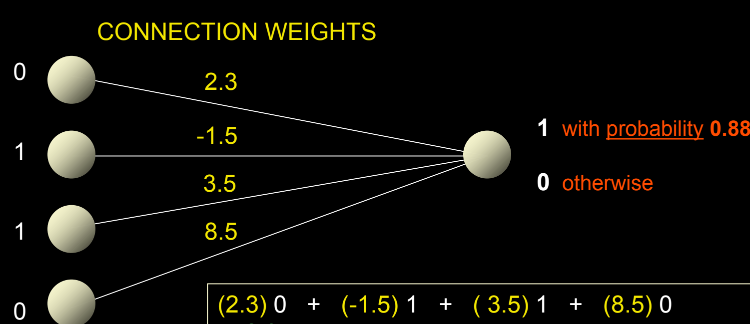
Probability

0 ———— 1

probability

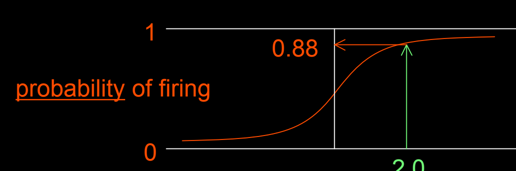


A Binary Stochastic Neuron

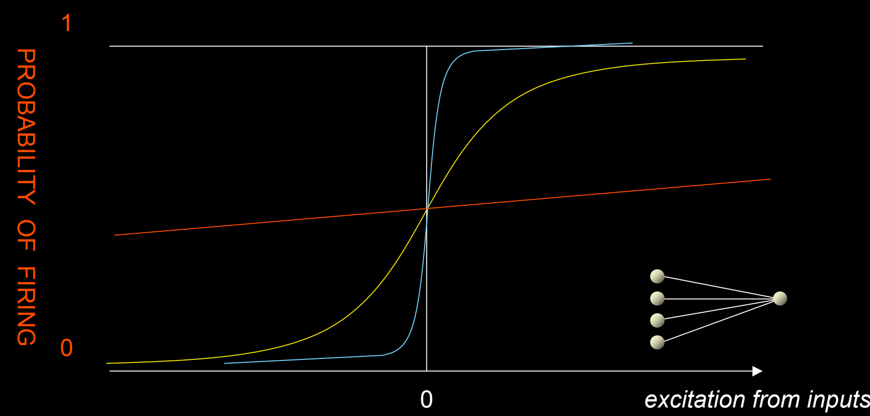


$(2.3)\,0 \;+\; (-1.5)\,1 \;+\; (3.5)\,1 \;+\; (8.5)\,0$
$= 2.0$

neurons are binary
(1 fire / 0 not fire)



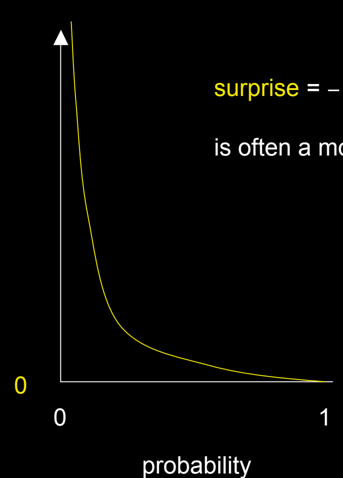
Temperature



low temperature: nearly deterministic-- fire if and only if input excitation is positive

high temperature: nearly random-- fire by flipping a 50-50 coin!

Surprise



surprise = – log probability

is often a more useful quantity than probability

Entropy

entropy = expected surprise

= probability of outcome 1 × surprise of outcome 1
+ probability of outcome 2 × surprise of outcome 2
+ probability of outcome 3 × surprise of outcome 3
+ probability of outcome 4 × surprise of outcome 4
+ …

PPPPPPPPPPPSSPPPPPPPPCPPPPSPPPCPPPPSPPPPPPCPPPPPSS

Outcome	*Count* [total = 50]
Phlegmatic	41
Sanguine	6
Choleric	3

H = – 41/50 log (41/50) – 6/50 log (6/50) – (3/50) log (3/50)
= 0.586

Entropy

entropy = expected surprise

= probability of outcome 1 × surprise of outcome 1
+ probability of outcome 2 × surprise of outcome 2
+ probability of outcome 3 × surprise of outcome 3
+ probability of outcome 4 × surprise of outcome 4
+ …

PPSPPPPPPSSCCSPPPCPPSSCCCSSPPPPSSSCCCSSCSSSSCCCCCC

Outcome	*Count* [total = 50]
Phlegmatic	17
Sanguine	17
Choleric	16

H = – 17/50 log (17/50) – 17/50 log (17/50) – (16/50) log (16/50)
= 1.10 *a situation fraught with greater surprise….*

Energy

Things like to roll downhill.
Things like to loose energy.
Build your neural network so that when it runs downhill it solves your problem.
This means: define an energy function in a clever way.

Some neural networks with clever energy functions:

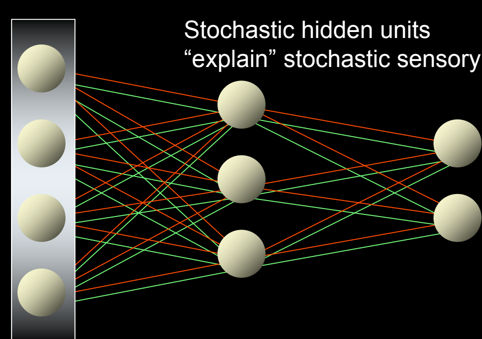
Hopfield network (1980), Boltzmann machine (1985)
Energy of global network pattern is $-\langle \text{pattern} \mid \text{Weights} \mid \text{pattern} \rangle$

Helmholtz machine (1995)
Energy is surprise level of the stochastic hidden neurons.

Helmholtz Machine

Stochastic hidden units
“explain” stochastic sensory data

sensory
data

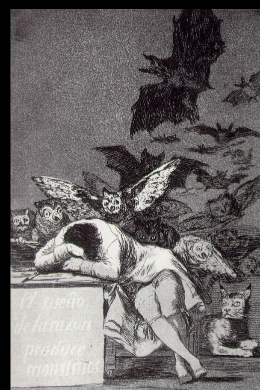


Evolve recognition weights
and generation weights
that allow hidden neurons to explain sense data
by minimizing this:

= surprise
Average Energy
– Temperature × Entropy
= 1

HELMHOLTZ FREE ENERGY

Dreams



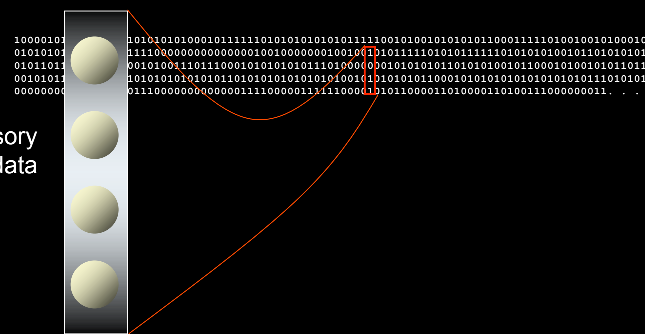
UNSUPERVISED!

Representational Learning

making sense of blooming buzzing confusion

Representational Learning

sensory
data

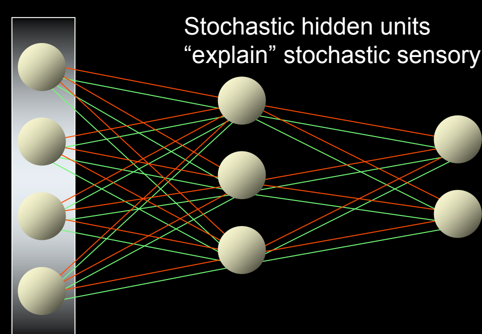


Blooming, buzzing confusion
Can the system learn to represent this sense data in a concise natural form?

Representational Learning

Stochastic hidden units
“explain” stochastic sensory data

sensory
data



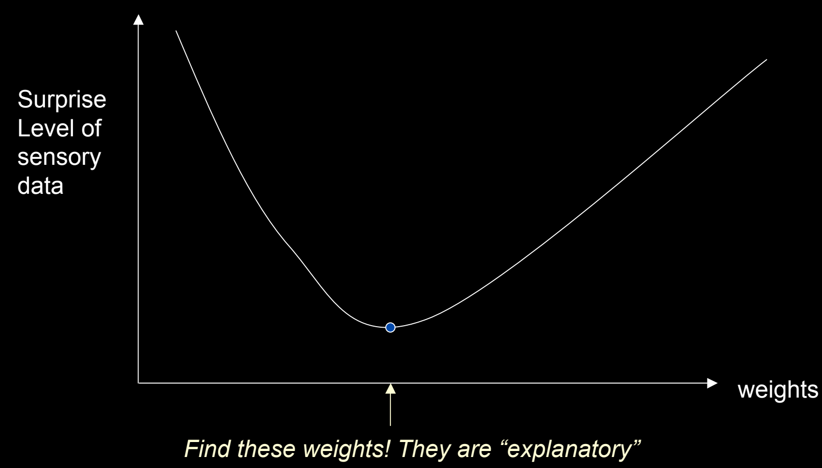
"causes"

The system learns to associate probability distributions rather than patterns.

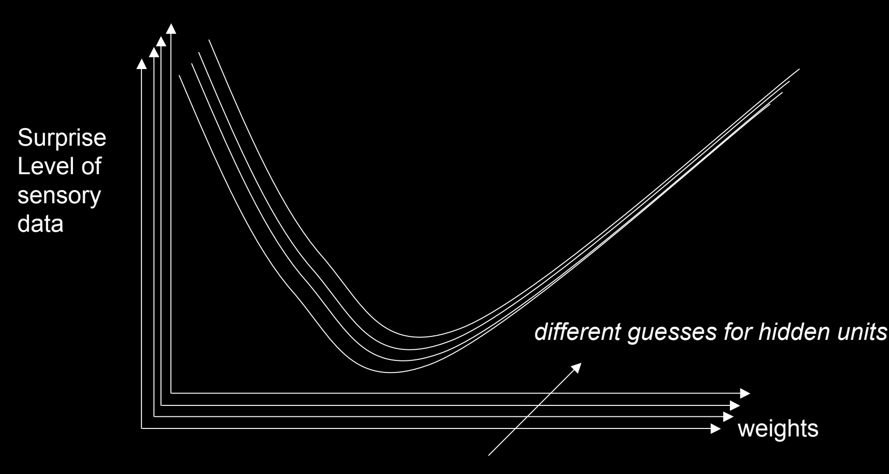
Find the weights that minimize surprise.

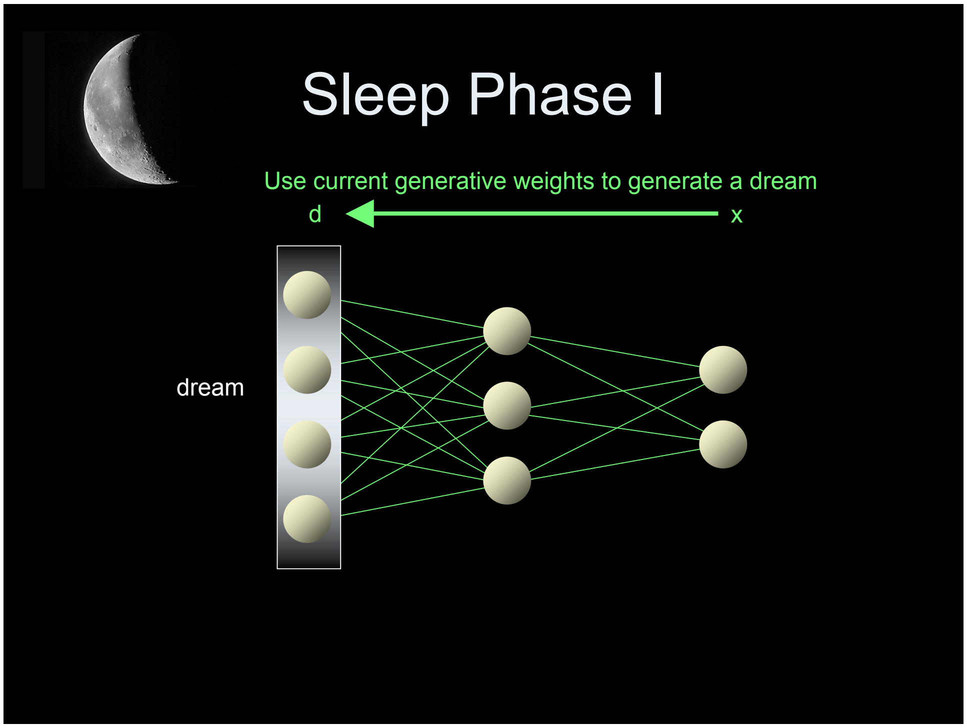
Surprise Minimization

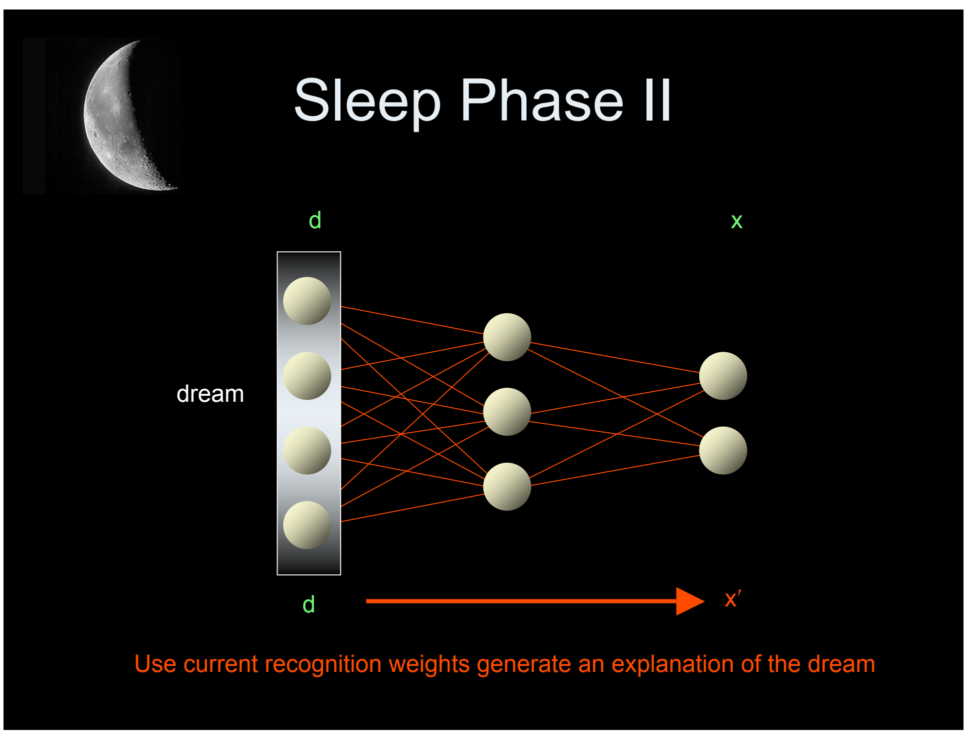
Better known as: Maximum (log) likelihood estimation (MLE)

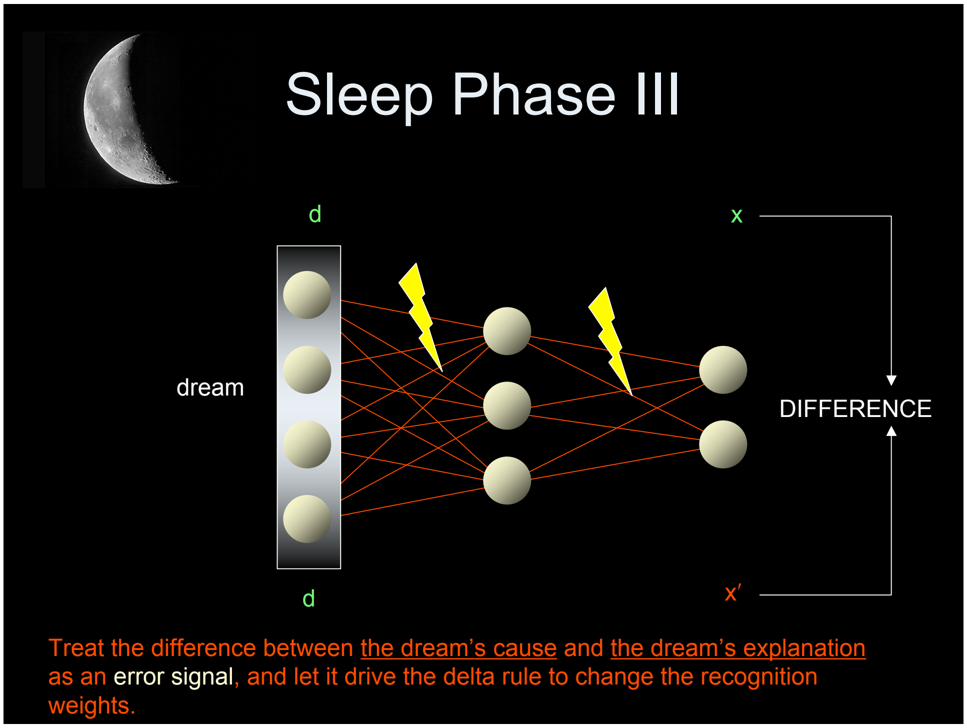


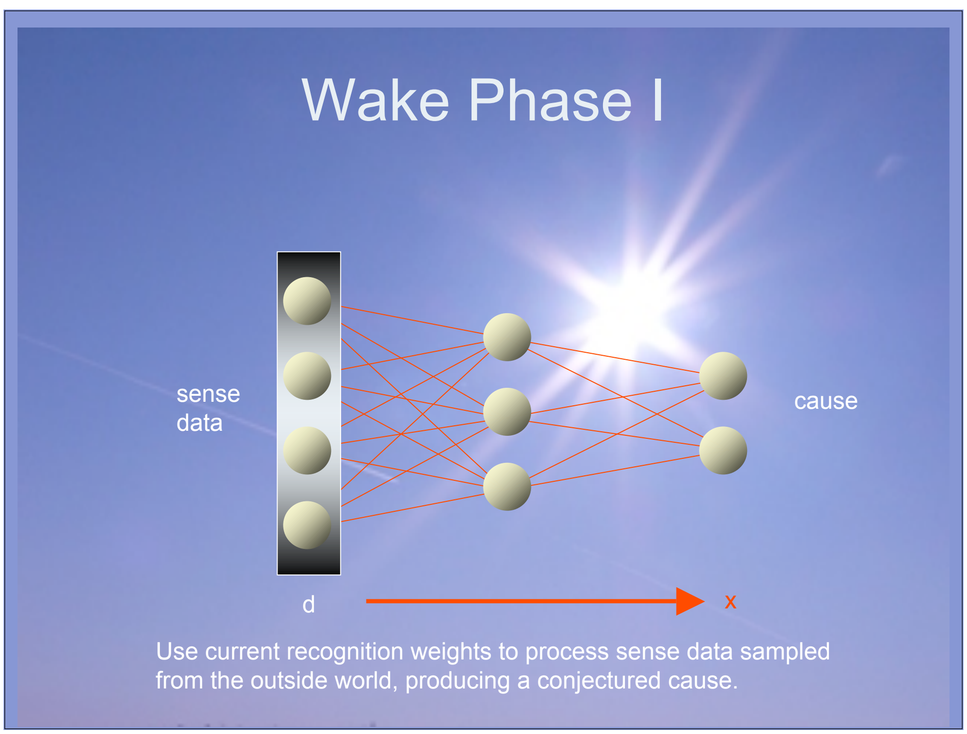
EM Algorithm

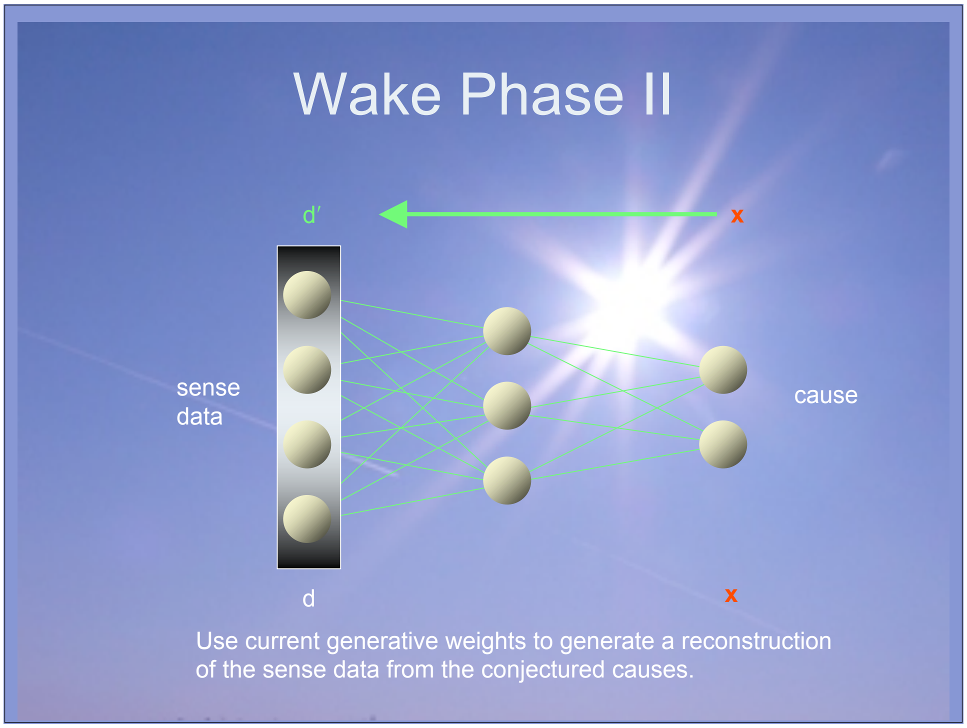


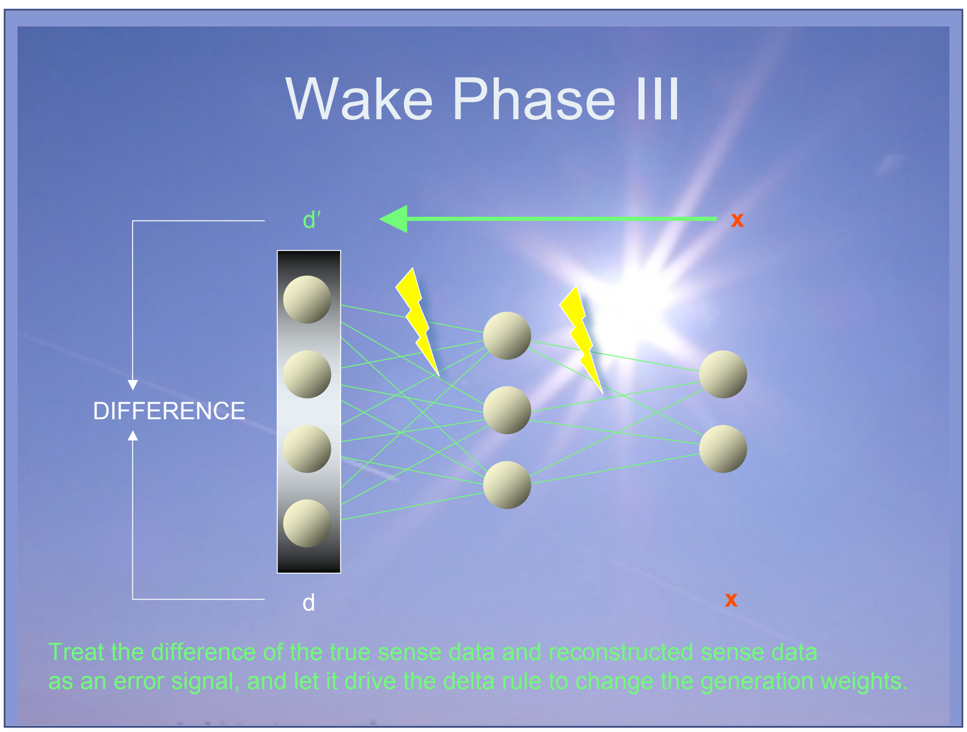

Sleep Phase I
Use current generative weights to generate a dream
d
x
dream


Sleep Phase II
d
x
dream
d
x′
Use current recognition weights generate an explanation of the dream


Sleep Phase III
d
x
dream
DIFFERENCE
d
x′
Treat the difference between the dream's cause and the dream's explanation as an error signal, and let it drive the delta rule to change the recognition weights.


Wake Phase I
sense data
cause
d
x
Use current recognition weights to process sense data sampled from the outside world, producing a conjectured cause.

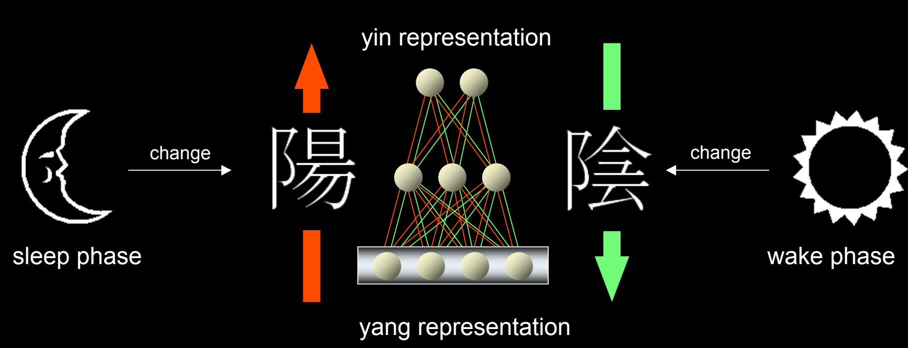

Wake Phase II
d′
x
sense
data
cause
d
x
Use current generative weights to generate a reconstruction
of the sense data from the conjectured causes.


Wake Phase III
d′
x
DIFFERENCE
d
x
Treat the difference of the true sense data and reconstructed sense data
as an error signal, and let it drive the delta rule to change the generation weights.

Wake-Sleep Minimizes Helmholtz Free Energy

wake
sleep
wake
sleep
wake
sleep
wake
sleep
wake
sleep
wake
sleep
.
.
.

Up-and-Down as a Key Mode of Information Processing



"Applications"

- Handwritten digit recognition / generation
- Small image recognition / generation
- Text analysis for document classification
- Explaining perceptual processing in the neocortex

"Applications?"

- Musical composition via internalization of pre-existing works
- Musical composition via real-time unfolding of generative model
- Compressed representations
- Communities of co-learning Helmholtz machines ("*you* live my dreams")

Further explorable connections:
- Harmonium (Smolensky)
- Information geometry (Amari)

Genealogy

Statistical mechanics (c. 1902)

Maximum likelihood (c. 1934)

Threshold unit neurons (1943)

EM Algorithm (1977)

Boltzmann machine (1985)

Multilayer Perceptron Backpropagation (1985)

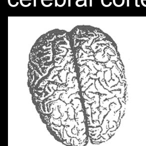
Statistical physics view of EM Algorithm (1992)

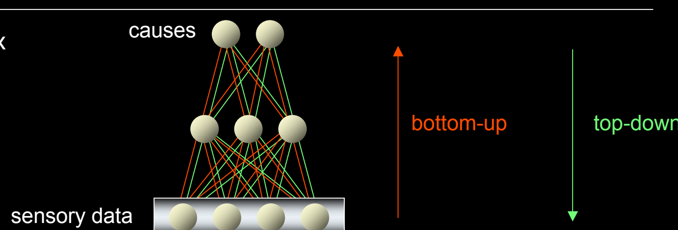
Helmholtz machine and wake-sleep learning (1995)

cf. Bayesian "ying-yang systems" (1995)

In the Brain

cerebral cortex

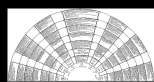




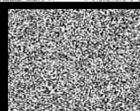
"When one understands the causes, all vanished images can easily be found again in the brain through the impression of the cause. This is the true art of memory…"

Rene Descartes, *Cogitationes privatae*.
Quoted in Frances Yates, *The Art of Memory* (1966)

Tomorrow



Closer looks at artificial neural memory



What exactly is free energy in a Helmholtz machine?

How do you code Helmholtz machines?

Helmholtzmaschinen im Klanglabor?
Does the sleep of reason bring forth monsters?

Belief dynamics?
Exaptability and torsion?
Quantum representations?
Evolutionary algorithms?
Rice's Theorem?

Related Readings

Dayan, P., G.E. Hinton, R.M. Neal, and R.S. Zemel. 1995. The Helmholtz machine. *Neural Computation 7*:5,889–904.

Dayan, P. and G.E. Hinton. 1996. Varieties of Helmholtz Machine. *Neural Networks 9*:8, 1385-1403.

Dayan, P. 2003. Helmholtz machines and sleep-wake learning. In M.A. Arbib, Ed., *Handbook of Brain Theory and Neural Networks*, *Second Edition*. MIT Press.

Hinton, G.E., P. Dayan, B.J. Frey and R.M. Neal. 1995. The wake-sleep algorithm for unsupervised neural networks. *Science 268*: 1158-1161.

Ikeda, S., S.-I. Amari and H. Nakahara. 1999. Convergence of the wake-sleep algorithm. In M.S. Kearns et al, Eds., *Advances in Neural Information Processing Systems* 11.

Kirby, K.G. 1997. Of memories, neurons, and rank-one corrections. *College Mathematics Journal* 28:1, 2–19.

Neal, R. and G. Hinton. 1998. A view of the EM algorithm that justifies incremental, sparse and other variants. In M.I. Jordan, Ed., *Learning in Graphical Models*, MIT Press.

Xu, Lei. 2003 Ying yang learning. In M.A. Arbib, Ed., *Handbook of Brain Theory and Neural Networks*, *Second Edition*. MIT Press.

And a new student thesis (March 26,2006!) → Pape, Leo. "Neural Machines for Music Recognition", Utrecht Univ.

Other Readings

THREE RELEVANT NEURAL NETWORK BOOKS

Dayan, P. and L.F. Abbott. 2001. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press.

Haykin, Simon. 1998. *Neural Networks: A Comprehensive Foundation*. Second Edition. Prentice-Hall.

Rumelhart, D.E., J.L. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press.

SOME OF MY PAPERS

K. Kirby, "Jeffrey-Bayes Dynamics as Natural Gradient Descent by a Recurrent Network" Submitted, *Neural Computation*, June 2005.

K. Kirby, "Biological Adaptabilities and Quantum Entropies." *BioSystems* 64, pp.31-41 (2002).

K. Kirby. "Exaptation and Torsion: Toward a Theory of Natural Computation". *BioSystems* 46, pp. 81-88. (1997.)

K. Kirby. "Hermeneutics and Biomolecular Computation." *Optical Memory and Neural Networks*, Vol. 4 No. 2, pp.111-117 (1995).

K. Kirby. "Duality in Sequential Associative Memory: A Simulationist Approach", *Proceedings, IEEE Int'l Conf. on Systems Engineering*, pp. 351-354. (1991).

K. Kirby and N. Day. "The Neurodynamics of Context Reverberation Learning." *Proceedings, IEEE Conference on Engineering in Medicine and Biology*, pp.1781-1782 (1990).

K. Kirby. "Information Processing In the Lorenz-Turing Neuron." *Proceedings of the 11th International IEEE Conference on Electronics in Medicine and Biology* (Seattle, November 9-12, 1989), Volume 11, 1358-1359.

K. Kirby and M. Conrad. "Intraneuronal Dynamics as a Substrate For Evolutionary Learning." *Physica D,* Vol. 22, 150-175 (1986).

K. Kirby and M. Conrad. "The Enzymatic Neuron as a Reaction-Diffusion Network of Cyclic Nucleotides." *Bulletin of Mathematical Biology*, Vol. 46, 765-782 (1984).

Northern Kentucky University