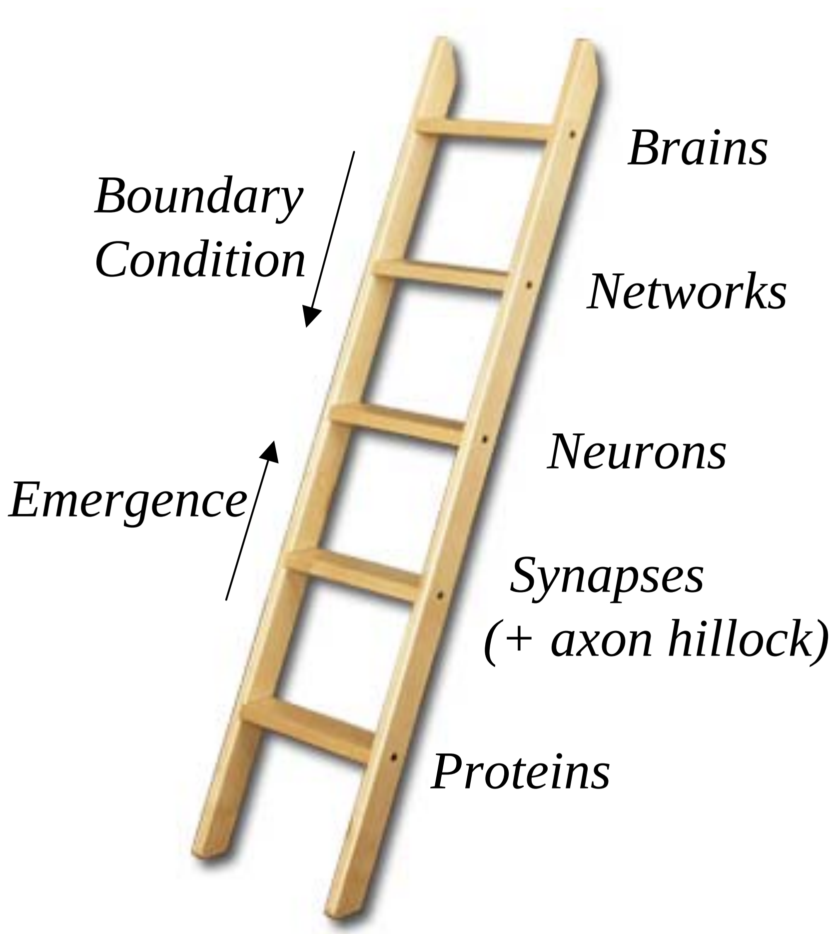
Biology: what’s the problem? The Levels Perspective

Tony Bell
Redwood Center for Theoretical Neuroscience
University of California at Berkeley

The Levels Ansatz

This is the idea that there is *no fundamental level in biology*.



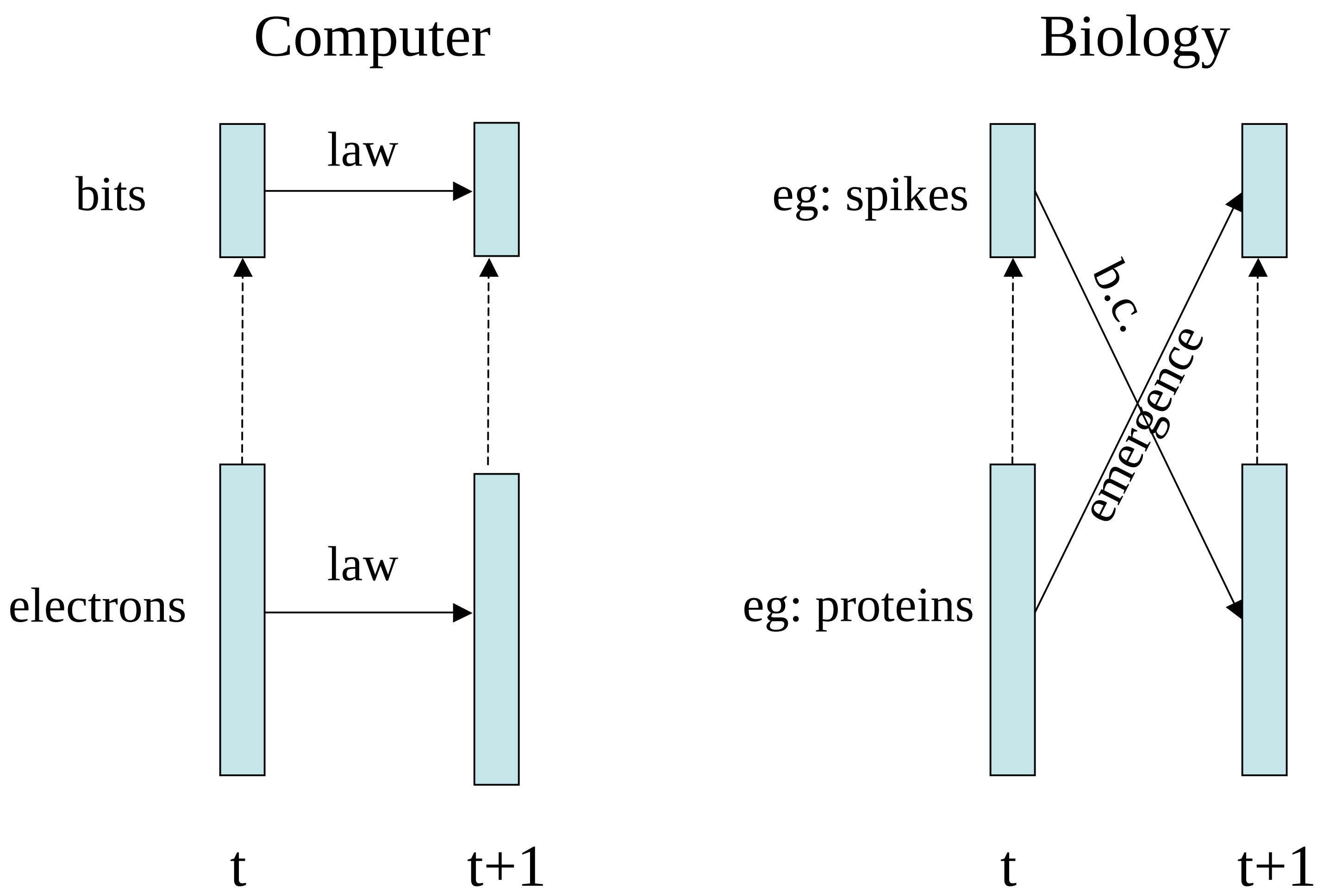
1. Learn about the rungs

2. Learn to walk up and down

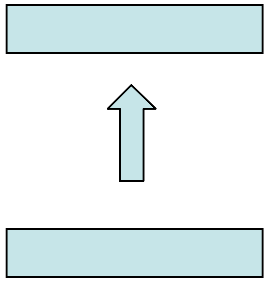
Because that is what the biological information is itself doing.

CLAIM: The adaptive power of biology comes from the inter-level information flows, not from the computation at any given level

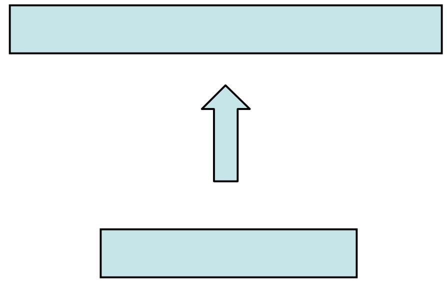
State vectors in machines and nature



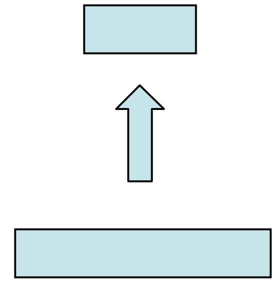
Unsupervised Learning (density estimation)



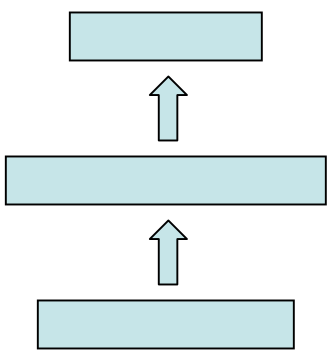
Complete:
ICA and extensions



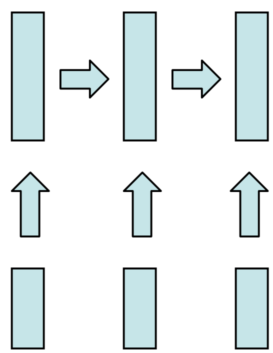
Overcomplete:
Sparse coding



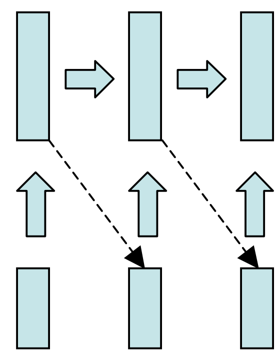
Undercomplete:
??



Multilayer:
Deep Belief Nets ?



Temporal:
Kalman, HMM,
Dynamic Bayes



Sensorimotor:
???

Sensorimotor density estimation

for

- $p(\mathbf{x})$: data distribution
- $q(\mathbf{x})$: model distribution
- $D[p \mid q]$: divergence of model from data
- w: a synaptic weight

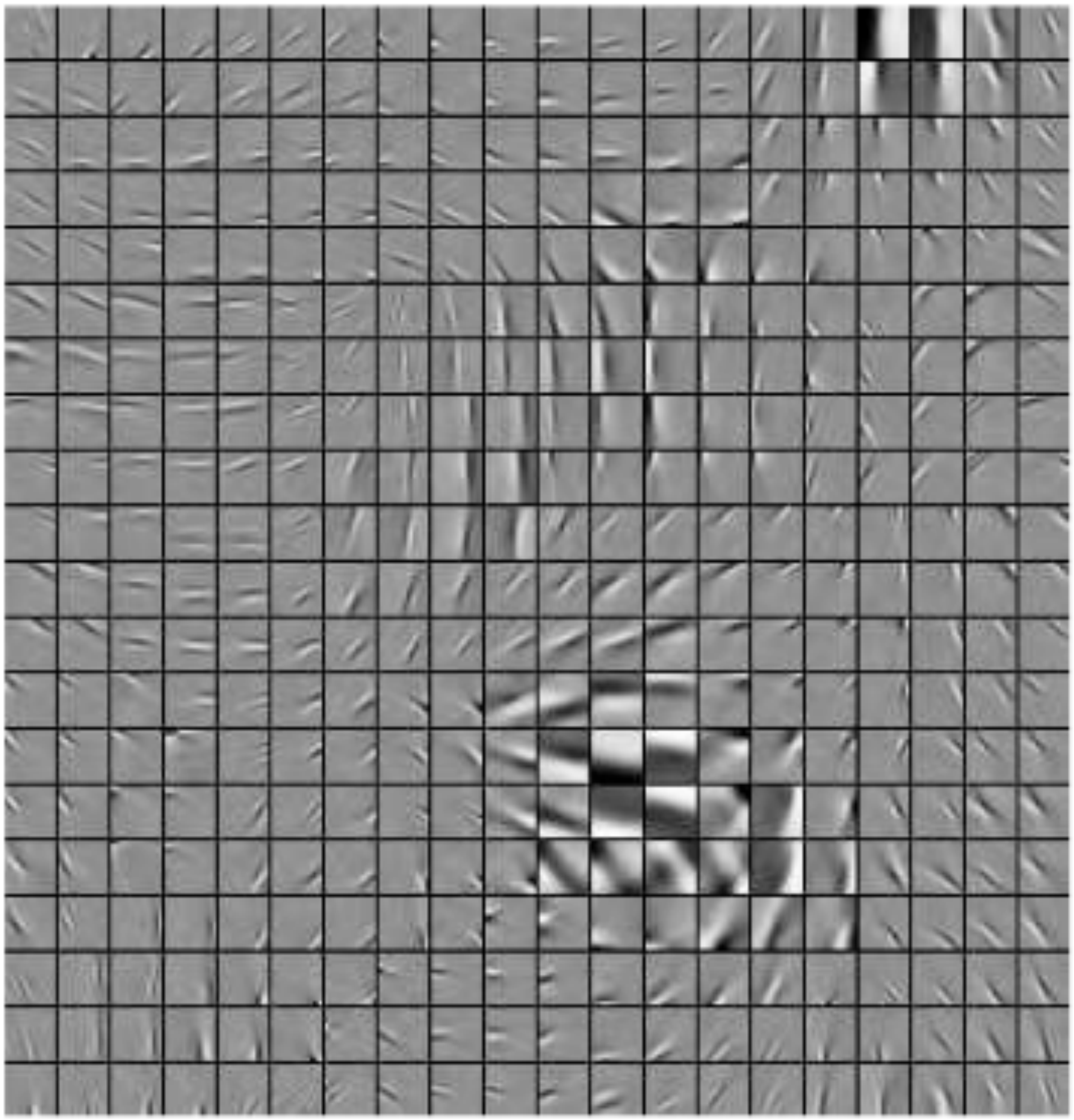
the learning gradient is:

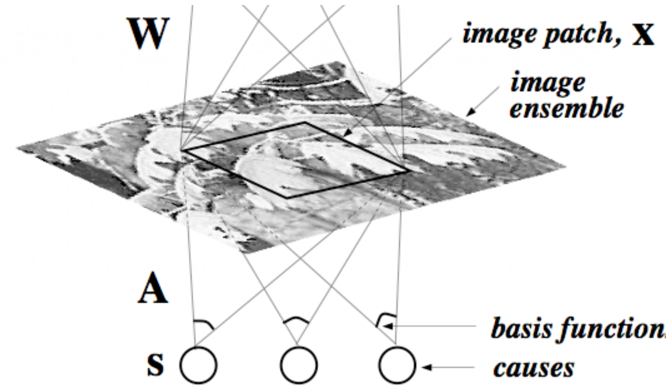
$$\partial_w D[p|q] = \left\langle \underbrace{\left(1+\log\frac{p}{q}\right)\partial_w \log p}_{\text{the motor problem}} - \underbrace{\partial_w \log q}_{\text{the sensory problem}} \right\rangle_p$$

change world
to fit model

change model
to fit world

Unsupervised learning from natural images





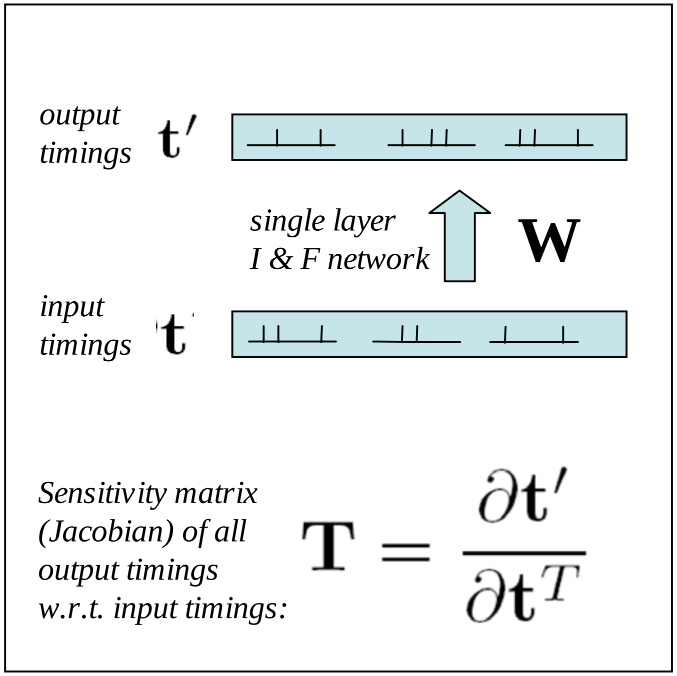
RESULTS:
simple cells
complex cells
V1-type topography
(‘orientation column’)

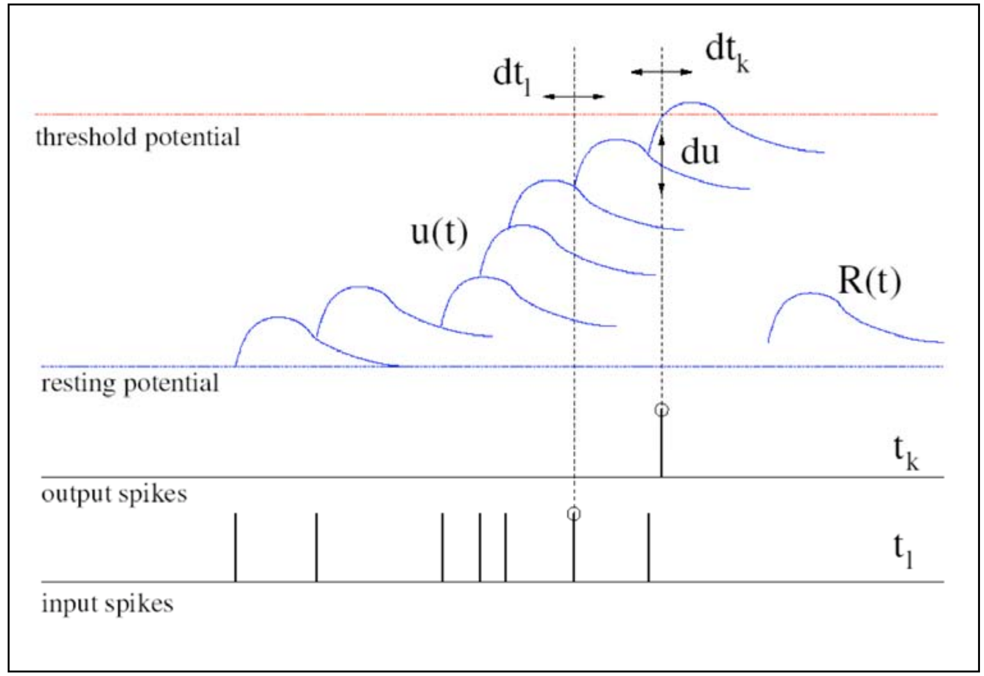
Density estimate16x16 image patches with assumptions:

(1) ‘Independence’ or sparseness
(2) 2D topography

Olshausen & Field 97
Bell & Sejnowski 97
Hyvarinen & Hoyer 01
(this result: Osindero et al 06)

Spikelihood (unsupervised learning with spikes)





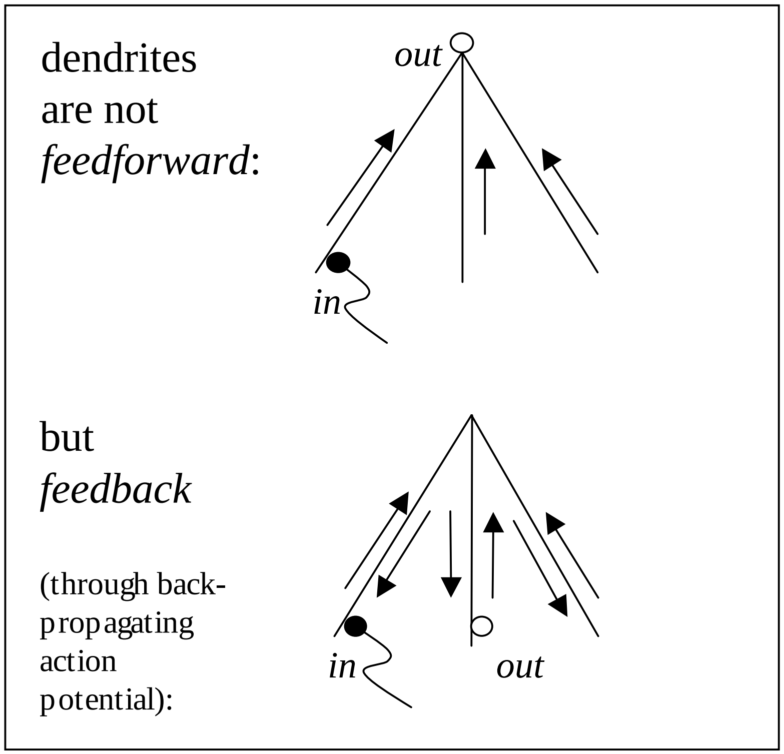
gives the most complicated unsupervised learning rule ever derived:

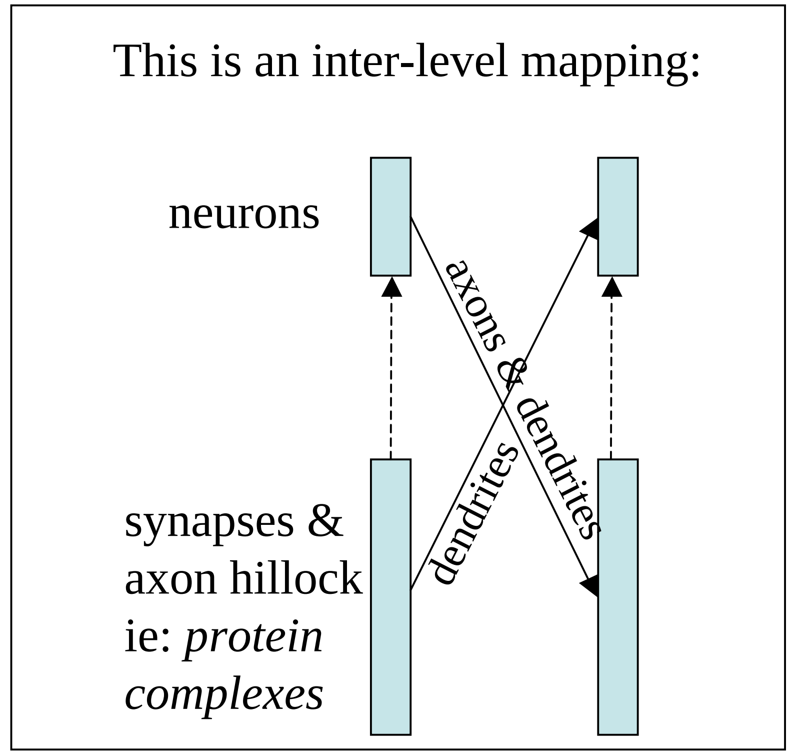
$$\Delta \mathbf{W}_{ij} \propto \frac{\mathbf{T}_{kl}}{\mathbf{W}_{ij}} \left(\left[\mathbf{T}^{T\#}\right]_{kl} - \left[\mathbf{T}\mathbf{T}^{T\#}\right]_{kk} \right) - f(r_i) r_j$$

input and output rates

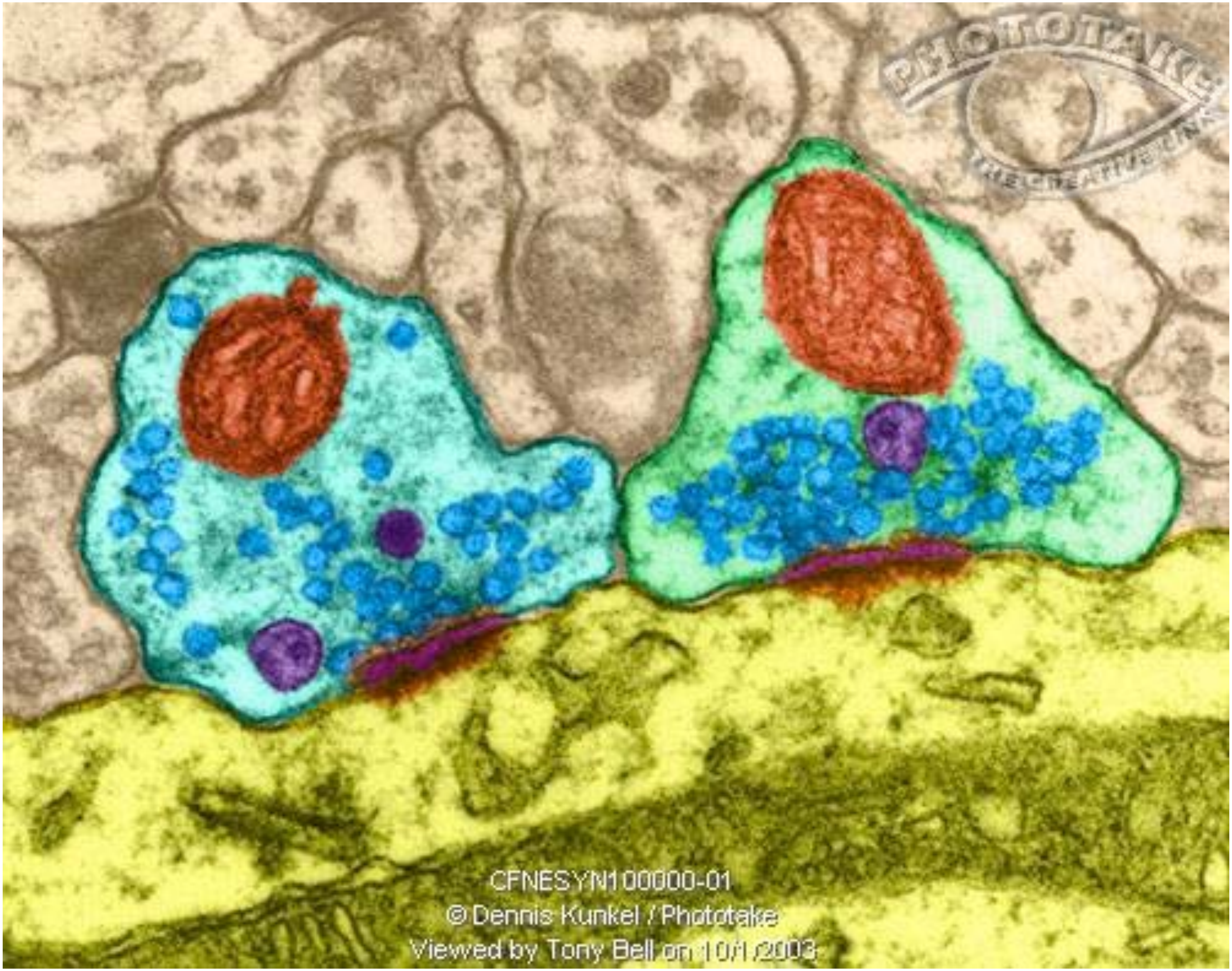
Bell & Parra *NIPS* 04, Parra, Beck & Bell, *Neur. Comp.* 08

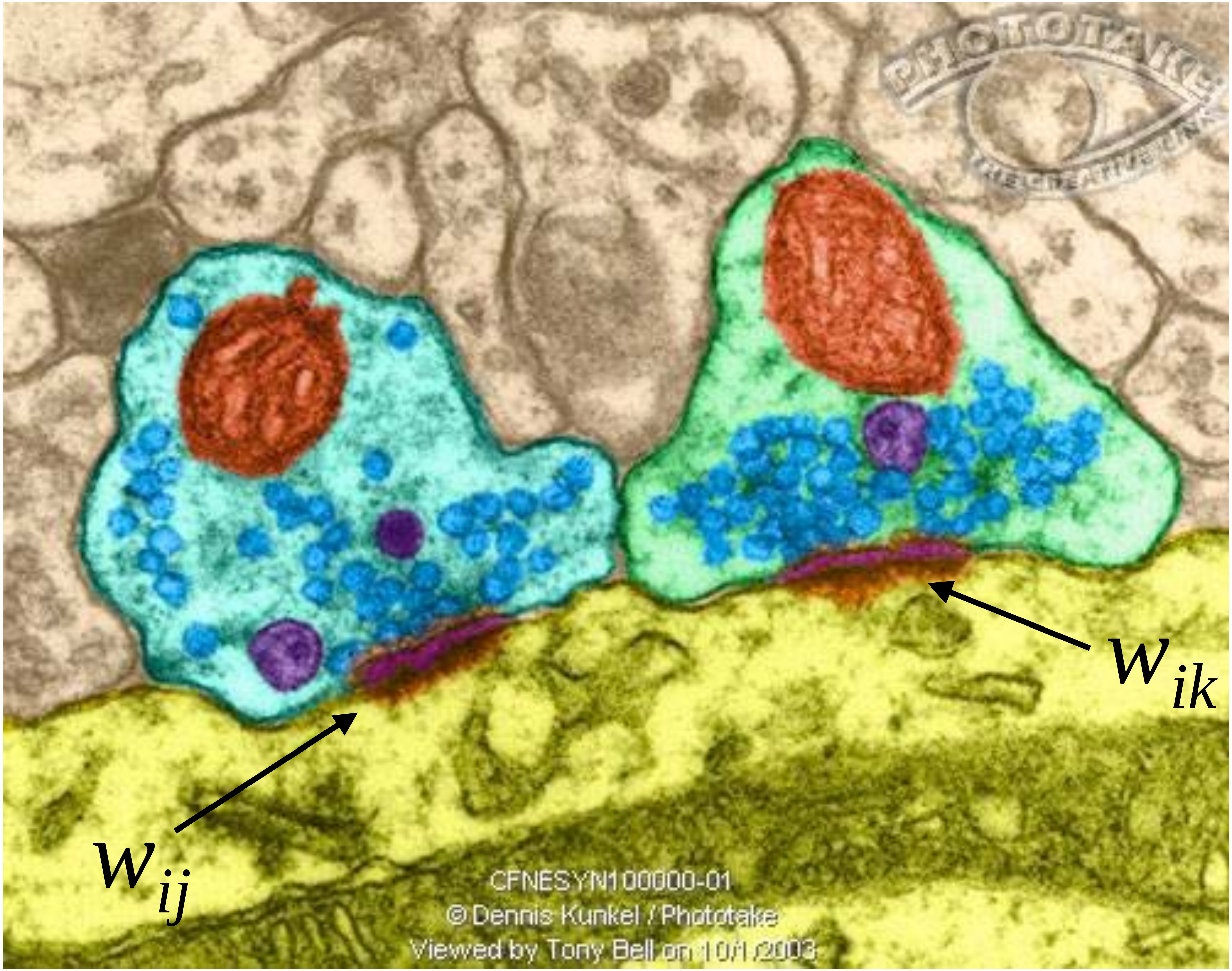
What went wrong?





Neurons map into an overcomplete, more microscopic, space (synapses)


PHOTOTAKE
CFNESYN100000-01
© Dennis Kunkel / Phototake
Viewed by Tony Bell on 10/1/2003


PHOTOTAKE
THE CREATIVE LINK
w_{ij}
w_{ik}
CFNESYN100000-01
© Dennis Kunkel / Phototake
Viewed by Tony Bell on 10/1/2003

and there are lots of these protein complexes in dendrites:

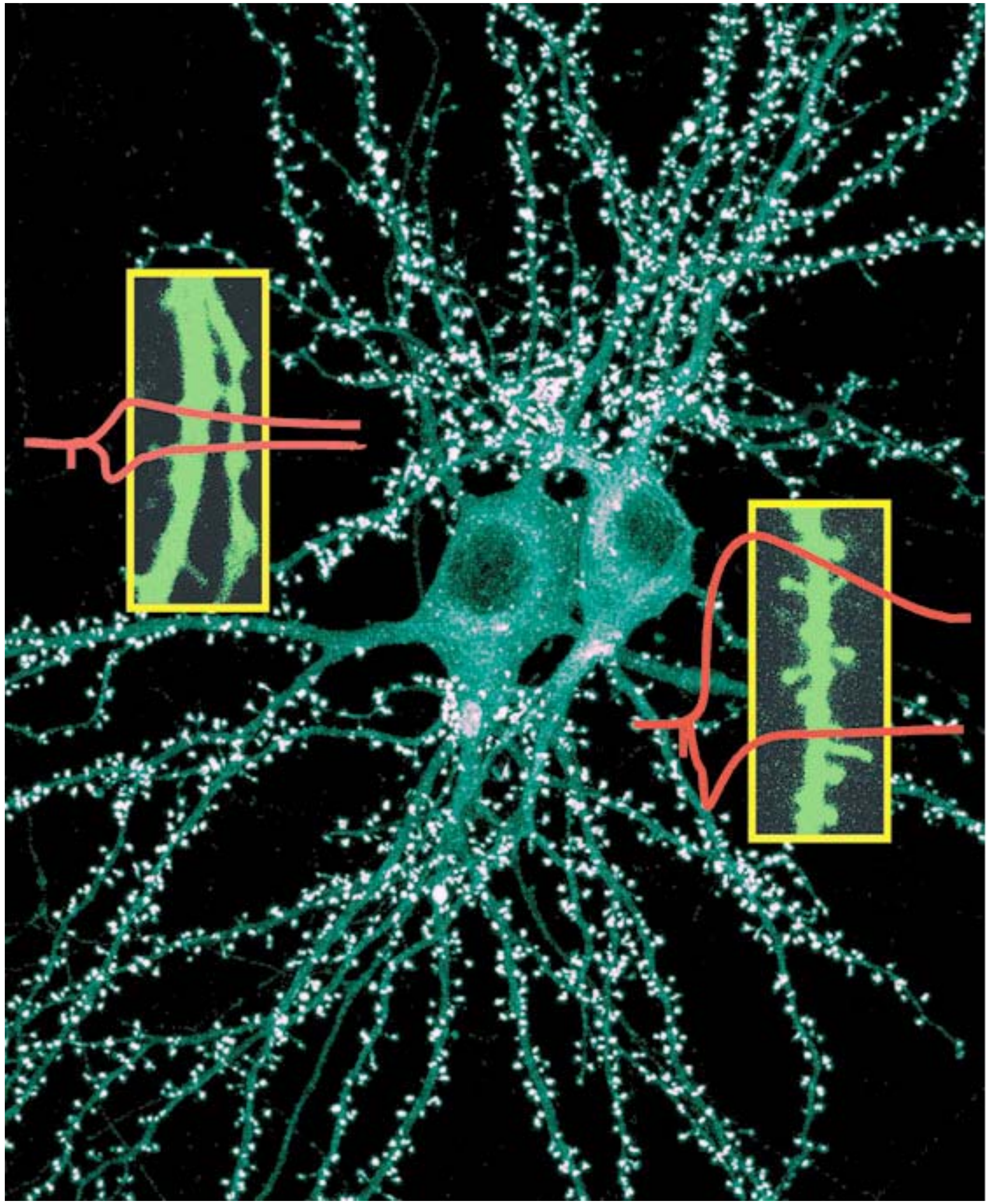
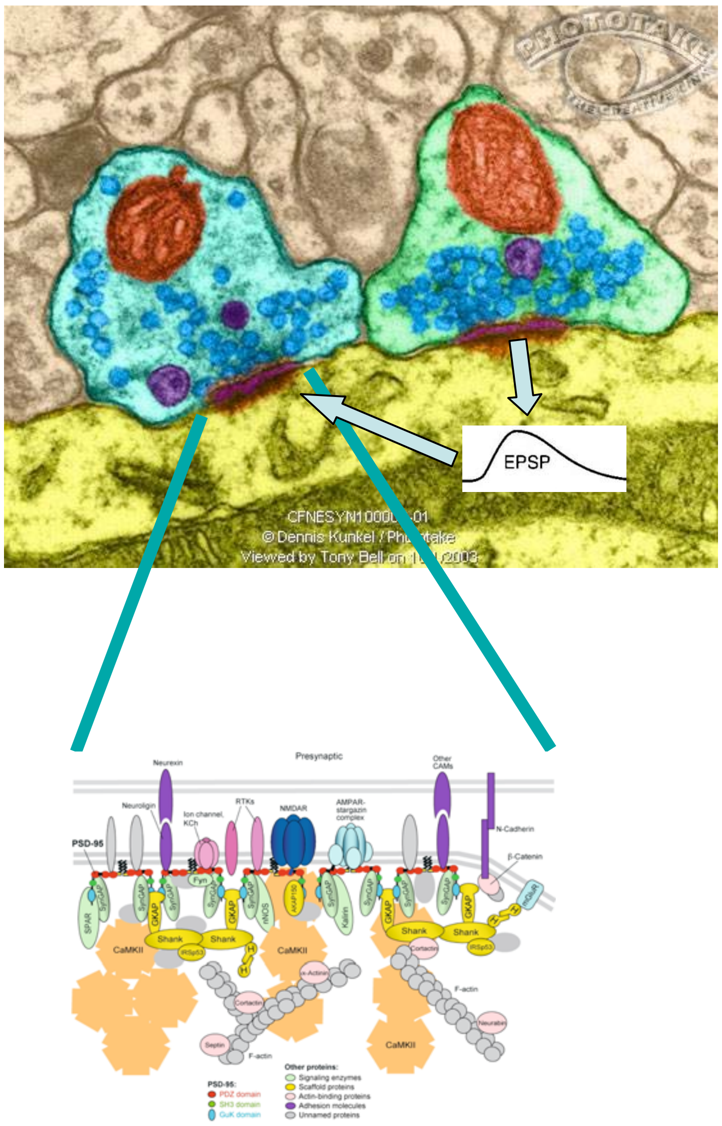
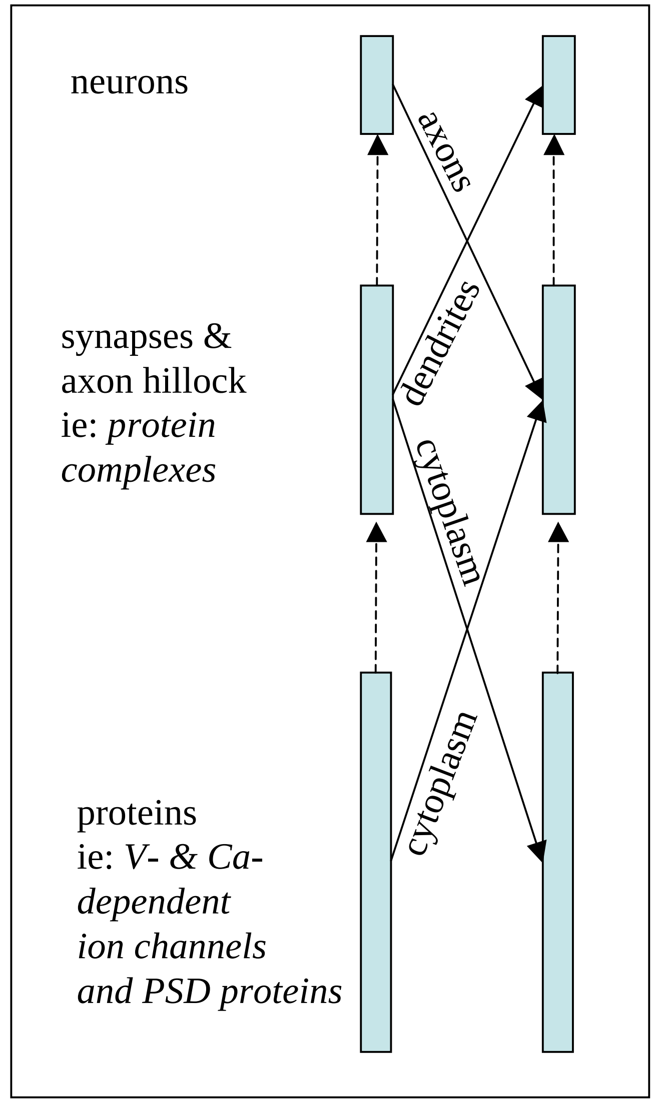


Figure: A hippocampal neuron with synapses stained for post-synaptic proteins Shank and Homer (white puncta). Overexpression of dominant negative form of Homer (Homer1a) causes loss of dendritic spines and suppression of postsynaptic responses. *Picture by Carlo Sala.*

Synapses *also* map into an overcomplete, more microscopic, space (macromolecules)

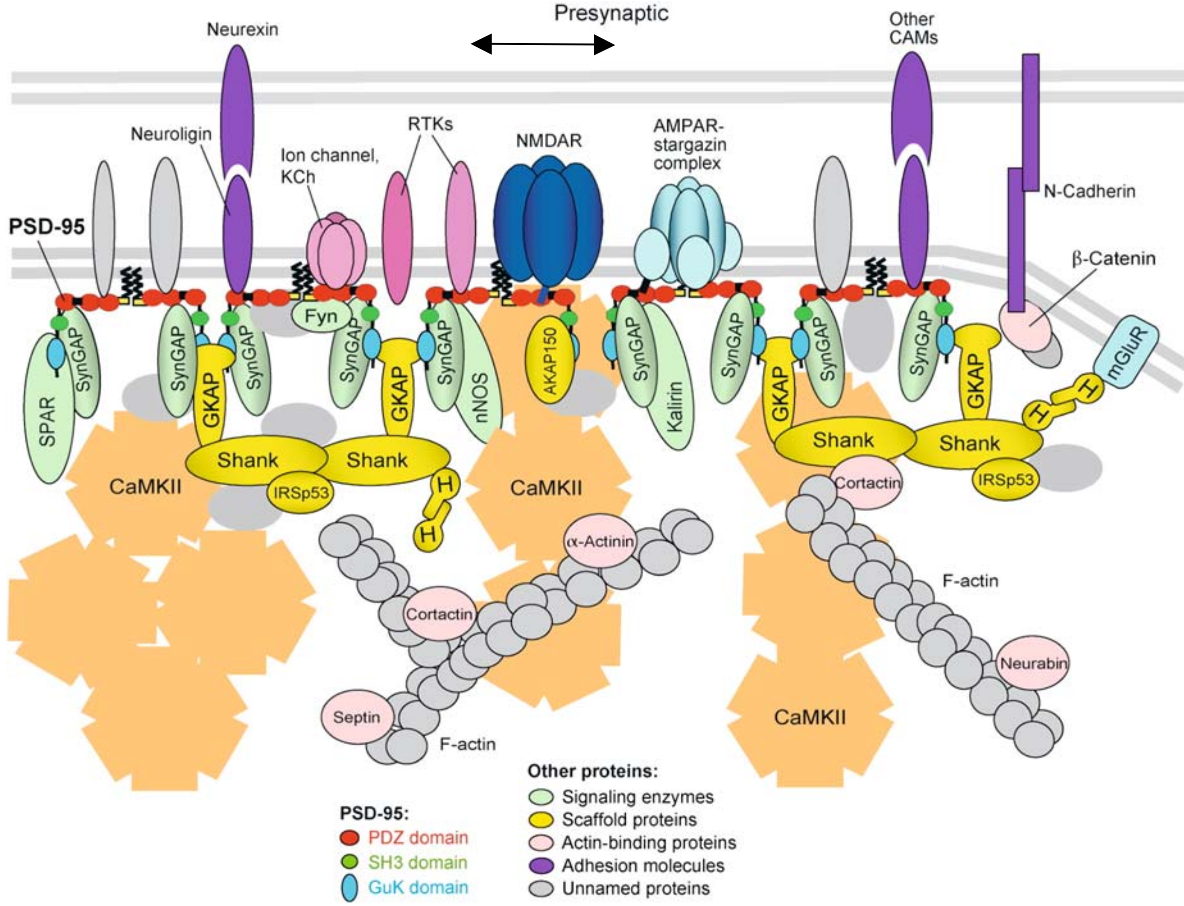




The synapse is itself a network, communicating through calcium. (calcium is the “voltage” of the PSD.)

calcium microdomain

20nm

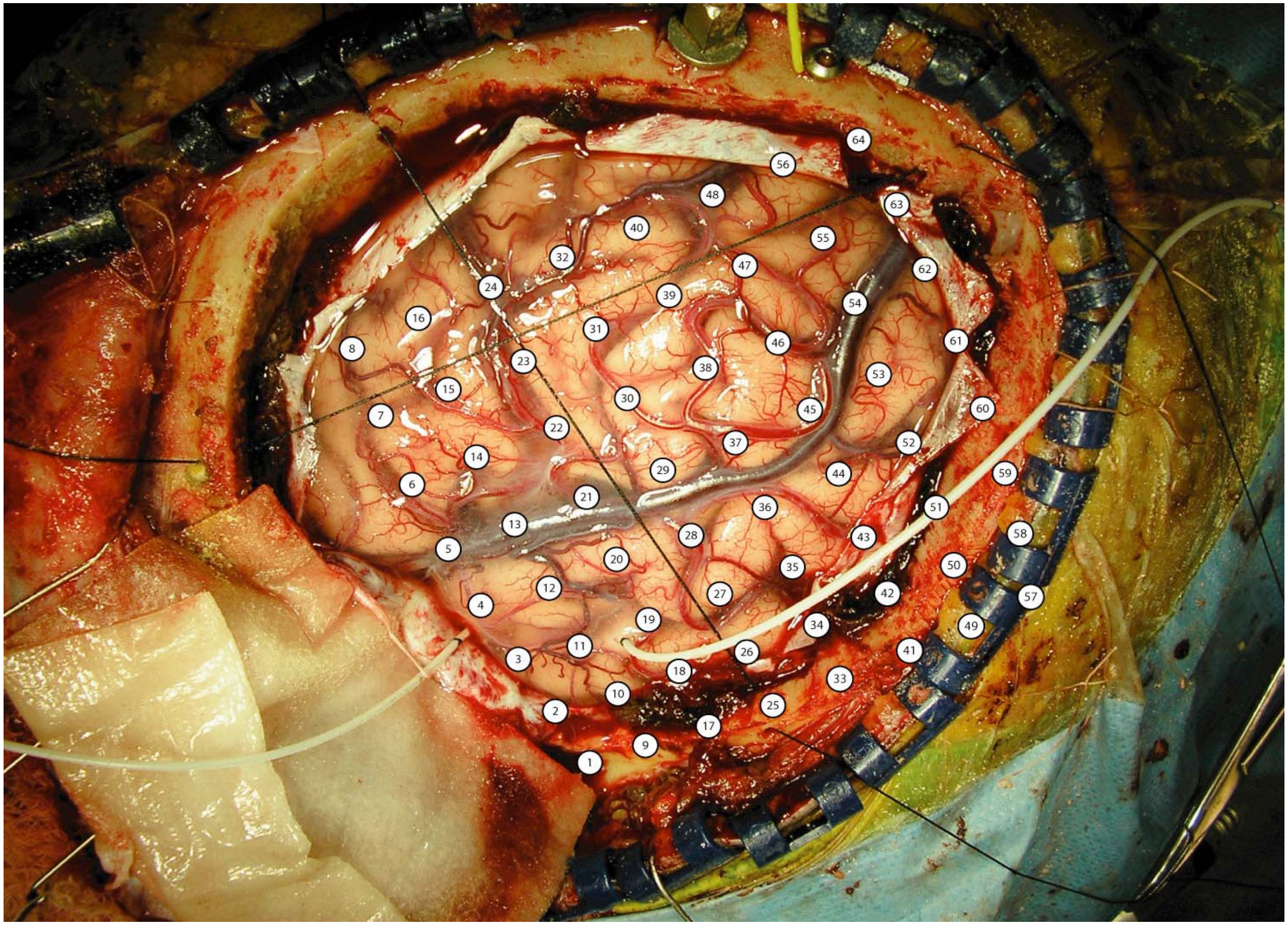


(image from Morgan Sheng)

Sheng M, Hoogenraad CC. 2007.
Annu. Rev. Biochem. 76:823–47

We could go deeper down (into the cytoplasm), but what about the brain?

Human Electro-corticogram with frequency–dependent coherences



gamma scale

theta scale

delta scale

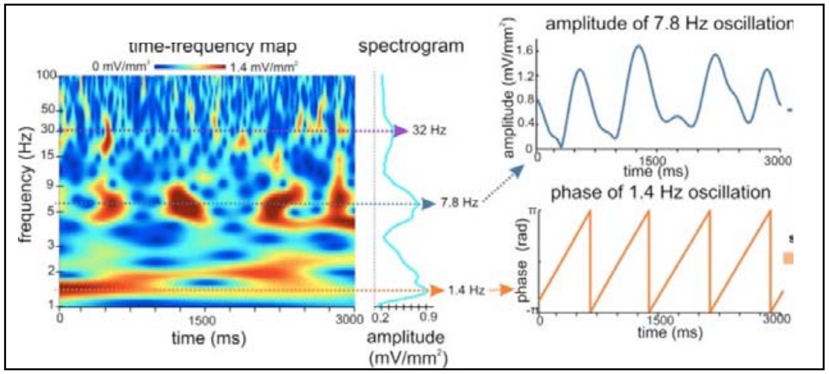
high gamma (80-150Hz) coherence: 0.3-3mm

theta (4-8Hz) coherence: 10-20mm

(from Canolty et al)

Brain networks communicate through oscillations.

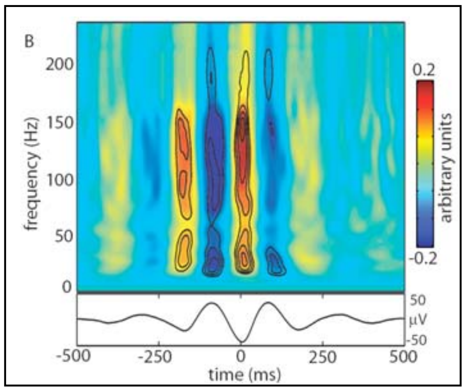
ie: *large-scale cell assemblies* map into an overcomplete space: *small assemblies* (Lakatos, Schroeder, Canolty)
ie: *small-scale cell assemblies* map into an overcomplete space: *neurons* (Fries, Koepsell)

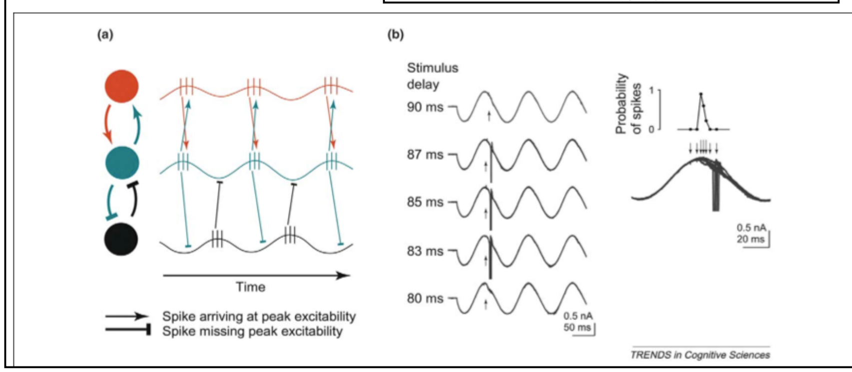


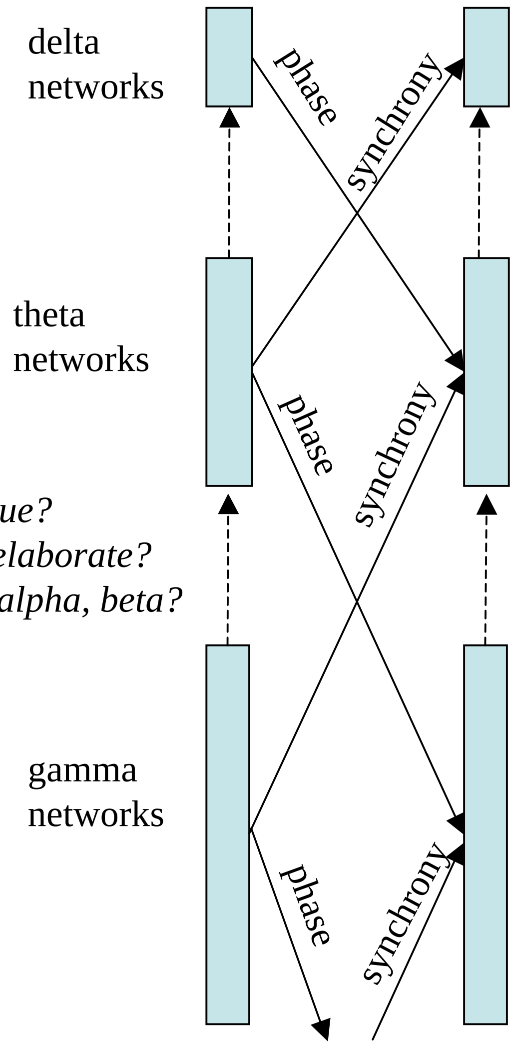
delta to theta coupling *(Lakatos et al)*

theta to hi-gamma coupling *(Canolty et al)*

gamma to spike coupling *(Fries et al)*





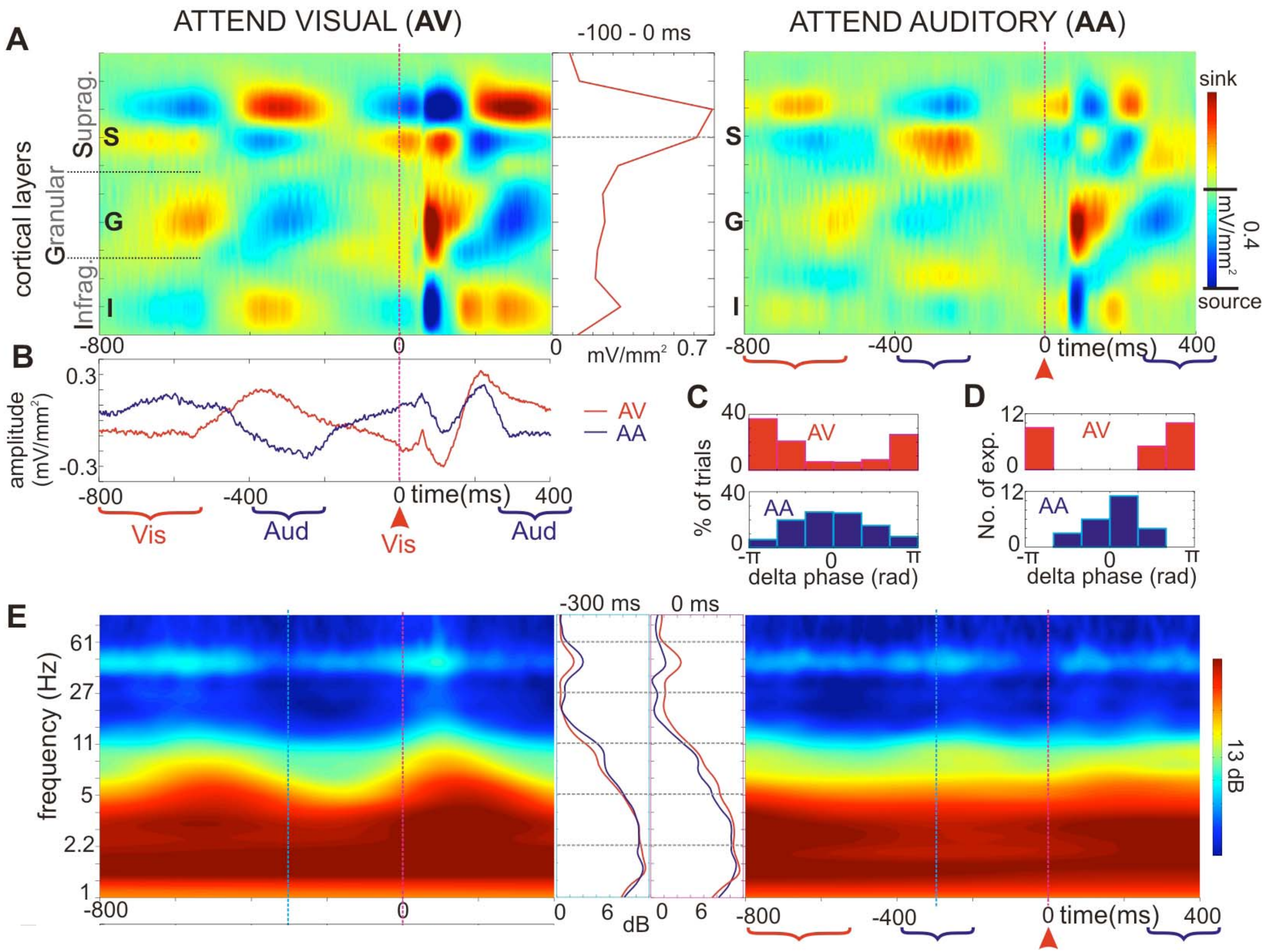


is this true?
can we elaborate?
what of alpha, beta?

Multisensory supragranular entrainment of delta in V1 by attention.

(Trial-averaged current source densities and time-freq. plots)

L II/III is pi out of phase when attending to auditory compared to attending to visual.



Theta and gamma amplitudes modulated in counterphase.

Lakatos et al. *Science* (in press)

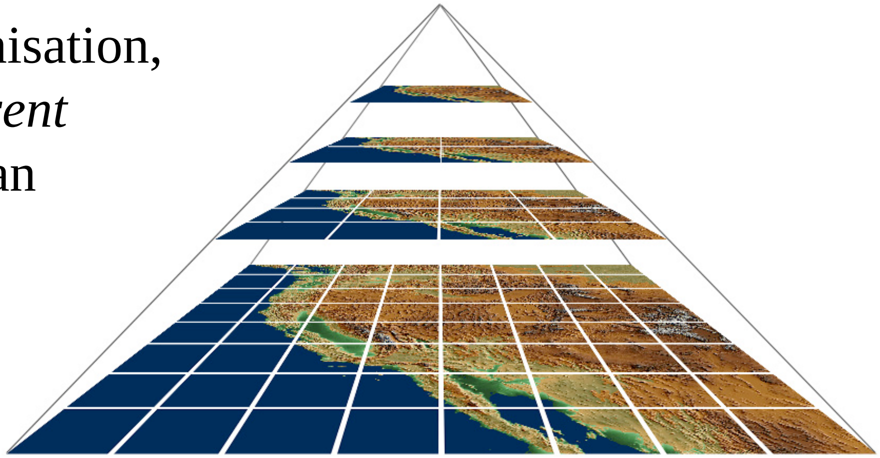
Summary

In the brain:

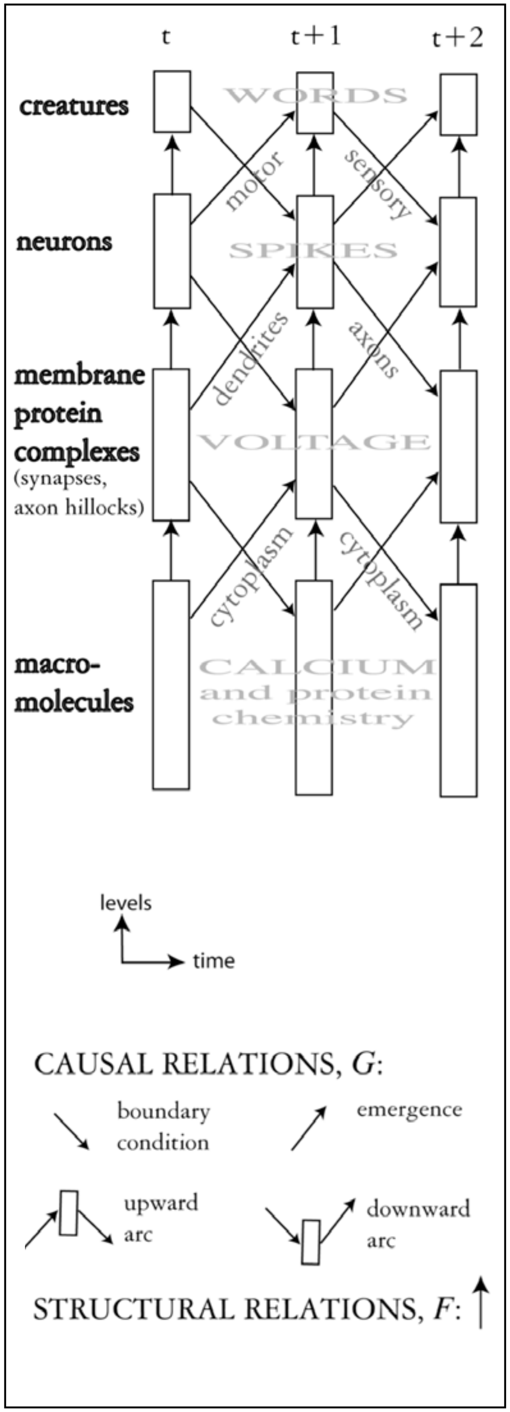
calcium is to networks of *proteins*
what *voltage* is to networks of *synapses/protein complexes*
what *spike timing* is to small networks of *neurons* (gamma circuits)
what *oscillation-phase* is to larger networks

Of course it is more complex than that, but this cartoon-view is a start.

These are not separate levels of organisation, but the *same thing expressed at different spatio-temporal resolutions*, as with an image pyramid:



(You could read my words from my calcium flows...)



Consequences of the Levels Perspective:

1. Biology consists of *networks within networks* with no “cutoff level”.
2. Modularity implies information flow is up and down, *not horizontal*.
3. The micro is an *overcomplete* space in which information can be stored.
4. A *question* is a macroscopic constraint.
5. An answer (a *memory*) is an emergence from the microscopic.
6. Emergence into *awareness* is probably emergence from the microscopic.
7. *Noise* is an experimental concept. It does not relate to reality. It is an emergence that is unwanted <u>by an experimenter</u>.
8. *Control* is a macroscopic b.c. disruptable by emergence or higher b.c.
9. What appear as loops are actually inter-level interactions.
10. *The sensorimotor loop* (eg) is inter-level and nested in the hierarchy.
11. *Reward* is an agent-centred concept which dissolves in the hierarchy.
12. *Sleep* is a chance for molecular and neural nets to converse without interference from the social network.
13. All processes are *the same thing* expressed at different resolutions.
14. There is thus *no friction* between explanations at different levels (for example, between evolution and self-organisation)

Scientific challenge:

To unify microscopic physics, biology and the modern theory
of probabilistic learning/inference (density estimation?),
in the light of these inter-level observations.

“Unfortunately, nature seems unaware of our intellectual need for convenience and unity, and very often takes delight in complication and diversity.” - Cajal

The levels perspective, which is diametrically opposed to the
von Neumann view, which still dominates, should not be depressing.

Rather, it should alert us to a different set of questions:

- Are there invariant characteristics in inter-level information flows in biology?

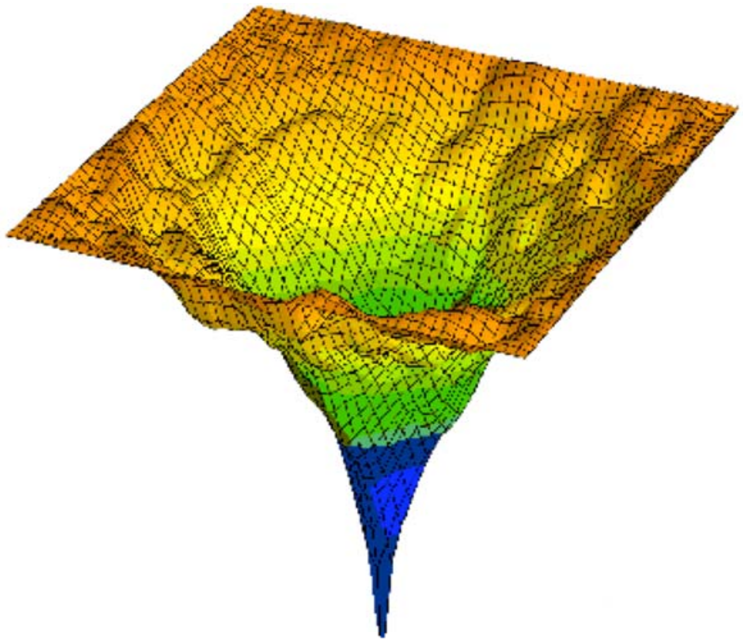
- Is there a multiresolution density estimation scheme involved?

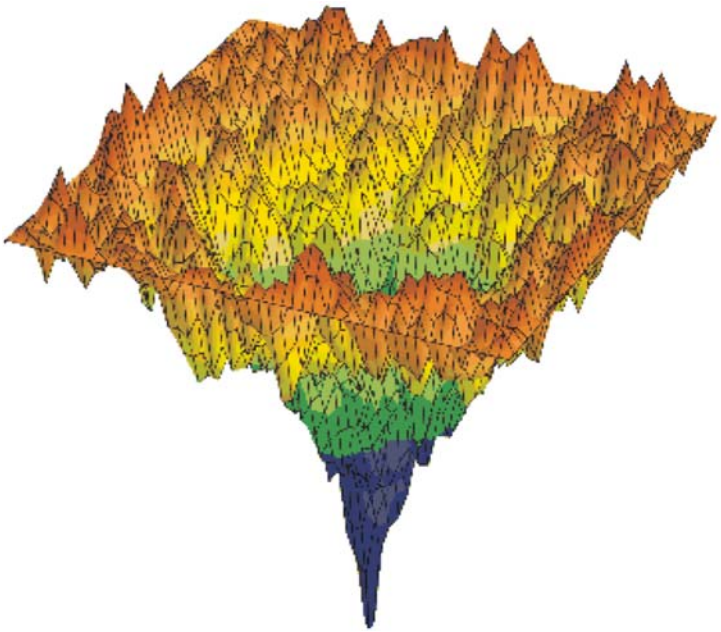
- Is there a connection to scale-invariant multi-level theories in physics (ie: Renormalisation Group)

- Are levels inter-defined? (in which case reductionism is wrong)

Thank you

Protein energy landscapes, wet and dry

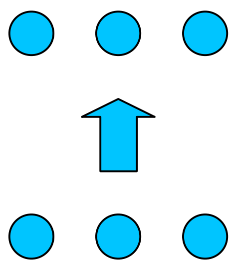




Multivariate case 1:
Independent Component Analysis.

$$\mathbf{u} = \mathbf{W}\mathbf{x}$$

$\mathbf{W}$

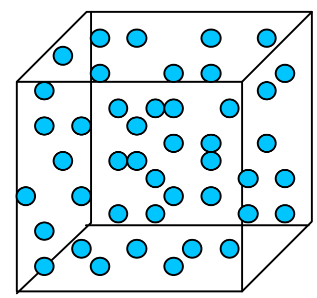


data

$$\text{Make } p(\mathbf{y}) = \frac{p(\mathbf{x})}{|\partial_{\mathbf{x}}\mathbf{y}|} \to \mathbf{1} :$$

$q(\mathbf{x})$

model

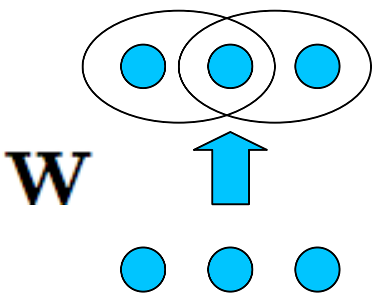


$$\Delta\mathbf{W} \propto \left(\mathbf{I} - \langle \mathbf{f}(\mathbf{u})\mathbf{u}^T \rangle_p\right) \mathbf{W}$$

Natural gradient infomax/maximum likelihood
(Bell & Sejnowski (1995), Amari, Cichocki & Yang (1996))

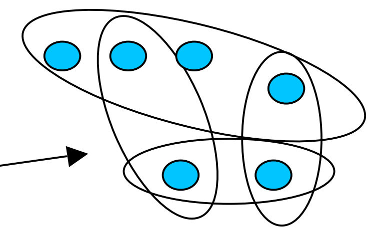
Multivariate case 2:
Dependent Component Analysis.

$$\mathbf{u} = \mathbf{W}\mathbf{x}$$



$$q(\mathbf{u}) = \text{whatever}$$

but if it is a loopy graphical model, like



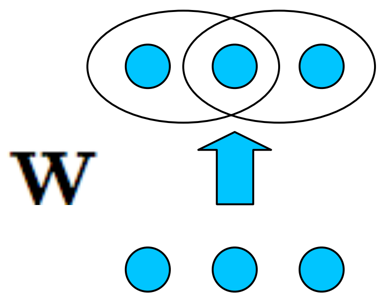
we get the gradient of the partition function (so we need to sleep)

$$\Delta \mathbf{W} \propto \left(\langle \mathbf{f}(\mathbf{u})\mathbf{u}^T \rangle_q - \langle \mathbf{f}(\mathbf{u})\mathbf{u}^T \rangle_p \right) \mathbf{W}$$

Hinton et al, A new view of ICA, *Proc. ICA* (2001)
Bell A. The co-information lattice, *Proc. ICA* (2002)

Multivariate case 2:
Dependent Component Analysis.

$$\mathbf{u} = \mathbf{W}\mathbf{x}$$



The Gibbs distribution:

$$q(\mathbf{u}) = \frac{1}{Z} e^{-E(\mathbf{u})}$$

gives us this very Boltzmann Machine-esque form

$\partial_{\mathbf{W}} \log Z$ $\quad$ $\partial_{\mathbf{W}} E(\mathbf{u})$

$$\Delta \mathbf{W} \propto \left(\langle \mathbf{f}(\mathbf{u})\mathbf{u}^T \rangle_q - \langle \mathbf{f}(\mathbf{u})\mathbf{u}^T \rangle_p \right) \mathbf{W}$$

Hinton et al, A new view of ICA, *Proc. ICA* (2001)
Bell A. The co-information lattice, *Proc. ICA* (2002)