Quantum Mechanics and Helmholtz Machines

George Chapline

Sept 10, 2009

References

"Quantum Mechanics, Pattern Recognition, and the Mammalian Brain"
AIP Conference Proceedings 1051, 15 (2008)

"Quantum Mechanics and Pattern Recognition"
Int. J Quantum Information 2, 295 (2004)

"Topological Quantum Image Analysis"
Proceedings of SPIE, 7342C-1 (2009)

Revisionist History

- Quantum mechanics appeared *dei gratia* in the 1920's as a result of the efforts of Heisenberg, Pauli, and Born to understand atomic spectra of multi-electron atoms
- However, in the 19th century Herman Helmholtz had a precursor premonition of the appearance of quantum mechanics

Helmholtz's Brilliant Idea

Helmholtz suggested that the human brain is a statistical inference engine whose function is to estimate the probabilities for alternative explanations of sensory data.

Helmholtz was correct!

T. Yang and M. Shadlen Nature 447, 1075 (2007)
U. Wash. Primate Research Center

Experiments with rhesus monkeys have shown by directly measuring the activity of neurons that process information from the visual cortex that the monkey brain uses standard probabilistic inference to combine evidence from multiple visual clues to select among alternative explanations for multiple inputs the one that is most likely to be correct.

Bayesian Pattern Recognition

- Best classification procedure is to choose class α such that posterior probability $P(\alpha \mid x)$ is largest, where

$$P(\alpha \mid x) = \frac{p(\alpha)\, p(x \mid \alpha)}{\sum_{\beta} p(\beta)\, p(x \mid \beta)}$$

“Physics” Interpretation for Posterior Probabilities

“Energy” cost for explanation α :

$$E_\alpha \equiv -\log[p(\alpha)p(\alpha \mid x)]$$

Then

$$P(\alpha) = \frac{e^{-E_\alpha}}{\sum_\alpha e^{-E_\alpha}}$$

Minimizes “free energy”

$$F(x) = \sum_\alpha [E_\alpha P(\alpha) - (-P(\alpha)\log P(\alpha))]$$

Maximum Likelihood Estimator

- The best model probability distribution $P_\theta(\alpha)$ minimizes the difference between the model free energy $F(x,\theta)$ and the exact free energy

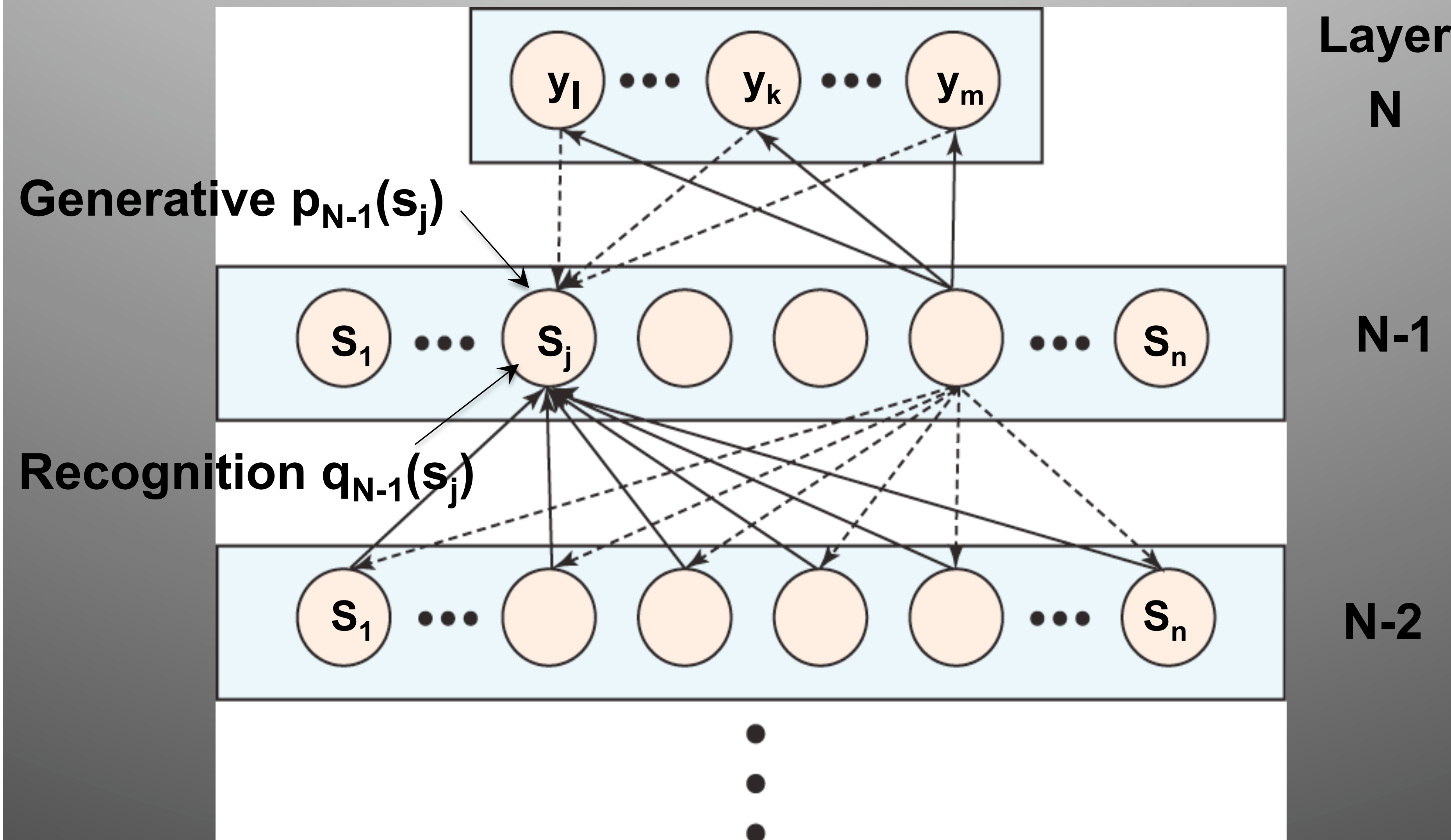
$$F(x) = F(x,\theta) - \sum_{\alpha} P_\theta(\alpha)\log[P_\theta(\alpha)/P(\alpha)]$$

Note: *F*(*x*) is just the cost (in bits) to needed to describe the state of the network

Bayesian Networks

- In Bayesian networks the “explanations” for input data are represented by the activation levels of the nodes in a layered network
- A hierarchical interpretation for input data can be introduced via the Markov property: conditional independence of node activations within in a given layer
- Practical implementation is difficult because Markov chain Monte Carlo methods are typically required to provide world model

A Helmholtz machine consists of two intertwined Bayesian networks



A Helmholtz machine constructs a Bayesian model of the world from scratch by "self-learning"

The Bayesian world model is contained in a "top-down" generative model that is used to generate activities in the network – which at first bear little resemblance to the activities generated by real world sensory inputs - but eventually with "training" come to closely resemble the activities generated by real world inputs

Wake-sleep Algorithm

G. Hinton, P. Dayan, B. Frey
Science 208, 1158 (1995)

1) In the "awake" phase the generative connection strengths are adjusted so that the generative probabilities $p_n(s_j)$ are better able to reproduce the activities produced in each layer by the recognition probabilities when the input layer is activated with real world inputs
2) In the "sleep" phase the recognition connection strengths are adjusted so that the free energy $F(x,\theta)$ calculated from the recognition probabilities is as similar as possible to the free energy $F(x)$ calculated from the a priori probabilities.

Historical note

- Schrodinger realized (1930) that in any discrete probabilistic process where the probabilities which describe the transition from one layer to the next layer satisfy the identities for Markov probabilities, there is a time symmetric representation

"Schrodinger" Representation for a Bayesian Network

In the time symmetric representation for a Bayesian network the transition probability to go from state **a** (n) to **a**(n+1)can be expressed in the form

$P_{forward}\,[\mathbf{a}(n+1)\,|\,\mathbf{a}(n)] = \phi_+^{-1}(\mathbf{a},n)\mathbf{K}(n,n+1)\phi_+(\mathbf{a},n+1)$

and to go from **a**(n+1) to **a**(n)

$P_{backward}\,[\mathbf{a}(n)\,|\,\mathbf{a}(n+1)] = \phi_-(\mathbf{a},n)\mathbf{K}(n,n+1)\phi_-^{-1}(\mathbf{a},n+1)$,

where $\phi_+ \equiv A\exp(S)$ and $\phi_- \equiv A\exp(-S)$ play the role of the complex wavefunction Aexp(iS)

Quantum Mechanics and the Helmholtz Machine

- *The Schrodinger representation for a Bayesian network provides an **exact** mapping of the Helmholtz machine into quantum dynamics*

- "Wake-sleep" algorithm can be viewed as analogous to the "inverse problem" of finding the Hamiltonian for a multi-channel Schrodinger equation (Newton-Jost eq's.)

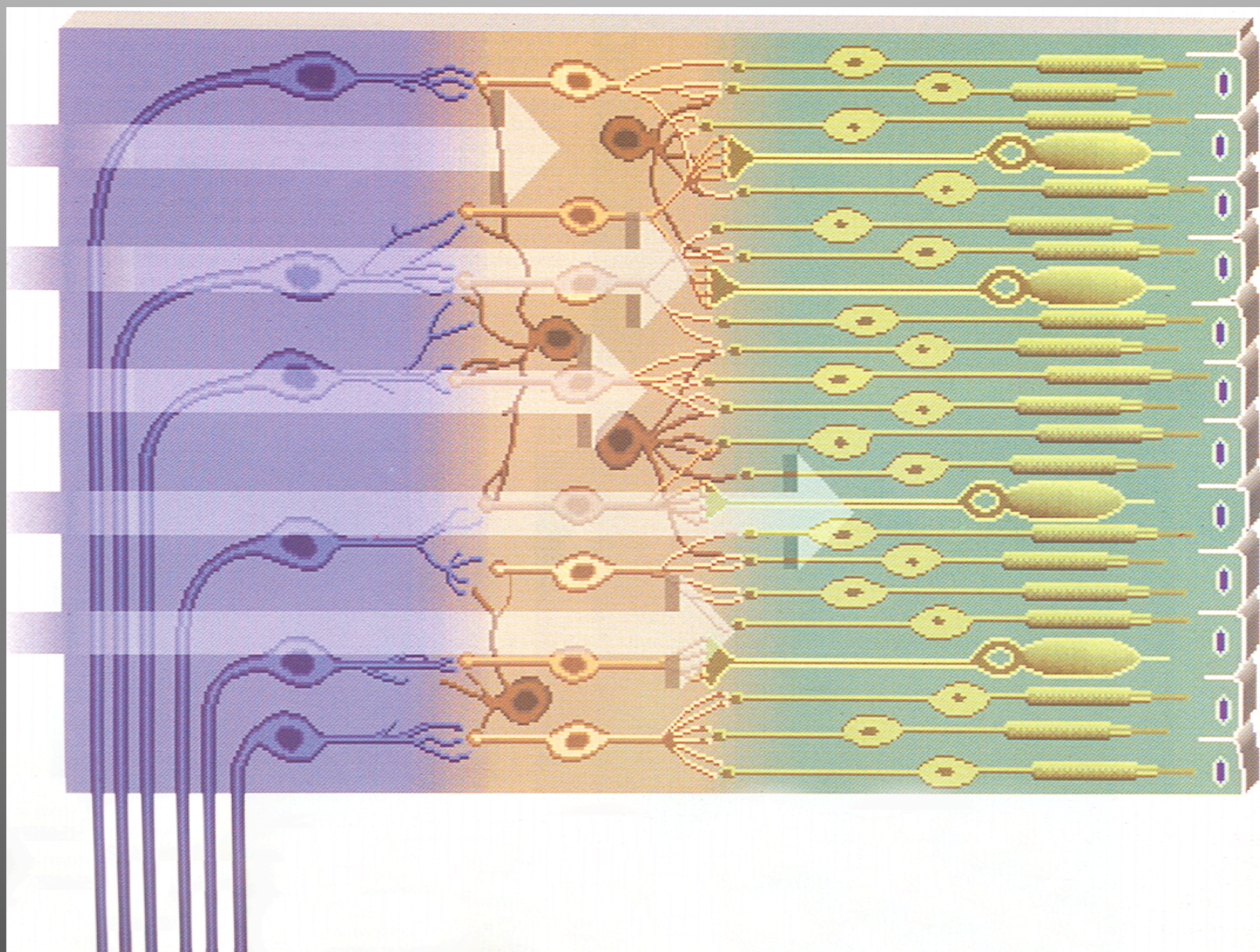
The Helmholtz Machine and Path Integrals

- The mapping of the Helmholtz machine into quantum dynamics raises the question as to whether mammalian pattern recognition can also be regarded as an extension a la Dirac/Feyman of classical mechanics.
  - The bit *cost E(a)* of describing the state of the network apparently plays the role of an "action"
- The actions that seem to be most relevant to mammalian pattern recognition are topological; e.g.

$$S_{top} = \int d^2\sigma Tr \varepsilon^{\alpha\beta} \left\{ \frac{1}{4} F^{\alpha\beta} [\Phi^+, \Phi] - D_\alpha \Phi^+ D_\beta \Phi \right\}$$

Network: Comput. Neural Syst. 8 , 185 (1997))

The mammalian brain consists of self-organized layers



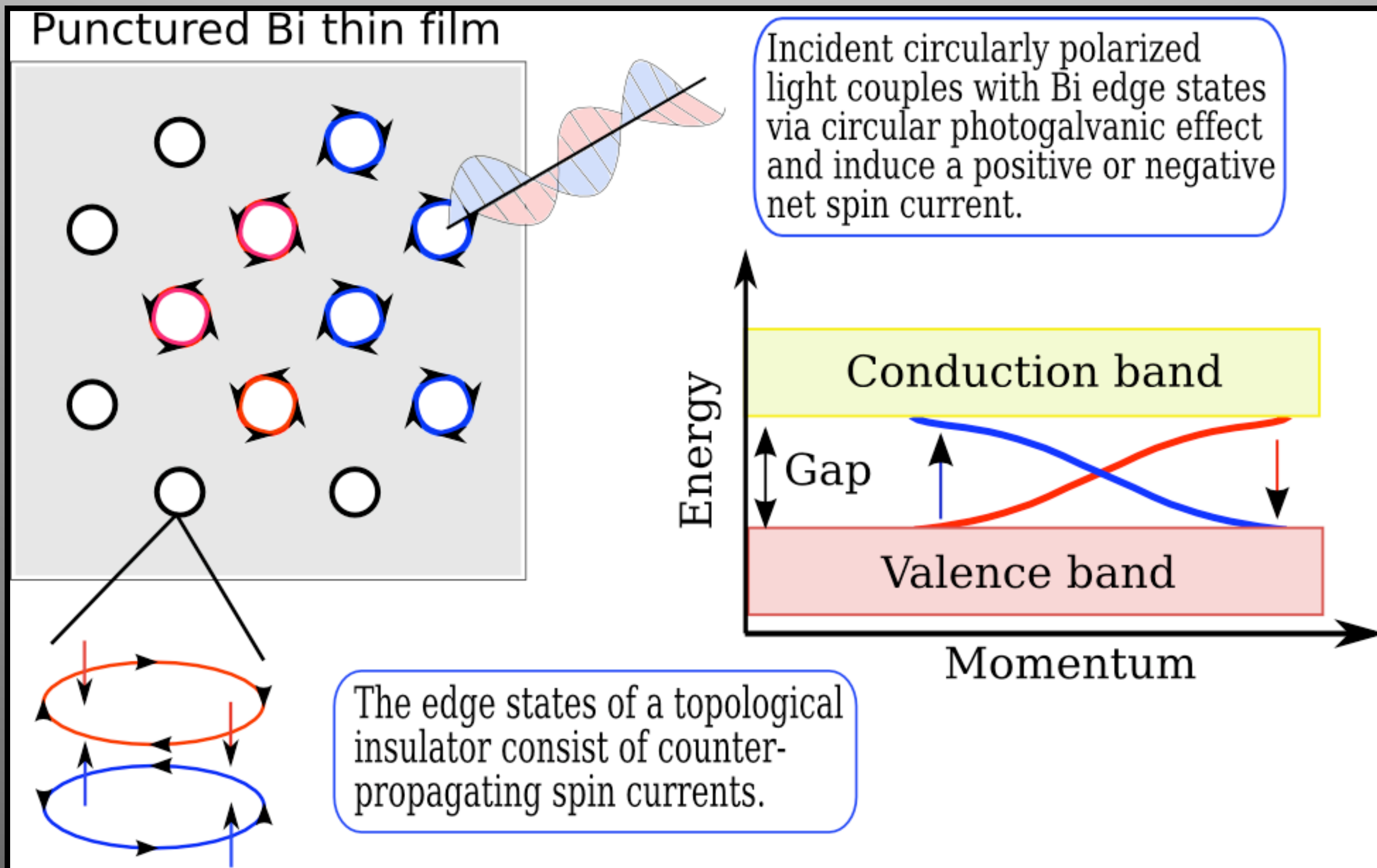
What are quantum computers good for?

- Although pattern recognition techniques don’t necessarily involve Bayesian analysis, it is remarkable that the mammalian brain uses pattern recognition methods closely related to (topological) quantum mechanics
- It’s a good bet that it will turn out that practical usefulness of quantum information processing lies in its ability to simulate mammalian-like pattern recognition

Quantum Image Analysis

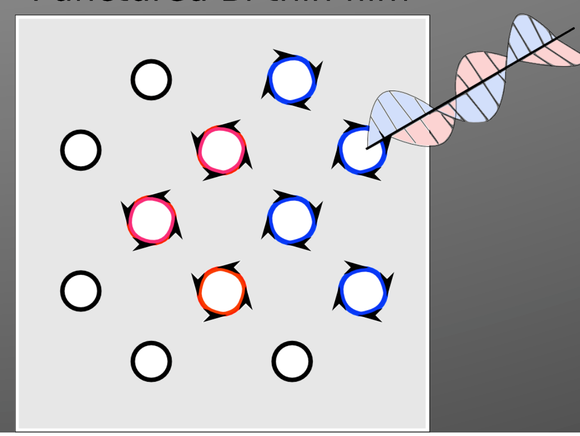
- One of the greatest challenges in computer vision is recognizing objects in visual images in real time.
  - Comparing an object in an image with a standard template requires topological transformations that are very difficult to implement with a classical computer
- Topological quantum information processing may offer a unique approach to this problem
- It has recently been discovered that materials with strong spin orbit interactions have topological discrete states
  - Thin bismuth films patterned with an array of nano-sized holes may provide a path to practical implementation

Manipulating topologically protected edge states in bismuth thin films



Prediction: In the future human object recognition will be replaced with integrated quantum sensors/information processing

Punctured Bi thin film



"Quantum Mechanics, Pattern Recognition, and the Mammalian Brain" (G. Chapline CASYS 2007 Award)

Lawrence Livermore
National Laboratory

Where do we stand?

- Quantum mechanical evolution is completely equivalent to the use of a Helmholtz machine for pattern recognition if one uses joint probability distributions to describe the states of the network within each layer
  - Non-local equal time correlations can be explained in terms of the time symmetry of the Helmholtz machine
  - Quantum "uncertainty" corresponds to the fact that pattern recognition involves ambiguities
  - Measurements correspond to information fusion

What is the meaning of consciousness?

- The fact that pattern recognition in mammals has a "quantum mechanical" interpretation suggests that consciousness has some natural interpretation in terms of quantum mechanics

- Does consciousness has something to do with quantum coherence; i.e., the emergence of classical behavior in a quantum system ?

Wigner was almost right!