David Kleinfeld

The mean-field brain

Modelling Brain Function: The World of Attractor Neural Networks **Daniel J Amit 1989 Cambridge University Press 504pp £25.00 hb**

MY THESIS advisor handed me a copy of John Hopfield's now seminal article 'Neural networks and physical systems with emergent selective computational properties' in the autumn of 1983 and added the remark 'This is just a side-line . . . his real meat and potatoes is still electron transfer (in biology)'. Some side-line. Hopfield suggested, in terms of a specific Hamiltonian, a correspondence between networks of neurons and spin systems. This work ignited the imagination of a very smart group of statistical mechanicians, Daniel Amit among them, and led them to ply their trade on neural networks.

What kind of world is 'the world of attractor neural networks' and what is its relevance to 'brain function'? The book begins with an examination of the assumptions and ideas leading to the Hopfield model. The prescription of what constitutes a physicist's neuron comes across clearly. We never quite learn what constitutes a biologist's neuron. Try one of the excellent texts, such as Shepherd's *The Synaptic Organization of the Brain* (OUP 1990), if you want to learn about neurobiology. The second and third chapters review aspects of cooperative dynamics in complex systems. Amit gives a thorough explanation of the essential physics involved in the emergence of attractors in networks whose dynamics are governed by feedback connections among the neurons. His illustration of asynchronous versus synchronous dynamics provides a nice introduction to the concepts of flow through state space and stable states.

The 'meat and potatoes' of the book begins with the fourth chapter. As noted elsewhere by Toulouse, two major pieces of research shaped the development of research on attractor networks. The first, by Amit, Gutfreund and Sompolinsky, showed that the Hopfield model could be solved exactly within the framework of mean-field theory. This work introduced the concept of noise temperature into attractor networks and delineated the rich set of phases that occur as a function of the temperature and the number of memory states. Amit brings a sense of perspective to his review, along with many intermediate results (particularly regarding the use of the replica trick) that make the book easier to read than the original papers. Alternative methods of analysis are not brought out, but can be found in the book *Spin Glass Theory and Beyond* by Mezard, Parisi and Virasoro (World Scientific 1987). Lastly, Amit brings together a host of important extensions to the original Hopfield model that tend to be scattered throughout the literature.

The second piece of major research, by the late Elizabeth Gardner, showed that the tools of statistical mechanics could be used to analyse the space of connectivity among physicists' neurons. This was a forceful idea. While most previous work on attractor networks had been concerned with finding the stable states that result for a given prescription of the neuronal connectivity, Gardner turned the problem around and solved for the space of connections that gave the desired memory states. Amit presents a succinct discussion of Gardner's results.

Amit approaches the relationship of attractor networks to higher brain function on two levels. The first is that of psychology – memory states as gestalts. The concept of stable states is applied to cognition in Amit's discussion of ideation in schizophrenic patients. Rather than leave the subject at the level of analogies, Amit makes a step toward abstract computation in terms of attractor networks that form temporal associations. The tie to abnormal psychology is not obvious, but at least we are left with a glimpse of how cognitive states can evolve.

The second level of approach to brain function is an attempt to wed attractor models to the experimental data that concerns the architecture and physiology of the cortex. The Hopfield model has features that are in apparent conflict with this body of data. Amit suggests necessary guidelines to bridge these differences and summarises tentative steps in this direction, e.g. the incorporation of neurons with purely excitatory or inhibitory connections and a plausible scheme for cortical-like oscillations within groups of such neurons. Other issues, such as the relatively low local levels of neuronal activity that occur in cortex but are difficult to achieve in the models, were approached only after the book was written.

When all is said and done, the impressive body of work that Amit discusses has not modelled brain function in a way that makes testable predictions for experimental neuroscience. But, accepting Amit's remark that 'attractors and their close relatives are the main message we bring from physics (to biology)', there is a chance that this work may lead to a necessary dogma for computation in large nervous systems. If you're interested in this business, read Amit's book. As a technical review, it is unsurpassed.

David Kleinfeld as at AT&T Bell Laboratories, Murray Hill, NJ, USA