Bioelectrochemistry and Bioenergetics, 28 (1992) 425–434
A section of *J. Electroanal. Chem.*, and constituting Vol. 343 (1992)
Elsevier Sequoia S.A., Lausanne

JEC BB 01527

On fitting dielectric spectra using artificial neural networks

Douglas B. Kell * and Christopher L. Davey

Department of Biological Sciences, University College of Wales, Aberystwyth, Dyfed SY23 3DA (UK)

(Received 20 September 1991; in revised form 30 December 1991)

Abstract

In biological dielectric spectroscopy, where dispersions are substantially broader than that expected from a purely Debye-like process, it is not always possible, because of technical limitations, to obtain data over a wide enough range of frequencies to encompass the entire dispersion(s) of interest. Similarly, because of the breadth of the dispersions, it is common to seek to characterize the dielectric behaviour of interest by means of the Cole–Cole function. Whilst it is possible to fit dielectric data to this equation using appropriate non-linear least-squares methods, these methods are computationally rather demanding, and must be performed iteratively for each set of data. We show here, for the first time, that it is possible to train an artificial neural network to learn to extract the parameters of the Cole–Cole equation from small sets of dielectric data (permittivities measured at various frequencies) which can thus give an essential instantaneous output of the limiting permittivities at frequencies that are both high and low with respect to the characteristic frequency.

INTRODUCTION

In the dielectric spectroscopy of biological and other systems, it is usual to find areas of strong frequency dependence, known as dielectric dispersions, in which the measured permittivity decreases with increasing frequency, with a shape (when the frequency is plotted logarithmically) approximating an inverse sigmoid [1–12]. To characterize the behaviour of the system of interest quantitatively, one fits the measurements (permittivity and conductivity at different frequencies) to an appropriate equation, that proposed by Cole and Cole [13] being perhaps the most

* To whom correspondence should be addressed.

popular in biological work. The Cole–Cole equation is a modification of the Debye [14] formulation of molecular dielectric behaviour which contains, in addition to the dielectric increment $\Delta \epsilon$, the characteristic frequency f_c and the high-frequency permittivity ϵ_h, an empirical parameter, the Cole–Cole α, which can be used to describe (if not to explain) the fact that real dielectric spectra are much broader than those due to a simple Debye-like dispersion. Whilst the Cole–Cole α has no theoretical justification (although it is widely interpreted in terms of a distribution of relaxation times), Schwan [15] showed that a great many types of relaxation–time distribution could accurately *fit* the Cole–Cole function. In addition, the Cole–Cole function permits one to extract the parameters describing an *entire* dielectric dispersion, even if, for technical reasons, one cannot measure over the whole frequency range across which it occurs. For these and other reasons, the Cole–Cole formulation remains very popular as a means of characterizing the dielectric properties of biological systems.

We have shown that the radio-frequency dielectric properties of biological cells at one or two appropriate frequencies can be used as a rapid (on-line) method for measuring levels of cellular biomass in fermentors and elsewhere, and for this purpose we have constructed a high-resolution dielectric spectrometer, capable of measuring in the range 0.2–10 MHz [16–20]. The method relies upon the fact that the β-dielectric dispersion exhibited by all biological cells is dominated by the charging of their plasma membrane(s), and that intact biological cells, but nothing else likely to be found in a fermentor, possess relatively non-conducting plasma membranes [16]. However, this approach requires that (i) at least one of the frequencies of measurement is low with respect to f_c, and (ii) the f_c does not change appreciably during the fermentation of interest. The f_c of the β dielectric dispersion depends upon both the internal and external conductivity [15], and whilst the former is likely to be a relatively constant property (at least for cells in a given medium) the latter may well change significantly [21].

One way round the above problem would be to take measurements at a number of frequencies and fit the data to the Cole–Cole equation, thereby obtaining the dielectric increment which is what truly reflects the biomass present [16–20]. Methods for fitting dielectric data are usually based on non-linear least-squares algorithms [1,22], and we ourselves (see later) have found that the popular Levenberg–Marquardt algorithm [23,24] provides excellent fits to real dielectric spectra. However, these methods are computationally rather intensive, and for the fermentor example would have to be carried out for every data set at every time point.

Artificial neural networks (ANNs) consists of highly interconnected parallel-processing elements known as nodes, which are arranged in layers representing a set of inputs, one or more so-called hidden layers and a set of outputs. Each node acts to sum its own inputs (which are the outputs of the elements of previous layers), and the sum is passed through a transfer function (which must be continuously differentiable and is normally non-linear) to the element(s) in the next layer. In the classical version, the transfer function is sigmoidal (via the

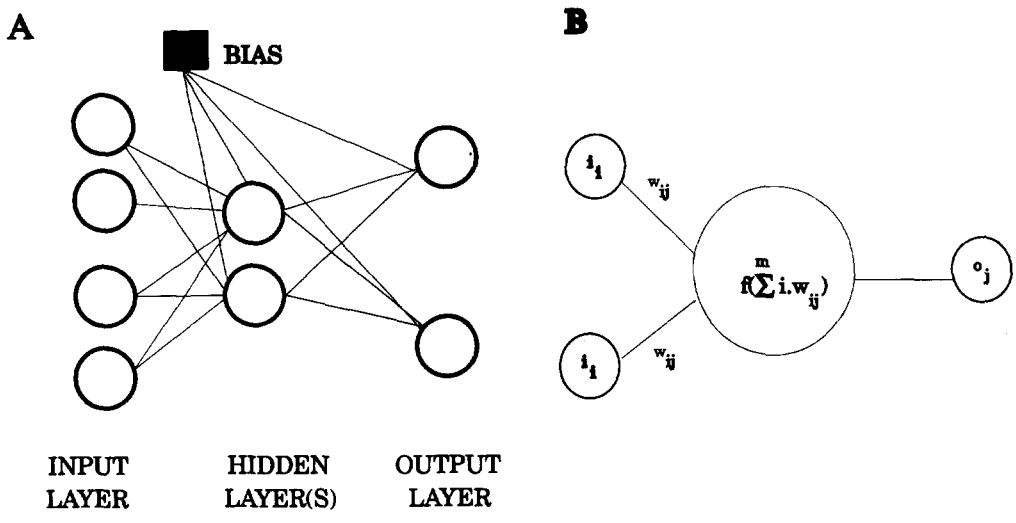

Fig. 1. The principle of a classical feedforward neural network. (A) The construction of a 4–2–2 neural network in which inputs and outputs are connected to each other via one or more hidden layers. Layers other than the input layer may also be connected to a bias. In the architecture shown, adjacent layers of the network are fully interconnected, although other architectures are possible. (B) Information processing by a neuron. An individual neuron sums its inputs from neurons in the previous layer, transforms them via a transfer function, and outputs them to the next neurons to which it is connected.

exponential term) and is normalized between 0 and 1. The output o_j of node j is given by

$$o_j = \frac{1}{1 + e^{-(x/\gamma)}} \tag{1}$$

where

$$x = \sum_{i=1}^{I} (w_{ij} o_i + \theta_j) \tag{2}$$

In eqn. (2), θ_j is a bias term, o_i is the output from the ith node of the previous layer, w_{ij} represents the so-called weight or strength between node i and node j and γ is known as the gain. Other popular functions include the sinh, tanh and sine functions, and the general principles of such networks are illustrated in Fig. 1.

It is possible to train such networks by initially setting the weights and biases to small random values, presenting the networks with known inputs and outputs, and comparing the output of the net with the "true" (known) outputs. By adjusting the weights using information based on the difference (the error e_i) between the output of the net and the true values, a principle known as the back-propagation of error (or, more simply and more commonly, back-propagation or back-prop), it is therefore possible to train the network accurately to deliver a desired output when

presented with a novel (previously unseen) input [25]. This process is repeated from the output through each hidden layer to the input. The actual weight updates for this so-called delta rule are

$$w'_{ij} = w_{ij} + \mathrm{LR}\, e_i x_{ij} + M m_{ij} \tag{3a}$$

$$m'_{ij} = w'_{ij} - w_{ij} \tag{3b}$$

where LR and M are user-defined values of the so-called learning rate and momentum respectively. In this way, weights are changed according to both the error and the input to the connection of interest. Training can be continued until a defined root mean square error (between the "true" outputs of the training set and the outputs of the network) is obtained, or simply for a fixed number of presentations of the training set.

The great interest in ANNs has therefore been aroused by their ability to act as pattern recognition or signal processing elements, among other applications, and ANNs are the subject of a number of books [25–35]. This is not the place to review in detail what is a very substantial literature, but the following outline comments are in order. First, it has been shown that (given sufficient time) an appropriately trained network with sufficient nodes can simulate any function to an arbitrary degree of accuracy [36]. Second, whilst back-propagation methods are nowadays considered to be computationally rather inefficient, and many other possible algorithms and architectures exist [37–42], they remain the most popular methods and their behaviour is reasonably well understood. Third, although the training of a network may be a lengthy procedure, *once trained the network processes the input into the output virtually instantaneously*. Fourth, it is widely believed that, owing to the parallel distributed processing that they effect, the networks are robust with respect to both noise in the inputs and "damage" to the neurons [43,44]. Finally, for our present purposes, it is pertinent to note that ANNs have been used with success in the analysis of nuclear magnetic resonance (NMR) spectra [45] and fluorescence spectra [46], and in chromatography [47].

The question therefore arose as to whether it might be possible to train an ANN using simulated dielectric data (permittivity values at a number of fixed frequencies) as the inputs and the parameters of the Cole–Cole equation that had generated the data as the outputs, and thereby teach the network to give (say) the dielectric increment of a dispersion when presented with a set of dielectric data that have (realistically) variable values of ϵ_h, f_c and α. In the present work, we show that this is indeed the case.

METHODS

All simulations and networks were run on a Viglen Vig III 80386-based PC-compatible microcomputer, incorporating an 80387 coprocessor. Simulations were run using programs written in-house in Microsoft QuickBasic (Version 4.5).

The neural networks were produced and run using the Neural Works Explorer package (Scientific Computers, Burgess Hill, UK), whilst non-linear least-squares fitting (according to the Levenberg–Marquardt algorithm) routines were performed and displayed, together with other plots, using GraFit version 2 (Erithacus Software, Staines, UK).

RESULTS AND DISCUSSION

Figure 2 gives an illustration of the present problem of interest. Figure 2(A) simulates the dielectric properties of a system in which $\Delta\epsilon$, ϵ_h and α are fixed (at values of 500, 100 and 0.1 respectively), whilst f_c is varied from 0.3 to 0.7 MHz. The effect on the permittivity measured at 0.3 MHz, a typical frequency of measurement in this application [16–20], is striking, with the apparent permittivity (which one might take as a measure of the biomass) varying from 350 to nearly 500. As mentioned, the dielectric *increment*, which is the true measure of biomass, is constant under these conditions.

Similar curves (in which the permittivity at 0.3 MHz varies under conditions of constant $\Delta\epsilon$) were easily generated as a function of changes in ϵ_h (which would occur due to changes in the gas hold-up of a culture) or in α (caused, for instance, by changes in the morphology or degree of heterogeneity of a culture), but are not displayed herein.

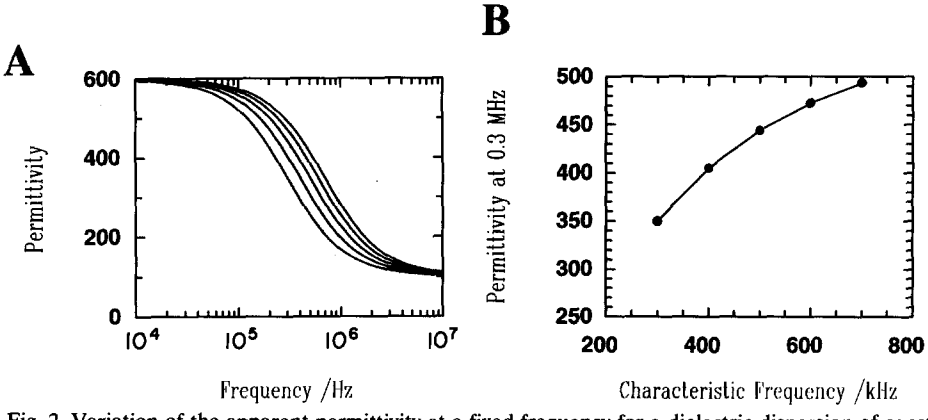

Fig. 2. Variation of the apparent permittivity at a fixed frequency for a dielectric dispersion of constant dielectric increment, high-frequency permittivity and Cole–Cole α. (A) Frequency dependence of the permittivity. Simulations of eqn. (4) were performed using the following values for the parameters: $\epsilon_h = 100$, $\Delta\epsilon = 500$ and $\alpha = 0.1$. The characteristic frequency f_c was varied in steps of 0.1 MHz between 0.3 and 0.7 MHz as indicated. (B) Dependence of the permittivity at 0.3 MHz on the characteristic frequency. Parameters were as in (A).

We therefore created various data sets for training an ANN to fit dielectric spectra to the Cole–Cole equation. The Cole–Cole equation [13] describing the variation of (the real part of the) permittivity with frequency is

$$\epsilon_f = \epsilon_h + \frac{\Delta\epsilon\left[1 + (f/f_c)^{1-\alpha} \sin(\alpha\pi/2)\right]}{1 + 2(f/f_c)^{1-\alpha} \sin(\alpha\pi/2) + (f/f_c)^{2-2\alpha}} \tag{4}$$

This equation was used to generate 500 sets of data for ϵ at ten different frequencies, logarithmically spaced in the range 0.2–10 MHz, using randomized values of the parameters in the range $0 < \Delta\epsilon < 3000$, $10^5 < f_c < 10^7$, $0 < \epsilon_h < 200$ and $0 < \alpha < 0.5$. Preliminary experiments (not shown) showed that suitable values of the learning rate and momentum were 0.3 and 0.4, and since the use of too many nodes in the hidden layer can cause the net to learn the training set (specifically, as opposed to learning and generalizing from it) we first used a single hidden layer with 15 nodes, i.e. a 10–15–4 topology, and the tanh transfer function. We also used an architecture in which the input nodes were connected directly to the output layer, in addition to their connection via the hidden layer, an architecture which we found (data not shown) to give better results. The net was shown the training set (in random order) for a total of 20000 examples (i.e. 40 epochs) and had converged to an effectively stationary state (as judged by changes in the root mean square error) after some 10000 examples. The net was then tested with a *separately generated* (i.e. unseen) set of 400 test data (representing permittivity values at the same frequencies used in the training set). Figure 3 shows a typical example of the results, and compares (i) the true data (used as the inputs

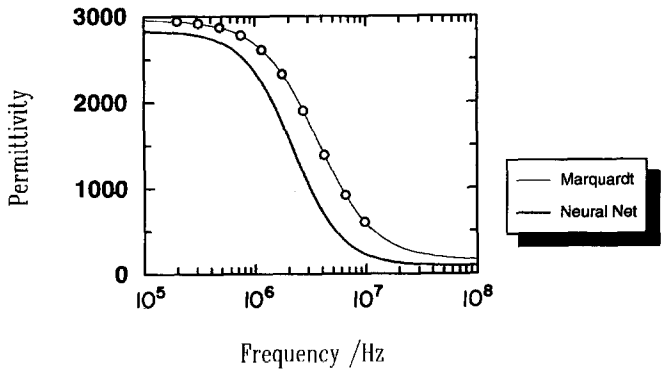

Fig. 3. Estimation of dielectric parameters by means of a neural network. A neural network with a 10–15–4 topology was trained as described in the text. The plot shows a typical example of the results, and compares the true data (○), the best fit to them using the Marquardt algorithm (———) and the neural network's estimate of the dielectric behaviour based on its estimate of the parameters of the Cole–Cole equation (▬▬). The actual values of $\Delta\epsilon$, f_c, ϵ_h and α were 2824, 3.62 MHz, 148 and 0.066, whilst the network's outputs were 2740, 2.17 MHz, 93 and 0.005 respectively.

to the network), (ii) the best fit to them using the Marquardt algorithm with simple weighting and (iii) the neural network's estimate of the dielectric behaviour plotted according to eqn. (4) and using its estimate of the parameters of the Cole–Cole equation. It is evident that the network had indeed learnt features of the training set and had become quite good at guessing the dielectric increment, but less so at determining the other dielectric parameters. This was even more evident when the net's guesses for all 400 (unseen) test sets of data are compared with the true values (data not shown). Although there was a clear trend for three of the four parameters, the net did not appear to detect any significant influence of ϵ_h and always assumed it to have a value somewhere near the mid-point of the range of training values to which it had been exposed. Further studies of this network (not shown) indicated that a sine transfer function was slightly preferable to tanh or sinh functions, that the number of nodes in the hidden layer had little effect on the overall accuracy when varied in the range 5–30 (or when two hidden layers were used instead of one), and that increasing the size of the training set to 10000 and the number of training examples presented to the net to 100000 also did little to change the accuracy of the network.

We felt that the chief cause of the net's inability to "notice" ϵ_h was the rather excessive (and unrealistic) range of values for f_c and α that we had expected it to learn (using what is presently a rather modest number of examples and epochs). We therefore created another training set of 1000 ten-frequency inputs suitable for dealing with the problem as displayed in Fig. 2 and in which the parameters were $0 < \Delta\epsilon < 2000$, $3 \times 10^5 < f_c < 7 \times 10^5$ Hz, $0 < \epsilon_h < 200$ and $0 < \alpha < 0.2$. Various networks were trained with 20000 examples, and the results from a 10–5–4 network are displayed in Fig. 4. Again we stress that the test data were different from the training data. Except at the very extremes of the training set, the values of $\Delta\epsilon$ and ϵ_h returned by the network are within 10% of the true values (and in most cases significantly better). In addition, the net has clearly learned to make use of the information embodied in f_c and α in providing estimates of $\Delta\epsilon$ and ϵ_h that are so much better than would have been surmised from Fig. 2, even though the net cannot determine f_c and α perfectly.

Two arguments can be advanced to show that the net has not simply learned the training set, but has actually learned to generalize from it: (i) the number of members in the training set greatly exceeded the content-addressable memory of our rather small networks (a fully connected 10–5–4 network has only 110 neurons), and (ii) the test data had never been seen by the networks during their training phase.

Whilst there are many other suitable architectures which might be tried in an attempt to improve both the speed of learning and the ability of our ANNs to fit dielectric data to the Cole–Cole equation accurately, the excellent results that we have obtained add weight to the view that this general approach provides a powerful means of solving inverse problems. Indeed, eqn. (4), containing four unknown parameters which interact in a complex non-linear fashion, provides a demanding test for ANNs which might prove useful to others in providing

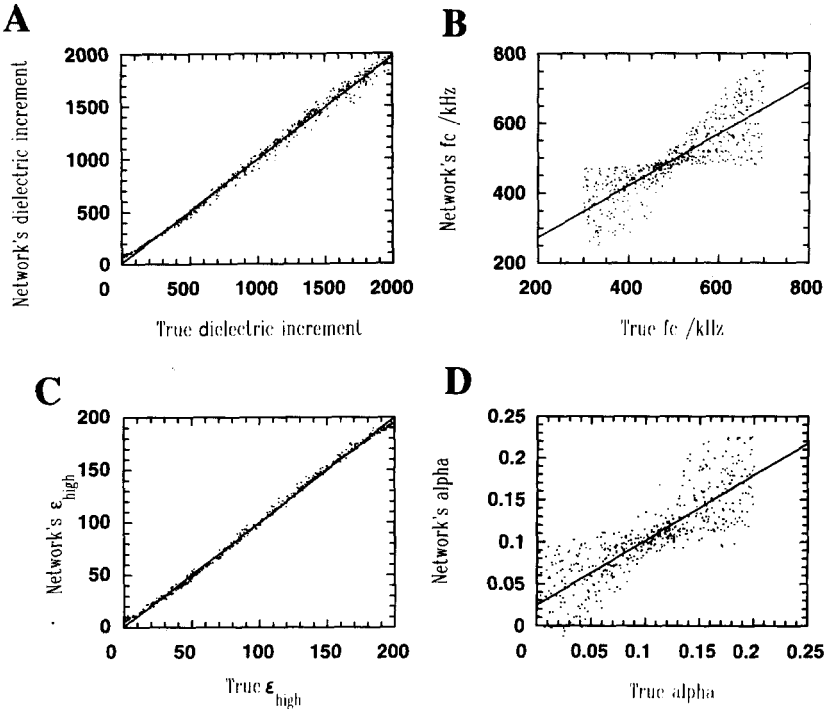

Fig. 4. Estimation of dielectric parameters by means of a neural network. A neural network with a 10–5–4 topology was trained as described in the text and tested with 400 unseen sets of permittivity data. The output of the network was compared with the true values for the parameters which had been used to train the network (according to eqn. (4)). Values for the parameters used in the training and test sets were $0 < \Delta\epsilon < 2000$, $3\times10^5 < f_c < 7\times10^5$, $0 < \epsilon_h < 200$ and $0 < \alpha < 0.2$. (A) Dielectric increment; (B) characteristic frequency; (C) high-frequency permittivity; (D) Cole–Cole α. In each case, the lines are the best linear least-squares fits to the data.

benchmarks for comparison of different neural architectures and learning algorithms. It seems evident that similar approaches might fruitfully be applied to other cognate "inverse" problems, such as that of obtaining the parameters of metabolic control theory [48–51] from experimentally measured variables such as fluxes and metabolite concentrations.

ACKNOWLEDGEMENTS

This work was supported by the Science and Engineering Research Council, Aber Instruments and FT Applikon as part of a LINK Award. DBK thanks Mike Forrester, Alan Hall and Mark Neal for useful and stimulating discussions.

REFERENCES

1 E.H. Grant, R.J. Sheppard and G.P. South, Dielectric Properties of Biological Molecules in Solution, Oxford University Press, Oxford, 1978.
2 O. Schanne and E.R.P. Ceretti, Impedance Measurements in Biological Cells. Wiley, Chichester, 1978.
3 R. Pethig, Dielectric and Electronic Properties of Biological Materials, Wiley, Chichester, 1979.
4 D.B. Kell and C.M. Harris, J. Bioelectr., 4 (1985) 317.
5 K.R. Foster and H.P. Schwan, in C. Polk and E. Postow (Ed), CRC Handbook of Biological Effects of Electromagnetic Fields, CRC Press, Boca Raton, FL, 1986, p. 27.
6 D.B. Kell, in A.P.F. Turner, I. Karube and G.S. Wilson (Ed.), Biosensors: Fundamentals and Applications, Oxford University Press, Oxford, 1987, p. 428.
7 R. Pethig and D.B. Kell, Phys. Med. Biol., 32 (1987) 933.
8 K.R. Foster and H.P. Schwan, CRC Crit Rev. Biomed. Eng., 17 (1989) 25.
9 C.L. Davey and D.B. Kell, in R. Paris (Ed.), Electric Field Phenomena in Biological Systems, IOP Short Meetings Series 21, Institute of Physics, London, 1989, p. 51.
10 S. Takashima, Electrical Properties of Biopolymers and Membranes, Adam Hilger, Bristol, 1989.
11 C.L. Davey and D.B. Kell, in M. O'Connor, R.H.C. Bentall and J.S. Monahan (Eds.), Emerging Electromagnetic Medicine, Springer, Berlin, 1990, p. 19.
12 D.B. Kell and C.L. Davey, in A.E.G. Cass (Ed.), Biosensors; A Practical Approach, Oxford University Press, Oxford, 1990, p. 125.
13 K.S. Cole and R.H. Cole, J. Chem. Phys., 9 (1941) 341.
14 P. Debye, Polar Molecules, Dover, New York, 1929.
15 H.P. Schwan, Adv. Biol. Med. Phys., 5 (1957) 147.
16 C.M. Harris, R.W. Todd, S.J. Bungard, R.W. Lovitt, J.G. Morris and D.B. Kell, Enzymol. Microbiol. Technol., 9 (1987) 181.
17 D.B. Kell, G.H. Markx, C.L. Davey and R.W. Todd, Trends Anal. Chem., 9 (1990) 190.
18 G.H. Markx and D.B. Kell, Biofouling, 2 (1990) 211.
19 G.H. Markx, C.L. Davey and D.B. Kell, J. Gen. Microbiol., 137 (1991) 735.
20 C.L. Davey, W. Peñaloza, D.B. Kell and J.N. Hedger, World J. Microbiol. Biotechnol., 7 (1991) 248.
21 R. Firstenberg-Eden and G. Eden, Impedance Microbiology, Research Studies Press, Letchworth, 1984.
22 J.R. Macdonald, J. Schoonman and A.P. Lehnen, J. Electroanal. Chem., 131 (1982) 77.
23 K. Levenberg, Q. Appl. Math., 2 (1944) 164.
24 D.W. Marquardt, J. Soc. Ind. Appl. Math., 11 (1963) 431.
25 D.E. Rumelhart, J.L. McClelland and the PDP Research Group, Parallel Distributed Processing. Experiments in the Microstructure of Cognition, MIT Press, Cambridge, MA, 1986.
26 J.L. McClelland and D.E. Rumelhart, Explorations in Parallel Distributed Processing, MIT Press, Cambridge, MA, 1989.
27 T. Kohonen, Self-Organization and Associative Memory (3rd edn.), Springer, Heidelberg, 1989.
28 Y.-H. Pao, Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, New York, 1989.
29 D.J. Amit, Modeling Brain Function. The World of Attractor Neural Networks, Cambridge University Press, Cambridge, 1989.
30 S. Brunak and B. Lautrup, Neural Networks, World Scientific, Singapore, 1990.
31 B. Müller and J. Reinhardt, Neural Networks, Springer, Heidelberg, 1990.
32 P.K. Simpson, Artificial Neural Systems. Pergamon, Oxford, 1990.
33 I. Aleksander and H. Morton, An Introduction to Neural Computing. Chapman and Hall, London, 1990.
34 J. Hertz, A. Krogh and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City, CA, 1991.
35 T. Kohonen, M. Mäkisara, O. Simula and J. Kangas (Eds.), Artificial Neural Networks, Vols. 1 and 2, Elsevier, Amsterdam, 1991.

36 K. Hornik, M. Stinchcombe and H. White, Neural Networks, 2 (1989) 359.
37 W.W. Armstrong and J. Gecsei, IEEE Trans. Syst. Man Cybern., SMC-9 (1979) 276.
38 D.H. Ackley, G.E. Hinton and T.J. Sejnowski, Cognitive Sci., 9 (1985) 147.
39 G.A. Carpenter and S. Grossberg, Comput. Vis. Graphics Image Proc., 37 (1987) 54.
40 G.A. Carpenter and S. Grossberg, Appl. Optics, 26 (1989) 4919.
41 K. Fukushima, Neural Networks, 1 (1988) 119.
42 S.E. Fahlman and C. Lebiere, The Cascade-Correlation Learning Architecture, Technical Report CMU-CS-90-100, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA.
43 G. Bolt, Investigating Fault Tolerance in Artificial Neural Networks, York Computer Science Report YCS 154, 1991.
44 M.J. Dzwonczyk,, Quantitative Failure Models of Feed-Forward Neural Networks, MSc Thesis, Draper Laboratory Publication CSDL-T-1068, Massachusetts Institute of Technology, 1991.
45 U. Thomsen and B. Meyer, J. Magn. Reson., 84 (1989) 212.
46 M. Bos and H.T. Weber, Anal. Chim. Acta, 247 (1991) 97.
47 J.R. Long, H.T. Mayfield, M.V. Henley and P.R. Kromann, Anal. Chem., 63 (1991) 1256.
48 H. Kacser and J.A. Burns, Symp. Soc. Exp. Biol., 27 (1973) 65.
49 R. Heinrich and T.A. Rapoport, Eur. J. Biochem., 42 (1974) 89.
50 D.B. Kell, K. van Dam and H.V. Westerhoff, Symp. Soc. Gen. Microbiol., 44 (1989) 63.
51 A. Cornish-Bowden and M. Cárdenas (Eds.), Control of Metabolic Processes, Plenum, New York, 1990.