

LA-UR- 98-1702

Title:

IMPLEMENTING SIZE-OPTIMAL DISCRETE NEURAL
NETWORKS REQUIRES ANALOG CIRCUITRY

CONF-980911--

Author(s):

Valeriu Beiu

RECEIVED
OCT 05 1998
OSTI

Submitted to:

IV European Signal Processing Conference
Rhodes, Greece
September 8-11, 1998

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

[Handwritten signature]



Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 R5
ST 2629 10/91

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

IMPLEMENTING SIZE-OPTIMAL DISCRETE NEURAL NETWORKS REQUIRE ANALOG CIRCUITRY

Valeriu Beiu *

Los Alamos National Laboratory, Division Space and Atmospheric Sciences NIS-1, MS D466

Los Alamos, New Mexico 87545, USA

Tel.: +1-505-667 2430; fax: +1-505-665 7395

e-mail: beiu@lanl.gov

ABSTRACT

This paper starts by overviewing results dealing with the approximation capabilities of neural networks, as well as bounds on the size of threshold gate circuits. Based on a constructive solution for Kolmogorov's superpositions we will show that implementing Boolean functions can be done using neurons having an identity transfer function. Because in this case the size of the network is minimised, it follows that size-optimal solutions for implementing Boolean functions can be obtained using analog circuitry. Conclusions and several comments on the required precision are ending the paper.

Keywords: neural networks, Kolmogorov's superpositions, threshold gate circuits, analog circuits, size, precision.

1 INTRODUCTION

In this paper a *network* is an acyclic graph having several input nodes, and some (at least one) output nodes. If a synaptic *weight* is associated with each edge, and each node computes the weighted sum of its inputs to which a nonlinear activation function is then applied: $f(\mathbf{x}) = f(x_1, \dots, x_n) = \sigma(\sum_{i=1}^n w_i x_i + \theta)$, the network is a *neural network* (NN), with $w_i \in \mathbb{R}$ the synaptic *weights*, $\theta \in \mathbb{R}$ known as the *threshold*, Δ being the *fan-in*, and σ a non-linear activation function. Because the underlying graph is acyclic, the network does not have feedback, and can be layered. That is why such a network is also known as a *multilayer feedforward neural network*, and is commonly characterised by: its *depth* (i.e., number of layers), and its *size* (i.e., number of neurons).

The paper starts by overviewing several of the results dealing with the approximation capabilities of NNs. They are followed by details on known upper and lower bounds on the size of *threshold gate circuits* (TGCs). They show that TGCs require exponential size if arbitrary Boolean functions (BFs) are to be implemented. Based on a constructive solution for Kolmogorov's superpositions we will prove that in order to obtain linear size NNs (i.e., size-optimal) for implementing any BF, the nonlinear activation function of the neurons has to be the identity function. Because both Boolean and TGCs require exponential size, it follows that size-optimal implementations of discrete NNs (i.e., implementing BFs) can be obtained only in analog circuitry. Conclusions, and several comments on the required precision are ending the paper.

2 PREVIOUS RESULTS

NNs have been experimentally shown to be quite effective in many applications (see *Applications of Neural Networks* in [2], together with *Part F: Applications of Neural Computation* and *Part G: Neural Networks in Practice: Case Studies* from [20]). This success has led researchers to undertake a rigorous analysis of their mathematical properties and has generated two directions of research for finding: (i) existence/constructive proofs for the '*universal approximation problem*'; (ii) tight bounds on the size needed by the approximation problem. The paper will focus on both aspects, for the case when the functions to be implemented are BFs.

2.1 Neural Networks as Universal Approximators

The first line of research has concentrated on the approximation capabilities of NNs [14, 22, 34, 35]. It was started in 1987 by Hecht-Nielsen [25] and Lippmann [46] who, together with LeCun [44], were probably the first to recognise that the specific format in [65, 66] of the form:

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \left\{ \Phi_q \left[\sum_{p=1}^n \alpha_p \psi(x_p + qa) \right] \right\}$$

of Kolmogorov's superpositions $f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q(y_q)$ [40] can be interpreted as a NN with one hidden layer. This gave an existence proof of the approximation properties of NNs. The first nonconstructive proof was given in 1988 by Cybenko [16, 17] using a continuous activation function, and was independently presented by Irie and Miyake [33]. Thus, the fact that NNs are computationally universal—with more or less restrictive conditions—when modifiable connections are allowed, was established. Different enhancements have been later presented (for details see [62], and *Chp. 1* in [9]):

- Funahashi [21] proved the same result but in a more constructive way and also refined the use of Kolmogorov's theorem in [25], giving an approximation result for two-hidden-layer NNs;
- Hornik *et al.* [30] showed that the continuity requirement for the output function can partly be removed;
- Hornik *et al.* [31] also proved that a NN can approximate simultaneously a function and its derivative;
- Park and Sandberg [56, 57] used radial basis functions in the hidden layer, and gave an almost constructive proof;
- Hornik [28] showed that the continuity requirement can be completely removed, the activation function having to be 'bounded and nonconstant';

* On leave of absence from the "Politehnica" University of Bucharest, Computer Science Department, Spl. Independentei 313, RO-77206 Bucharest, Romania.

- Geva and Sitte [23] proved that four-layered NNs with sigmoid activation function are universal approximators;
- Kürková [42, 43] has demonstrated the existence of approximate superposition representations within the constraints of NNs, *i.e.* ψ and Φ_q can be approximated with functions of the form $\sum a_i \sigma(b_i x + c_i)$, where σ is an arbitrary activation sigmoidal function;
- Mhaskar and Micchelli [48, 49] approach was based on the Fourier series of the function, by truncating the infinite sum to a finite set, and rewriting e^{ikx} in terms of the activation function (which now has to be periodic);
- Koiran [39] presented a new proof on the line of Funahashi's proof [21], but more general in that it allows the use of units with 'piecewise continuous' activation functions; these include the particular but important case of *threshold gates* (TGs);
- Leshno *et al.* [45] relaxed the condition for the activation function to 'locally bounded piecewise continuous' (*i.e.*, if and only if the activation function is not a polynomial), thus embedding as special cases almost all the activation functions that have been reported in the literature;
- Hornik [29] added to these results by proving that: (*i*) if the activation function is locally Riemann integrable and nonpolynomial, the *weights* and the *thresholds* can be constrained to arbitrarily small sets; and (*ii*) if the activation function is locally analytic, a single universal *threshold* will do;
- Funahashi and Nakamura [22] showed that the universal approximation theorem also holds for trajectories of patterns;
- Sprecher [67] has demonstrated that there are universal hidden layers that are independent of the number of input variables n ;
- Barron [4] described spaces of functions that can be approximated by the relaxed algorithm of Jones [36] using functions computed by single-hidden-layer networks of perceptrons;
- Attali and Pagès [3] provide an elementary proof based on the Taylor expansion and the Vandermonde determinant, yielding bounds for the design of the hidden layer and convergence results for the derivatives.

All these results—with the partial exception of [4, 39, 42, 56, 57]—were obtained “provided that sufficiently many hidden units are available” (*i.e.*, with no claims on the *size* minimality). More constructive solutions have been obtained in very small *depth* later [37, 53, 54], but their *size* still grows fast with respect to the number of dimensions and/or examples, or with the required precision.

Recently, an explicit numerical algorithm for superpositions has been detailed [68–70].

2.2 Threshold Gate Circuits

The other line of research was to find the smallest *size* NN which can realise an arbitrary function given a set of m vectors from \mathbb{R}^n . Many results have been obtained for TGs [50]. The first lower bound of $\text{size} \geq 2(2^n/n)^{1/2}$ on the *size* of a TGC for “almost all” n -ary BFs ($f: \mathbb{B}^n \rightarrow \mathbb{B}$) was given in

[52]. Later [47] a very tight upper bound was proven in $\text{depth} = 4$: $\text{size} \leq 2(2^n/n)^{1/2} \times \{1 + \Omega[(2^n/n)^{1/2}]\}$. A similar existence exponential lower bound of $\Omega(2^{n/3})$ for arbitrary BFs can be found in [63], which also gives bounds for many particular but important BFs (see also [61]).

For classification problems ($f: \mathbb{R}^n \rightarrow \mathbb{B}^k$), the first result was that a NN of *depth* = 3 and *size* = $m - 1$ could compute an arbitrary dichotomy. The main improvements have been:

- Baum [5] presented a TGC with one hidden layer having $\lceil m/n \rceil$ neurons capable of realising an arbitrary dichotomy on a set of m points *in general position* in \mathbb{R}^n ; if the points are on the corners of the n -dimensional hypercube ($f: \mathbb{B}^n \rightarrow \mathbb{B}$), $m - 1$ nodes are still needed;
- a slightly tighter bound of only $\lceil 1 + (m-2)/n \rceil$ neurons in the hidden layer for realising an arbitrary dichotomy on a set of m points which satisfy a more relaxed topological assumption was proven in [32]; the $m - 1$ nodes condition was shown to be the least upper bound needed;
- Arai [1] showed that $m - 1$ hidden neurons are necessary for arbitrary separability, but improved the bound for the dichotomy problem to $m/3$ (without any condition on the inputs);
- Beiu and De Pauw [10] have detailed existence lower and upper bounds by estimating the entropy of the dataset $2m/(n \log n) < \text{size} < 1.44m/n$ (see also [11, 13]).

Other existence lower bounds for the arbitrary dichotomy problem can be found in [24, 58]:

- a *depth*-2 TGC requires at least $m/\{n \log(m/n)\}$ TGs;
- a *depth*-3 TGC requires at least $2(m/\log m)^{1/2}$ TGs in each of the two hidden layer (if $m \gg n^2$);
- an arbitrarily interconnected TGC without feedback needs $(2m/\log m)^{1/2}$ TGs (if $m \gg n^2$).

One study [15] has tried to unify these two lines of research by first presenting analytical solutions for the general NN problem in one dimension (having infinite *size*), and then giving practical solutions for the one-dimensional cases (*i.e.*, including an upper bound on the *size*). Extensions to the n -dimensional case using three- and four-layers solutions were derived under piecewise constant approximations (having constant or variable width partitions), and under piecewise linear approximations (using ramps instead of sigmoids).

2.3 Boolean Functions

The particular case of BFs has been studied intensively [9, 55]. Many results have been obtained for particular BFs [61, 63], but a *size*-optimal result for BFs that have exactly m groups of ones in their truth table $\mathbb{F}_{n,m}$ (BFs which are defined on the m groups) was detailed by Red'kin in 1970 [60].

Theorem from [60] *The complexity realisation (*i.e.*, number of threshold elements) of $\mathbb{F}_{n,m}$ is at most $2(2m)^{1/2} + 3$.*

This result—as are all the previous mentioned ones—is valid for unlimited *fan-in* TGs. Departing from these lines, Horne and Hush [27] have detailed a solution for limited *fan-in* TGCs.

Theorem from [27] *Arbitrary Boolean functions of the form $f: \{0, 1\}^n \rightarrow \{0, 1\}^m$ can be implemented in a NN of perceptrons restricted to *fan-in* 2 with a node complexity of $\Theta\{m 2^n/(n + \log m)\}$ and requiring $O(n)$ layers.*

3 SIZE-OPTIMAL IMPLEMENTATIONS

It is known that implementing arbitrary BFs using classical Boolean gates (*i.e.*, AND and OR gates) requires exponential *size* circuits. As has been seen from all the previous results, the known bounds for *size* are also exponential if TGCs are used for solving arbitrary BFs [6]. It is true that these bounds reveal exponential gaps (thus encouraging research efforts to reduce them), and also suggest that TGCs with more layers (*depth* ≠ small constant [8, 12]) might have a smaller *size*.

A different approach is to use Kolmogorov's superpositions, which shows that there are NNs having only $2n + 1$ neurons which can approximate any function. Such a solution would clearly be *size*-optimal. We start from [68-70], where a constructive solution for the general case was detailed.

Theorem from [68] Define the function $\psi: \mathcal{E} \rightarrow \mathcal{D}$ such that for each integer $k \in N$:

$$\psi\left(\sum_{r=1}^k i_r \gamma^{-r}\right) = \sum_{r=1}^k \tilde{i}_r 2^{-m_r} \gamma^{-\frac{n-m_r-1}{n-1}} \quad (1)$$

where $\tilde{i}_r = i_r - (\gamma - 2) \langle i_r \rangle$ and

$$m_r = \langle i_r \rangle (1 + \sum_{s=1}^{r-1} [i_s] \times \dots \times [i_{r-1}])$$

for $r = 1, 2, \dots, k$.

Here $\gamma \geq 2n + 2$ is a base, $\mathcal{E} = [0, 1]$ is the unit interval, \mathcal{D} is the set of terminating rational numbers $d_k = \sum_{r=1}^k i_r \gamma^{-r}$ defined on $k \in N$ digits ($0 \leq i_r \leq \gamma - 1$). Also, $\langle i_1 \rangle = [i_1] = 0$, while for $r \geq 2$: $\langle i_r \rangle = 0$ when $i_r = 0, 1, \dots, \gamma - 2$, $\langle i_r \rangle = 1$ when $i_r = \gamma - 1$, $[i_r] = 0$ when $i_r = 0, 1, \dots, \gamma - 3$, while $[i_r] = 1$ when $i_r = \gamma - 2, \gamma - 1$.

If we limit the functions to be 'approximated' to BFs, one digit is enough ($k = 1$), which gives $\psi(0.i_1) = 0.i_1$, which is the identity function $\psi(x) = x$. Such a solution builds quite simple analog neurons. They have *fan-in* $\Delta \leq 2n + 1$, for which the known *weight* bounds (holding for any *fan-in* $\Delta \geq 4$) are [51, 55, 59, 64]:

$$2^{(\Delta-1)/2} < \text{weight} < (\Delta+1)^{(\Delta+1)/2} / 2^\Delta.$$

Thus, a *precision* of between Δ , and $\Delta \log \Delta$ bits per *weight* would be expected. Unfortunately, the constructive solution for Kolmogorov's superpositions requires a double exponential precision for $\psi(1)$, and for the *weights*:

$$\alpha_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\frac{n^r-1}{n-1}}.$$

For BFs precision is reduced to $(2n+2)^{-n}$, or $2n \log n$ bits per *weight*. Analog implementations are limited to just several bits of precision [41], this being one of the reasons for investigations on precision [18, 26, 71, 72] and on algorithms relying on limited integer *weights* [7, 19, 38].

Due to the limitation on precision an optimal solution for implementing BFs should decompose the given BF in simpler BFs which can be efficiently implemented based on Kolmogorov's superpositions (*i.e.*, we have to reduce n to small values). The partial results from this first layer of analog building blocks can be combined using again Kolmogorov's superpositions. The final analog implementation will requires more than three layers.

4 CONCLUSIONS

Arbitrary BFs can be implemented using:

- classical Boolean gates, but require exponential *size*;
- TGS, but (again) in exponential *size* (still, there are exponential gaps between classical Boolean solutions and TG ones);
- analog building blocks in linear *size* (having linear *fan-in* and polynomial precision *weights* and *thresholds*); the nonlinear activation function is the identity function.

The main conclusion is that *size*-optimal hardware implementations of BFs can be obtained only in analog circuitry. The high precision required by the solution based on Kolmogorov's superpositions can be tackled by decomposing a BFs into simpler BFs.

References

- [1] M. Arai. Bounds on the Number of Hidden Units in Binary-valued Three-layer Neural Networks. *Neural Networks*, 6(6), 855-860, 1993.
- [2] M.A. Arbib (ed.). *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, 1995.
- [3] J.-G. Attali & G. Pages. Approximations of Functions by a Multilayer Perceptron: a New Approach. *Neural Networks*, 10(6), 1069-1081, 1997.
- [4] A.R. Barron. Universal Approximation Bounds for Superpositions of a Sigmoidal Function. *IEEE Trans. Info. Theory*, 39(3), 930-945, 1993.
- [5] E.B. Baum. On the Capabilities of Multilayer Perceptrons. *J. Complexity*, 4, 193-215, 1988.
- [6] V. Beiu. Digital Integrated Circuit Implementations. *Chapter E1.4* in [20], 1996.
- [7] V. Beiu. Reduced Complexity Constructive Learning Algorithm. *Tech. Rep. LA-UR-97-515*, Los Alamos Natl. Lab., USA; *Neural Network World*, 8(1), 1-38, 1998.
- [8] V. Beiu. On the Circuit and VLSI Complexity of Threshold Gate *COMPARISON Tech. Rep. LA-UR-96-3591*, Los Alamos Natl. Lab., USA; *Neurocomputing*, 19(1-3), 77-98, 1998.
- [9] V. Beiu. *VLSI Complexity of Discrete Neural Networks*. Gordon & Breach, Newark, NJ, 1998.
- [10] V. Beiu & T. De Pauw. Tight Bounds on the Size of Neural Networks for Classification Problems. *Tech. Rep. LA-UR-97-60*, Los Alamos Natl. Lab., USA; in J. Mira, R. Moreno-Díaz & J. Cabestany (eds.): *Biological and Artificial Computation*, Springer, Berlin, 743-752, 1997.
- [11] V. Beiu & S. Draghici. Limited Weights Neural Networks: Very Tight Entropy Based Bounds. *Tech. Rep. LA-UR-97-294*, Los Alamos Natl. Lab., USA; in D.W. Pearson (ed.): *Proc. ICSC Symp. on Soft Computing*, ICSC Acad. Press, Canada, 111-118, 1997.
- [12] V. Beiu & H.E. Makaruk. Deeper Sparser Nets Can Be Optimal. *Tech. Rep. LA-UR-97-4413*, Los Alamos Natl. Lab., USA, 1997; to appear in *Neural Proc. Lett.*, 7/8, 1998.
- [13] V. Beiu, S. Draghici & T. De Pauw. A Constructive Approach to Calculating Lower Entropy Bounds. *Tech. Rep. LA-UR-97-884*, Los Alamos Natl. Lab., USA; to appear in *Neural Proc. Lett.*, 7/8, 1998.
- [14] E. Blum & K. Li. Approximation Theory and Feedforward Networks. *Neural Networks*, 4(3), 511-515, 1991.
- [15] A. Bulsari. Some Analytical Solutions to the General Approximation Problem for Feedforward Neural Networks. *Neural Networks*, 6(7), 991-996, 1993.
- [16] G. Cybenko. Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient. *Tech. Rep.*, Tufts Univ., Medford, MA, 1988.
- [17] G. Cybenko. Approximations by Superpositions of a Sigmoid Function. *Math. of Control, Signals and Systems*, 2, 303-314, 1989.
- [18] J.S. Denker & B.S. Wittner. Network Generality, Training Required, and Precision Required. In D.Z. Anderson (ed.): *Neural Information Processing Systems (NIPS*87)*, AIP, NY, 219-222, 1988.

[19] S. Draghici & I.K. Sethi. On the Possibilities of the Limited Precision Weights Neural Networks in Classification Problems. In J. Mira, R. Moreno-Díaz & J. Cabestany (eds.): *Biological and Artificial Computation*, Springer, Berlin, 753-762, 1997.

[20] E. Fiesler & R. Beale (eds.). *Handbook of Neural Computation*. Oxford Univ. Press & the Inst. of Physics, NY, 1996.

[21] K.-I. Funahashi. On the Approximate Realization of Continuous Mapping by Neural Networks. *Neural Networks*, **2**(2), 183-192, 1989.

[22] K.-I. Funahashi and Y. Nakamura. Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks. *Neural Networks*, **6**(6), 801-806, 1993.

[23] S. Geva & J. Sitte. A Constructive Method for Multivariate Function Approximation by Multilayered Perceptrons. *IEEE Trans. Neural Networks*, **3**(4), 621-623, 1992.

[24] M.H. Hassoun (ed.). *Fundamentals of Artificial Neural Networks*. MIT Press, Cambridge, MA, 1995.

[25] R. Hecht-Nielsen. Kolmogorov's Mapping Neural Network Existence Theorem. *Proc. Intl. Conf. on Neural Networks*, IEEE CS Press, CA, 11-14, 1987.

[26] J.L. Holt & J.-N. Hwang. Finite Precision Error Analysis of Neural Network Hardware Implementations. *IEEE Trans. Comp.*, **42**(3), 281-290, 1993.

[27] B.G. Horne & D.R. Hush. On the Node Complexity of Neural Networks. *Neural Networks*, **7**(9), 1413-1426, 1994.

[28] K. Hornik. Approximation Capabilities of Multilayer Feedforward Networks. *Neural Network*, **4**(2), 251-257, 1991.

[29] K. Hornik. Some New Results on Neural Network Approximation. *Neural Network*, **6**(8), 1069-1072, 1993.

[30] K. Hornik, M. Stinchcombe & H. White. Multilayer Feedforward Neural Networks Are Universal Approximators. *Neural Network*, **2**(3), 359-366, 1989.

[31] K. Hornik, M. Stinchcombe & H. White. Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks. *Neural Network*, **3**(4), 551-560, 1990.

[32] S.-C. Huang & Y.-F. Huang. Bounds on the Number of Hidden Neurons of Multilayer Perceptrons in Classification and Recognition. *IEEE Trans. Neural Networks*, **2**(1), 47-55, 1991.

[33] B. Irie & S. Miyake. Capabilities of Three-Layered Perceptrons. *Proc. Intl. Conf. on Neural Networks*, IEEE CS Press, CA, vol. 1, 641-648, 1988.

[34] Y. Ito. Approximation of Functions on a Compact Set by Finite Sums of Sigmoid Functions without Scaling. *Neural Networks*, **4**(7), 817-826, 1991.

[35] Y. Ito. Approximation Capabilities of Layered Neural Networks with Sigmoid Units on Two Layers. *Neural Computation*, **6**(6), 1233-1243, 1994.

[36] L.K. Jones. A Simple Lemma on Greedy Approximation in Hilbert Space and Convergence Rates for Projection Pursuit Regression and Neural Network Training. *Ann. Statist.*, **20**(1), 608-613, 1992.

[37] H. Katsuura & D.A. Sprecher. Computational Aspects of Kolmogorov's Superposition Theorem. *Neural Networks*, **7**(3), 455-461, 1994.

[38] A.H. Khan & E.L. Hines. Integer-Weight Neural Networks. *Electr. Lett.*, **30**(15), 1237-1238, 1994.

[39] P. Koiran. On the Complexity of Approximating Mappings Using Feedforward Networks. *Neural Networks*, **6**(5), 649-653, 1993.

[40] A.N. Kolmogorov. On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition. *Dokl. Akad. Nauk SSSR*, **114**, 953-956, 1957; English transl.: *Transl. Amer. Math. Soc.*, **2**(28), 55-59, 1963.

[41] A.H. Kramer. Array-Based Analog Computation; Principles, Advantages and Limitations. *Proc. Microelectronics for Neural Networks*, IEEE CS Press, CA, 68-79, 1996.

[42] V. Kůrková. Kolmogorov's Theorem and Multilayer Neural Networks. *Neural Networks*, **5**(4), 501-506, 1992.

[43] V. Kůrková, P.C. Kainen & V. Kreinovich. Estimates of the Number of Hidden Units and Variations with Respect to Half-Spaces. *Neural Networks*, **10**(6), 1061-1068, 1997.

[44] Y. LeCun. Models connexionnistes de l'apprentisage. *MSc thesis*, Université Pierre et Marie Curie, Paris, 1987.

[45] M. Leshno, V.Y. Lin, A. Pinkus & S. Schocken. Multilayer Feedforward Neural Networks with a Nonpolynomial Activation Function Can Approximate any Function. *Neural Networks*, **6**(6), 861-867, 1993.

[46] R.P. Lippmann. An Introduction to Computing with Neural Nets. *IEEE ASSP Mag.*, **4**(2), 4-22, 1987.

[47] O.B. Lupanov. The Synthesis of Circuits from Threshold Elements. *Problemy Kibernetiki*, **20**, 109-140, 1973.

[48] H.N. Mhaskar & C. Micchelli. Approximation by Superposition of Sigmoidal and Radial Basis Functions. *Adv. in Appl. Maths.*, **13**, 350-373, 1992.

[49] H.N. Mhaskar & C. Micchelli. Dimension Independent Bounds on the degree of Approximation by Neural Networks. *IBM J. Res. and Dev.*, **38**(3), 277-283, 1994.

[50] R.C. Minnik. Linear-Input Logic. *IRE Trans. Electr. Comp.*, **10**, 6-16, 1961.

[51] J. Myhill & W.H. Kautz. On the Size of Weights Required for Linear-Input Switching Functions. *IRE Trans. Electr. Comp.*, **10**, 288-290, 1961.

[52] E.I. Nekopork. The Synthesis of Networks from Threshold Elements. *Soviet Mathematics—Doklady*, **5**(1), 163-166, 1964. English transl.: *Automation Express*, **7**(1), 35-39 & **7**(2), 27-32, 1964.

[53] M. Nees. Approximate Versions of Kolmogorov's Superposition Theorem, Proved Constructively. *J. Comp. and Appl. Math.*, **54**(2), 239-250, 1994.

[54] M. Nees. Chebyshev Approximation by Discrete Superposition. Application to Neural Networks. *Adv. in Comp. Maths.*, **5**(2), 137-152, 1996.

[55] I. Parberry. *Circuit Complexity and Neural Networks*. MIT Press, Cambridge, MA, 1994.

[56] J. Park & I.W. Sandberg. Universal Approximation Using Radial-Basis-Function Networks. *Neural Computation*, **3**(2), 246-257, 1991.

[57] J. Park & I.W. Sandberg. Approximation and Radial-Basis-Function Networks. *Neural Computation*, **5**(3), 305-316, 1993.

[58] H. Paugam-Moisy. *Optimisation des réseaux des neurones artificiels*. PhD dissertation, LIP-IMAG, Lyon, France, 1992.

[59] P. Raghavan. Learning in Threshold Networks: A Computational Model and Applications. *Tech. Rep. RC 13859*, IBM Res., 1988. Also in *Proc. on Comput. Learning Th.*, ACM Press, NY, 19-27, 1988.

[60] N.P. Red'kin. Synthesis of Threshold Circuits for Certain Classes of Boolean Functions. *Kibernetika*, **5**(1), 6-9, 1970. English transl.: *Cybernetics*, **6**(5), 540-544, 1973.

[61] V.P. Roychowdhury, A. Orlitsky & K.-Y. Siu. Lower Bounds on Threshold and Related Circuits Via Communication Complexity. *IEEE Trans. Info. Theory*, **40**(2), 467-474, 1994.

[62] F. Scarselli & A.C. Tsoi. Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results. *Neural Networks*, **11**(1), 15-37, 1998.

[63] K.-Y. Siu, V. Roychowdhury & T. Kailath. Depth-Size Tradeoffs for Neural Computations. *IEEE Trans. Comp.*, **40**(12), 1402-1412, 1991.

[64] E.D. Sontag. Shattering All Sets of k Points in "General Position" Requires $(k-1)/2$ Parameters. *Report SYCON 96-01*, Maths.Dept., Rutgers Univ., 1996; *Neural Computation*, **9**(2), 337-348, 1997.

[65] D.A. Sprecher. On the Structure of Continuous Functions of Several Variables. *Trans. American Math. Soc.*, **115**, 340-355, 1965.

[66] D.A. Sprecher. On the Structure of Representations of Continuous functions of Several Variables as Finite Sums of Continuous Functions of One Variable. *Proc. American Math. Soc.*, **17**, 98-105, 1966.

[67] D.A. Sprecher. A Universal Mapping for Kolmogorov's Superposition Theorem. *Neural Networks*, **6**(8), 1089-1094, 1993.

[68] D.A. Sprecher. A Numerical Implementation of Kolmogorov's Superpositions. *Neural Networks*, **9**(5), 765-772, 1996.

[69] D.A. Sprecher. A Numerical Construction of a Universal Function for Kolmogorov's Superpositions. *Neural Network World*, **6**(4), 711-718, 1996.

[70] D.A. Sprecher. A Numerical Implementation of Kolmogorov's Superpositions II. *Neural Networks*, **10**(3), 447-457, 1997.

[71] M. Stevenson & S. Huq. On the Capability of Threshold Adalines with Limited-Precision Weights. *Neural Computation*, **8**(8), 1603-1610, 1996.

[72] J. Wray & G.G.R. Green. Neural Networks, Approximation Theory, and Finite Precision Computation. *Neural Networks*, **8**(1), 31-37, 1995.