LA-UR- 98-3463

Approved for public release; distribution is unlimited.

Title: ON AUTOMATIC SYNTHESIS OF ANALOG/DIGITAL CIRCUITS

CONF-9809116--

Author(s): Valeriu C. Beiu, NIS-1

Submitted to: International Symposium on Nonlinear Theory and its Applications
September 14-17, 1998 Crans-Montana, Switzerland

Los Alamos
NATIONAL LABORATORY

DISCLAIMER

On Automatic Synthesis of Analog / Digital Circuits

Valeriu Beiu *

Division NIS–1, Los Alamos National Laboratory, MS D466, Los Alamos, NM 87545, USA
beiu@lanl.gov

Abstract— The paper builds on a recent explicit numerical algorithm for Kolmogorov's superpositions, and will show that in order to synthesise minimum *size* (*i.e.*, *size*-optimal) circuits for implementing any Boolean function, the nonlinear activation function of the gates has to be the identity function. Because classical AND-OR implementations, as well as threshold gate implementations require exponential *size*, it follows that *size*-optimal solutions for implementing arbitrary Boolean functions can be obtained using analog (or mixed analog/digital) circuits. Conclusions and several comments are ending the paper.

I. INTRODUCTION

In this paper a *network* is an acyclic graph having several input nodes, and some 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 applied: $f(x_1,\ldots,x_\Delta) = \sigma(\sum_{i=1}^{\Delta} w_i x_i + \theta)$, 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. The underlying graph is acyclic, thus the network can be layered, *i.e.*, it is a *multilayer feedforward NN*. It is characterised by: *depth* (*i.e.*, number of layers), and *size* (*i.e.*, number of neurons).

The paper starts by overviewing results dealing with the approximation capabilities of NNs, and known upper and lower bounds on the *size* of *threshold gate circuits* (TGCs). They show that TGCs require exponential *size* for implementing arbitrary *Boolean functions* (BFs). 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 TG circuits require expo- nential *size*, it follows that *size*-optimal implementations of BFs can be obtained in analog circuitry. Conclusions, and comments on the required precision are ending the paper.

II. 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 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.

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

A. Neural Networks as Universal Approximators

The first line of research has concentrated on the approximation capabilities of NNs [14,22,35,36]. It was started in 1987 by Hecht-Nielsen [26] and Lippmann [47] who, together with LeCun [45], were probably the first to recognise that the specific format of Kolmogorov's superpositions [41] $f(x_1,\ldots,x_n) = \sum_{q=1}^{2n+1} \Phi_q(y_q)$ in [67,68]:

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

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], and was independently presented by Irie&Miyake [34]. Similar results for radial basis functions were shortly reported [24,59]. Different enhancements have been later presented (see [9,64]):

- Funahashi [21] proved the same result and also refined the use of Kolmogorov's theorem in [26], giving an approximation result for two-hidden-layer NNs;
- Hornik *et al.* [31] showed that the continuity requirement for the output function can partly be removed;
- Hornik *et al.* [32] also proved that a NN can approximate simultaneously a function and its derivative;
- Park&Sandberg [57, 58] used radial basis functions in the hidden layer (almost constructive proof);
- Hornik [29] showed that the continuity requirement can be completely removed, the activation function having to be 'bounded and nonconstant';
- Geva&Sitte [23] proved that four-layered NNs with sigmoid activation are universal approximators;
- Kůrková [43,44] has demonstrated the existence of approximate superposition representations within the constraints of NNs, *i.e.* ψ and Φ_q being of the form $\sum a_r \sigma(b_r x + c_r)$, where σ is an arbitrary activation sigmoidal function;
- Mhaskar&Micchelli [49,50] approach was based on the Fourier series of the function, by truncating the infinite sum, and rewriting e^{ikx} in terms of the activation function (which now has to be periodic);
- Koiran [40] presented a proof on the line of Funahashi's proof [21], which allows the use of units with 'piecewise continuous' activation functions;
- Leshno *et al.* [46] relaxed the condition for the activation function to 'locally bounded piecewise continuous', thus embedding as special cases almost all the activation functions that have been reported;
- Hornik [30] 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&Nakamura [22] showed that the universal approximation theorem also holds for trajectories;

- Sprecher [69] has demonstrated that there are universal hidden layers that are independent of n;
- Barron [4] described spaces of functions that can be approximated by the relaxed algorithm of Jones [37] using functions computed by single-hidden-layer networks of perceptrons;
- Attali&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 were obtained *provided that sufficiently many hidden units are available.* More constructive solutions have been obtained in very small *depth* [38,54,55], 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 [70–72].

B. 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 [51]. The first lower bound of:

$$size \geq 2\,(2^n/n)^{1/2} \tag{2}$$

on the *size* of a TGC for 'almost all' n-ary BFs (*i.e.*, $f: \mathbb{B}^n \rightarrow \mathbb{B}$) was given in [53]. Later [48] a very tight upper bound was proven in $depth = 4$:

$$size \leq 2\,(2^n/n)^{1/2} \times \{1 + \Omega\,[(2^n/n)^{1/2}]\}. \tag{3}$$

A similar existence exponential lower bound of $\Omega\,(2^{n/3})$ for arbitrary BFs can be found in [65] (see also [63]).

For classification problems, 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 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 tighter bound of $\lceil 1 + (m-2)/n \rceil$ 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 [33]; 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 $m/3$;
- Beiu&De Pauw [10] have detailed existence lower and upper bounds: $2m/(n\log n) < size < 1.44m/n$ by estimating the entropy of the data-set ([11,13]).

Other existence lower bounds for the arbitrary dichotomy problem can be found in [25,59]:

- a *depth*-2 TGC requires $m/\{n\log(m/n)\}$ TGs;
- a *depth*-3 TGC requires $2\,(m/\log m)^{1/2}$ 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, and under piecewise linear approximations.

C. Boolean Functions

The particular case of BFs has been studied intensively [9,56]. Many results have been obtained for particular BFs [63,65], but a *size*-optimal result for BFs that have m groups of ones in their truth table $\mathbb{F}_{n,m}$ (*i.e.*, BFs defined on the m groups) was detailed by Red'kin in 1970 [62].

Theorem from [62] *The complexity realisation (i.e., number of elements) of $\mathbb{F}_{n,m}$ is at most $2\,(2m)^{1/2} + 3$.*

This result is valid for unlimited *fan-in* TGs. Departing from these lines, Horne&Hush [28] have detailed a solution for limited *fan-in* TGCs.

Theorem from [28] *Arbitrary BFs $f: \{0,1\}^n \rightarrow \{0,1\}^m$ 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.*

II. SIZE-OPTIMAL IMPLEMENTATIONS

Implementing arbitrary BFs using classical Boolean gates (*i.e.*, AND and OR) requires exponential *size*. The known bounds for *size* are also exponential if TGCs are used for solving arbitrary BFs [6]. These bounds reveal exponential gaps, and also suggest that TGCs with more layers ($depth \neq$ constant [8,12]) might have a smaller *size*.

Another 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 is clearly *size*-optimal. We start from [70–72], where a constructive solution for the general case was detailed.

Theorem from [70] *Define the function $\psi : \mathcal{E} \rightarrow \mathcal{D}$ such that for each integer $k \in N$:*

$$\psi\left(\sum_{r=1}^{k} i_r \gamma^{-r}\right) = \sum_{r=1}^{k} \tilde{i}_r\, 2^{-m_r}\, \gamma^{-\frac{n^{r-m_r}-1}{n-1}} \tag{4}$$

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

$$m_r = \langle i_r \rangle\,\{1 + \textstyle\sum_{s=1}^{r-1} [i_s] \times \ldots \times [i_{r-1}]\} \tag{5}$$

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

Here $\gamma \geq 2n + 2$ is a base, $\mathcal{E} = [0, 1]$ is the unit interval, $\mathcal{D}$ is the set of 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 $i_r = 0, 1, \ldots, \gamma - 2$, $\langle i_r \rangle = 1$ when $i_r = \gamma - 1$, $[i_r] = 0$ when $i_r = 0, 1, \ldots, \gamma - 3$, 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 $\psi\,(0.i_1) = 0.i_1$, *i.e.*, the identity function $\psi\,(x) = x$. Such a solution builds 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 [52,56,61,66]:

$$2^{(\Delta-1)/2} < weight < (\Delta + 1)^{(\Delta+1)/2} / 2^{\Delta}. \tag{6}$$

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 ψ (eq. 4), and for the *weights*:

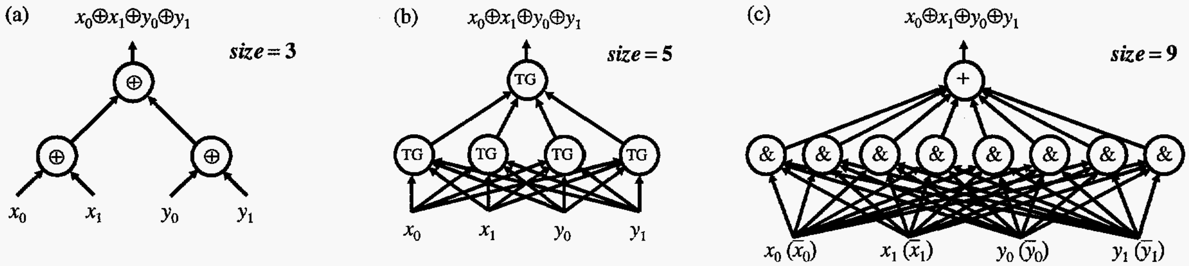

Figure 1. The PARITY problem: (a) solution using XORs gates; (b) solution using TGs; (c) solution using AND-OR gates.

$$\alpha_p = \sum_{r=1}^{\infty} \gamma^{-(p-1)\frac{n^r-1}{n-1}}. \tag{7}$$

For BFs the precision is reduced to $(2n+2)^{-n}$, or $2n\log n$ bits per *weight*. Analog implementations are limited to just several bits [42], this being one of the reasons for investigations on precision [18,27,73,74], and on algorithms relying on limited integer *weights* [7,19,39]. As an example, let us consider the PARITY function of four bits. It is known that PARITY can be implemented with three 2-input XOR gates (Fig. 1.a), or with five 4-input TGs (Fig. 1.b). It is also known that a 4-input BF requires $2\sqrt{16}+3 = 11$ TGs [62]. A classical Boolean solution requires eight AND gates (Fig. 1.c). A brute force solution would approximate a 4-input BF (Fig. 2.a and 2.b). Another solution having COMPARISON as the ψ function, and translated inputs is presented in Fig. 2.c; it . The $2n+1=5$ hidden functions (Φ_q) are simple AND functions, while the addition is an OR function.

Due to the limited precision, an optimal solution for implementing BFs should decompose the given BF in simpler BFs (which can be implemented based on Kolmogorov's superpositions). The partial results from this first layer can be combined using (again) Kolmogorov's superpositions. Such an analog implementation will requires more than three layers. A systematic solution which would utilise silicon to the best advantage would be to rewrite a given computation (*i.e.*, set of BFs) in a base larger than 2, and use Kolmogorov's superpositions for implementing the digit-wise computations in this larger base.

IV. CONCLUSIONS

Arbitrary BFs can be implemented using: (*i*) Boolean circuits, but require exponential *size*; (*ii*) TGs, but (again) in exponential *size*; (*iii*) analog building blocks in linear *size* (having linear *fan-in* and polynomial precision).

The main conclusion is that *size*-optimal hardware implementations of BFs can be obtained using analog (or mixed analog/digital) circuitry. The high precision required by the solution based on Kolmogorov's superpositions can be tackled by decomposing a BFs into simpler BFs. This is mathematically equivalent to computing in a larger base. Due to the reduced number of inputs, Kolmogorov's superpositions can be used to design the analog implementations of the digit-wise computations in such larger bases.

References

[1] M. Arai. Bounds on the Number of Hidden Units in Binary-valued Three-layer Neural Networks. *Neural Networks*, **6**(6), pp. 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. Pagès. Approximations of Functions by a Multilayer Perceptron: a New Approach. *Neural Networks*, **10**(6), pp. 1069-1081, 1997.

[4] A.R. Barron. Universal Approximation Bounds for Superpositions of a Sigmoidal Function. *IEEE Trans. Info. Theory*, **39**(3), pp. 930-945, 1993.

[5] E.B. Baum. On the Capabilities of Multilayer Perceptrons. *J. Complexity*, **4**, pp. 193-215, 1988.

[6] V. Beiu. Digital Integrated Circuit Implementations. *Chapter E1.4* in [20], 1996.

[7] V. Beiu. Reduced Complexity Constructive Learning Algorithm. *Neural Network World*, **8**(1), pp. 1-38, 1998.

[8] V. Beiu. On the Circuit and VLSI Complexity of Threshold Gate COMPARISON. *Neurocomputing*, **19**(1), pp. 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. In J. Mira, R. Moreno-Díaz & J. Cabestany (eds.): *Biological and Artificial Computation*, Springer, Berlin, pp. 743-752, 1997.

[11] V. Beiu & S. Drăghici. Limited Weights Neural Networks: Very Tight Entropy Based Bounds. In D.W. Pearson (ed.): *Proc. ICSC Symp. on Soft Computing SOCO'97*, ICSC Acad. Press, Canada, pp. 111-118, 1997.

[12] V. Beiu & H.E. Makaruk. Deeper Sparser Nets Can Be Optimal. *Neural Proc. Lett.*, **7/8**, 1998.

[13] V. Beiu, S. Drăghici & T. De Pauw. A Constructive Approach to Calculating Lower Entropy Bounds. *Neural Proc. Lett.*, **7/8**, 1998.

[14] E. Blum & K. Li. Approximation Theory and Feedforward Networks. *Neural Networks*, **4**(3), pp. 511-515, 1991.

[15] A. Bulsari. Some Analytical Solutions to the General Approxima-

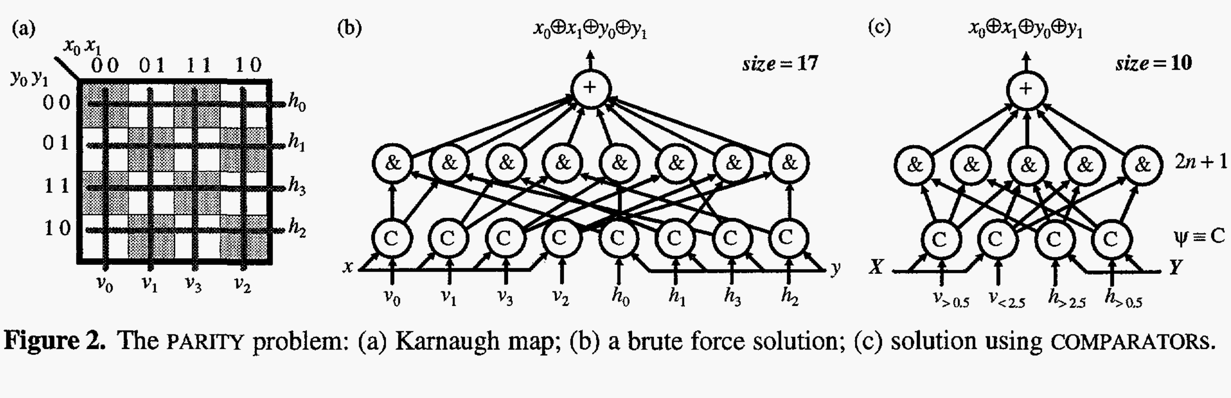

Figure 2. The PARITY problem: (a) Karnaugh map; (b) a brute force solution; (c) solution using COMPARATORS.

tion Problem for Feedforward Neural Networks. *Neural Networks*, **6**(7), pp. 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**, pp. 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*, AIP, NY, pp. 219-222, 1988.

[19] S. Drăghici & 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, pp. 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), pp. 183-192, 1989.

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

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

[24] E. Hartman, J.D. Keeler & J.M. Kowalski. Layered Neural Networks with Gaussian Hidden Units as Universal Approximations. *Neural Computation*, **2**(2), pp. 210-215, *1989.*

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

[26] R. Hecht-Nielsen. Kolmogorov's Mapping Neural Network Existence Theorem. *Proc. Intl. Conf. on Neural Networks ICNN'87*, IEEE CS Press, CA, pp. 11-14, 1987.

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

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

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

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

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

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

[33] 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), pp. 47-55, 1991.

[34] B. Irie & S. Miyake. Capabilities of Three-Layered Perceptrons. *Proc. Intl. Conf. on Neural Networks ICNN'88*, IEEE CS Press, CA, vol. **1**, pp. 641-648, 1988.

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

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

[37] 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), pp. 608-613, 1992.

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

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

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

[41] 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**, pp. 953-956, 1957; English transl.: *Trans. Amer. Math. Soc.*, **2**(28), pp. 55-59, 1963.

[42] A.H. Kramer. Array-Based Analog Computation; Principles, Advantages and Limitations. *Proc. Microelectronics for Neural Networks MicroNeuro'96*, IEEE CS Press, CA, pp. 68-79, 1996.

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

[44] 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), pp. 1061-1068, 1997.

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

[46] 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), pp. 861-867, 1993.

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

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

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

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

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

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

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

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

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

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

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

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

[59] T. Poggio & F. Girosi. A Theory of Networks for Approximation and Learning. *Tech. Rep. AI Memo 1140*, MIT, MA, *1989*; short version as: Networks for Approximation and Learning, *Proc. IEEE*, **78**(9), pp. 1481-1497, 1990.

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

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

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

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

[64] 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), pp. 15-37, 1998.

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

[66] 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.

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

[68] 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**, pp. 98-105, 1966.

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

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

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

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

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

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