On the Sample Complexity of Learning for Networks of Spiking Neurons with Nonlinear Synaptic Interactions*

Michael Schmitt†

Lehrstuhl Mathematik und Informatik, Fakultät für Mathematik
Ruhr-Universität Bochum, D–44780 Bochum, Germany
http://www.ruhr-uni-bochum.de/lmi/mschmitt/
mschmitt@lmi.ruhr-uni-bochum.de

Abstract

We study networks of spiking neurons that use the timing of pulses to encode information. Nonlinear interactions model the spatial groupings of synapses on the neural dendrites and describe the computations performed at local branches. Within a theoretical framework of learning we analyze the question of how many training examples these networks must receive to be able to generalize well. Bounds for this sample complexity of learning can be obtained in terms of a combinatorial parameter known as the pseudo-dimension. This dimension characterizes the computational richness of a neural network and is given in terms of the number of network parameters. Two types of feedforward architectures are considered: constant-depth networks and networks of unconstrained depth. We derive asymptotically tight bounds for each of these network types: Constant depth networks are shown to have an almost linear pseudo-dimension, whereas the pseudo-dimension of general networks is quadratic.

Networks of spiking neurons that use temporal coding are becoming increasingly more important in practical tasks such as computer vision,

†Results in this article have been presented at the International Conference on Artificial Neural Networks ICANN 2001 in Vienna [29]. Part of the research has been conducted while the author was with the Universität Osnabrück.

speech recognition, and motor control. The question of how well these networks generalize from a given set of training examples is a central issue for their successful application as adaptive systems. The results show that, although coding and computation in these networks is quite different and in many cases more powerful, their generalization capabilities are at least as good as those of traditional neural network models.

Keywords: pulsed neural networks, spike timing, computational learning theory, generalization capability, pseudo-dimension

1 Introduction

One of the computational features that is most common to artificial neural networks is the combination of synaptic input signals using a summation. This is a remarkable fact, as results from neurobiology have exhibited several exemplars of nerve cells that accomplish complex tasks through nonlinear synaptic interactions [10, 16, 17, 18]. Although in recent years artificial neurons have come closer to the biological paradigm and, in particular, spiking neuron models that use temporal coding are currently a highly active area of research [11, 14, 20, 25, 27], nonlinearity in dendritic processing has not played a major role, despite experimental evidence.

In this article, we study a model of spiking neurons that uses temporal coding and employs nonlinear operations for synaptic interactions. This model encodes information in terms of single firing events. Thus, it takes account of the fact that examples of biological information processing have been found supporting the hypothesis that, in contrast to the firing rate, the timing of the first spike is the crucial quantity [9, 13, 24]. Models with linear interactions are sufficient for capturing the passive properties of the dendritic membrane where synaptic inputs occur as currents that are combined using a summing operation. The nonlinearities of synaptic computations included in the model considered here reflect the spatial groupings of synapses on the dendrites (also called "synaptic clusters" [22]) and model the information processing steps performed at local branches. Rectangular pulses are used to represent the postsynaptic responses evoked by input spikes. The nonlinearities employed for synaptic interactions are given in terms of arithmetic operations that can easily be implemented in analog VLSI circuits and thus be used for pulsed neural hardware [2, 23, 26]. We allow the nonlinear synaptic interactions to be from a comprehensive class: the set of rational functions. These nonlinearities comprise not only multiplicative but also divisive operations and account for the finding that dendrites are able to compute division [5, 8].

Networks of spiking neurons have recently been studied with regard to their capabilities for learning [6, 7, 19, 28]. In addition to the tuning of the synaptic efficacies, the ability to adjust the transmission delays between neurons is a feature that is considered to be relevant for learning mechanisms [15]. Our model includes not only the classical synaptic weights as parameters, but also the transmission delays.

We investigate the complexity of learning for these networks in terms of a combinatorial parameter known as the pseudo-dimension. The Vapnik-Chervonenkis (VC) dimension and its generalization to real-valued functions, the pseudo-dimension, are well established measures for the sample complexity of learning [1, 3, 31]. They provide bounds for the number of training examples that are required by a wide class of learning algorithms for producing results with good generalization performance. We consider feedforward networks and calculate tight bounds for their pseudo-dimension. In particular, we show that the pseudo-dimension is bounded by $\Theta(W \log W)$ for networks with constant depth and degree-limited synaptic interactions, where W is the number of network parameters. Further, we derive the bound $\Theta(W^2)$ for networks without depth constraints. These two major results indicate that the property of small depth, which is prevailing in biological as well as in artificial neural networks, must be taken into account for obtaining good pseudo-dimension bounds. For traditional networks consisting of neurons with polynomial activation functions but without adjustable delays, analogous results have been obtained previously [4, 12]. The results established here imply that, although temporal coding, synaptic nonlinearities, and adjustable delays enhance the computational power of these networks considerably, the sample complexity of learning does not become prohibitive.

In the following section, we give the definition of the network model. Section 3 introduces the VC dimension and the pseudo-dimension. The results for depth-restricted networks are derived in Section 4, while Section 5 deals with arbitrary architectures. Section 6 discusses the results and Section 7 concludes with some closing remarks.

2 Spiking Neurons with Nonlinear Synaptic Interactions

In a network of *spiking neurons*, each node receives inputs as spikes through its synaptic connections from other nodes. We characterize each synapse by a weight $w_i \in \mathbb{R}$ and a transmission delay $d_i \in \mathbb{R}^+$, where we let $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$. If the neuron that is presynaptic to connection i emits a spike at time $\tau_i \in \mathbb{R}^+$, this generates after a delay of d_i in the postsynaptic neuron a rectangular pulse that has unit duration and a height corresponding to the strength, or efficacy, w_i of the synapse. This pulse is called the *postsynaptic pulse* of synapse i and described by a function $t \mapsto h_i(t - \tau_i)$, where $h_i : \mathbb{R}^+ \to \mathbb{R}$ is defined as

$$h_i(t) \quad = \quad \begin{cases} w_i & \text{if } d_i \leq t < d_i + 1, \\ 0 & \text{otherwise.} \end{cases}$$

We consider the delays d_i and the weights w_i as adjustable parameters of the neuron.

Assume that neuron v has n input connections. A *synaptic cluster* of v represents a group of synapses that interact nonlinearly. It is specified by a subset $I \subseteq \{1, \ldots, n\}$. A synapse may be a member of more than one cluster as it may be the case in a biological neuron that a synapse participates in the computation of several spatial groups on the dendritic tree. The nonlinearity is described for every synaptic cluster $\{i_1, \ldots, i_l\} \subseteq I$ by a rational function $f_{\{i_1,\ldots,i_l\}}$ in l variables. We call the function $f_{\{i_1,\ldots,i_l\}}$ the *synaptic interaction* of $\{i_1, \ldots, i_l\}$ and assume without loss of generality that the denominator of $f_{\{i_1,\ldots,i_l\}}$ is never zero. (This requirement does not impose constraints on the modeling of biologically realistic nonlinearities.)

We consider the computations of the network of spiking neurons during a given time interval $[t_0, t_1] \subseteq \mathbb{R}^+$. Let $F \subseteq \{1, \ldots, n\}$ be the set of synapses of neuron v that receive a spike during this interval. For every synaptic cluster I we introduce the function $J_I : [t_0, t_1] \to 2^I$ defined in such a way that $J_I(t)$ is the set of synapses in I that are simultaneously active at time t, that is,

$$J_I(t) \quad = \quad \{i \in I \cap F : h_i(t - \tau_i) \neq 0\}.$$

The interaction of the synapses in I is specified by the function $M_I : [t_0, t_1] \to \mathbb{R}$ with

$$M_I(t) = f_{\{i_1,\ldots,i_l\}}(w_{i_1}, \ldots, w_{i_l}), \quad \text{where } \{i_1, \ldots, i_l\} = J_I(t).$$

Assume that neuron v has k synaptic clusters $I_1, \ldots, I_k$. The *membrane potential* $P_v : [t_0, t_1] \to \mathbb{R}$ of v is the sum

$$P_v(t) \quad = \quad \sum_{j=1}^{k} M_{I_j}(t). \tag{1}$$

Thus, the time course of the membrane potential is represented by a sequence of pulses. An example is shown in Figure 1. If the synaptic clusters are pairwise disjoint and the synaptic interactions have degree one, we regain the model with linear synaptic interactions introduced in [21].

Neuron v fires when its membrane potential reaches a value specified by another parameter of v: the threshold θ. The *firing time* t_v^f of v is defined by

$$t_v^f \quad = \quad \min\{t \in [t_0, t_1] : P_v(t) \geq \theta\}.$$

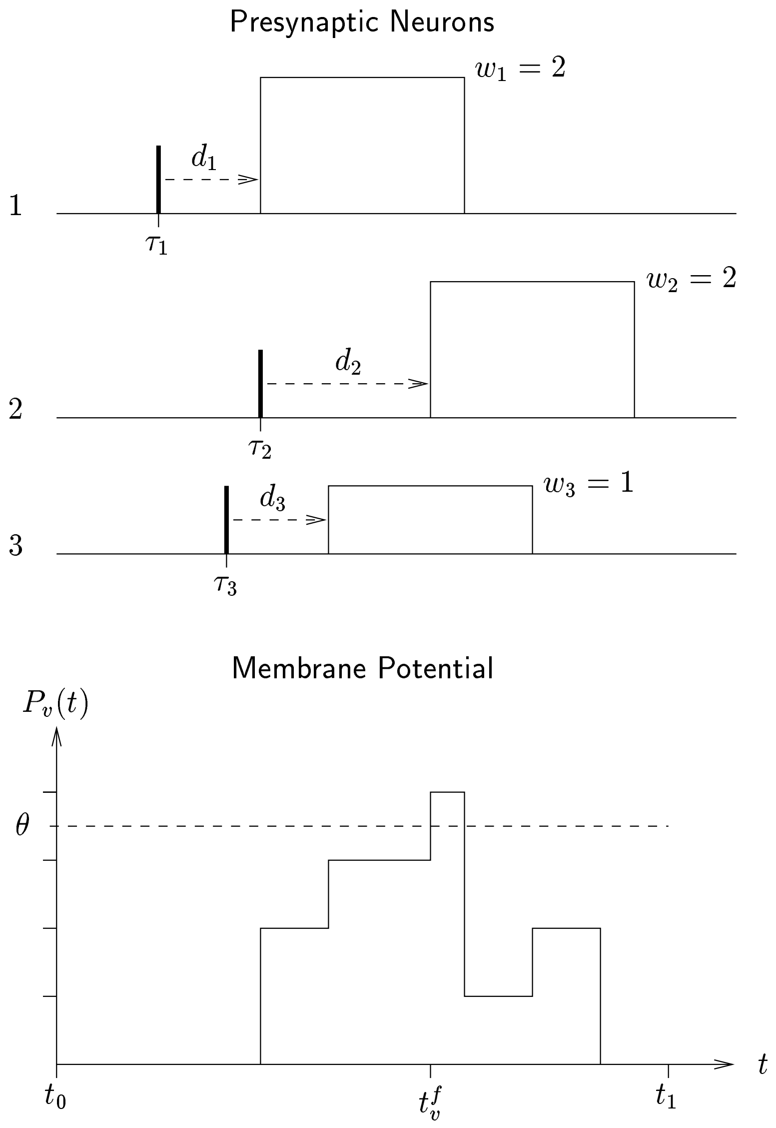

Figure 1: The time course of the membrane potential $P_v(t)$ for a neuron v with three presynaptic neurons and a single synaptic cluster $I = \{1,2,3\}$, implying that $P_v(t) = M_{\{1,2,3\}}(t)$. Each presynaptic neuron $i = 1,2,3$ emits a spike at time τ_i which generates a postsynaptic pulse at time $\tau_i + d_i$. The value of $P_v(t)$ is determined by the synapses that are active at time t (see equation (1)). The synaptic interactions are defined as $f_{\{1\}}(w_1) = w_1, f_{\{1,3\}}(w_1, w_3) = 2w_1w_3 - w_1/2, f_{\{1,2,3\}}(w_1, w_2, w_3) = w_1w_2w_3, f_{\{2,3\}}(w_2, w_3) = w_2w_3/2$, and $f_{\{2\}}(w_2) = w_2$.

If t_v is undefined, v does not fire. The networks of spiking neurons that we consider use temporal coding, that is, they encode information in the timing of single firing events. Precisely, if neuron v fires at time $t_v^f \in \mathbb{R}^+$, this event is supposed to encode the real number t_v^f. We consider feedforward networks with a given number of input nodes and one output node. Input vectors enter the network in terms of firing times of the input nodes. The output value of the network is the firing time of the output node. If the output node does not fire, the network output is defined to be 0. We also impose restrictions on the *depth* of networks, that is, the length of the longest path, measured as the number of connections, from an input node to the output node.

3 VC Dimension and Pseudo-Dimension

The VC dimension and the pseudo-dimension are based on the notion of shattering: A class $\mathcal{F}$ of $\{0,1\}$-valued functions in n variables is said to *shatter* a set $S \subseteq \mathbb{R}^n$ if $\mathcal{F}$ induces all dichotomies on S, that is, if for every partition of S into two disjoint subsets (S_0, S_1) there is some $f \in \mathcal{F}$ satisfying

$$f(S_0) \subseteq \{0\} \quad \text{and} \quad f(S_1) \subseteq \{1\}.$$

The *Vapnik-Chervonenkis (VC) dimension* of $\mathcal{F}$ is the cardinality of the largest set shattered by $\mathcal{F}$.

If $\mathcal{F}$ is a class of real-valued functions in n variables, the *pseudo-dimension* of $\mathcal{F}$ is defined as the VC dimension of the function class

$$\begin{aligned}\{g : \mathbb{R}^{n+1} \to \{0,1\} \mid & \text{ there is some } f \in \mathcal{F} \\ & \text{such that for all } x \in \mathbb{R}^n \text{ and } y \in \mathbb{R} : g(x,y) = \operatorname{sgn}(f(x) - y)\},\end{aligned}$$

where $\operatorname{sgn} : \mathbb{R} \to \{0,1\}$ is the function satisfying $\operatorname{sgn}(z) = 1$ if $z \geq 0$, and 0 otherwise. Thus, the pseudo-dimension can be considered as the VC dimension of a function class that is forced to be $\{0,1\}$-valued by imposing a threshold on the output value, where each input vector x may have its own threshold y.

The VC dimension of a network of spiking neurons is defined to be the VC dimension of the set of functions computed by the network with all possible assignments of values to its adjustable parameters, that is, its delays, weights, and thresholds. The same holds for the pseudo-dimension. To define the VC dimension for a network with real-valued output we assume that the output values are mapped to $\{0,1\}$ by specifying some constant threshold. Clearly, the VC dimension of a network is not larger than its pseudo-dimension.

4 Networks of Bounded Depth

In the following we provide an upper bound on the pseudo-dimension for networks of spiking neurons with depth restrictions. The bound is almost linear in the number of network parameters (delays, weights, and thresholds) and the network depth. At the end of the section we show that for constant-depth networks with rational synaptic interactions of bounded degree, the result is asymptotically tight. We recall that the degree of a rational function is defined as the sum of the degrees of the numerator and denominator polynomial.

Theorem 1. *Let $\mathcal{N}$ be a network of spiking neurons with W parameters and depth D. Assume that each neuron employs rational synaptic interactions with degree no larger than p. Then the pseudo-dimension of $\mathcal{N}$ is at most*

$$2(2WD+W)\log\left(\frac{2(2WD+W)}{e\ln 2}\right) + 2WD\log(16e^2W(p+1)) \\ + 2(W\log(2e)+D),$$

and, hence, $O(WD\log(WDp))$. For fixed depth D and degree p, this entails the bound $O(W\log W)$.

The proof is based on the method of solution set components bounds [1]. Assuming a set of input vectors, we derive a set of rational functions in the parameter variables of $\mathcal{N}$. Then we calculate a bound for the number of connected components into which the parameter domain is partitioned by the zero-sets of these rational functions. (A connected component of a set R is a maximal subset Q of R such that any two points are connected by a continuous curve that is contained in Q.) This bound limits the number of dichotomies that $\mathcal{N}$ induces on the set of input vectors. Assuming that this set is shattered, we obtain the bound on the pseudo-dimension.

We require some definitions from [1]: A set $\{f_1,\ldots,f_k\}$ of differentiable real-valued functions on $\mathbb{R}^d$ has *regular zero-set intersections* if for every non-empty set $\{i_1,\ldots,i_l\} \subseteq \{1,\ldots,k\}$ the Jacobian (that is, the matrix of the partial derivatives) of $(f_{i_1},\ldots,f_{i_l}) : \mathbb{R}^d \to \mathbb{R}^l$ has rank l at every point of the set

$$\{a \in \mathbb{R}^d : f_{i_1}(a) = \cdots = f_{i_l}(a) = 0\}.$$

A class $\mathcal{G}$ of real-valued functions defined on $\mathbb{R}^d$ has *solution set components bound B* if for every $k \in \{1,\ldots,d\}$ and every $\{f_1,\ldots,f_k\} \subseteq \mathcal{G}$ that has regular zero-set intersections, the number of connected components of the set

$$\{a \in \mathbb{R}^d : f_1(a) = \cdots = f_k(a) = 0\}$$

is at most B. In the proof of the theorem, we make use of the following solution set components bound.

Lemma 2. *Consider the class of functions from $\mathbb{R}^d$ to $\mathbb{R}$ that can be represented as a finite sum where each term is a rational function with degree no more than p and has a denominator that is never zero. This class has solution set components bound $2(2p+2)^d$.*

Proof. The result is a special case of [1, Lemma 8.16]. A rational function q/r can be expressed in terms of a polynomial equation by introducing a new variable z and a new function g in z and the variables x of q/r defined as

$$g(x, z) \quad = \quad zr(x) - q(x).$$

Then, g is a polynomial of degree at most $p+1$. Thus, sums of rational functions of degree at most p can be represented by polynomials of degree no more than $p+1$. The class of polynomials of degree at most $p+1$ in d variables has solution set components bound $2(2p+2)^d$ [1, Corollary 8.2]. The assumptions together with the reasoning provided in [1, Lemma 8.16] yield that this value is also a solution set components bound for the class considered here. □

A class $\mathcal{F}$ of real-valued functions is *closed under addition of constants* if for every $c \in \mathbb{R}$ and $f \in \mathcal{F}$, the function $z \mapsto f(z) + c$ is also a member of $\mathcal{F}$. A proof of the following result is provided in [30, Lemma 3].

Lemma 3. *Let $\mathcal{F}$ be a class of real-valued functions $(y_1, \ldots, y_d, x_1, \ldots, x_n) \mapsto f(y_1, \ldots, y_d, x_1, \ldots, x_n)$ that is closed under addition of constants and where each function in $\mathcal{F}$ is C^d in the variables $y_1, \ldots, y_d$. If the class $\mathcal{G} = \{(y_1, \ldots, y_d) \mapsto f(y_1, \ldots, y_d, s) : f \in \mathcal{F}, s \in \mathbb{R}^n\}$ has solution set components bound B then for any sets $\{f_1, \ldots, f_k\} \subseteq \mathcal{F}$ and $\{s_1, \ldots, s_m\} \subseteq \mathbb{R}^n$, where $m \geq d/k$, the function $T : \mathbb{R}^d \to \{0,1\}^{mk}$ defined by*

$$T(a) = (\mathrm{sgn}(f_1(a, s_1)), \ldots, \mathrm{sgn}(f_1(a, s_m)), \ldots, \mathrm{sgn}(f_k(a, s_1)), \ldots, \mathrm{sgn}(f_k(a, s_m)))$$

partitions $\mathbb{R}^d$ into at most

$$B \sum_{i=0}^{d} \binom{mk}{i} \quad \leq \quad B(emk/d)^d$$

equivalence classes (where $a_1, a_2 \in \mathbb{R}^d$ are equivalent if and only if $T(a_1) = T(a_2)$).

We now have all that is needed for the proof of the main result.

Proof of Theorem 1. Given $\mathcal{N}$ as supposed, let $\{s_1, \ldots, s_m\}$ be a set of input vectors and $u_1, \ldots, u_m$ real numbers. The main aim is to derive a bound on the number of dichotomies induced on $S = \{(s_1, u_1), \ldots, (s_m, u_m)\}$ by functions of the form $(x, y) \mapsto \mathrm{sgn}(f(x) - y)$ where f is computed by $\mathcal{N}$. We proceed inductively through the levels of the network, from the input nodes towards the output node. The level of a node v is defined as the length of the longest path from an input node to v. Thus, input nodes have level 0 and the level of the output node is equal to the depth of the network.

We derive sets $\mathcal{G}_\lambda$, $\lambda \geq 0$, of functions in the network parameters that satisfy the following condition: The set $\mathcal{G}_\lambda$ partitions the parameter domain $\mathbb{R}^W$ into equivalence classes such that for all parameter vectors within the same class, the computations of the nodes of level $0, \ldots, \lambda$ on inputs from S remain unchanged in the following sense: For each node, consider the successive subsets of synapses that are simultaneously active. This sequence of subsets remains the same and, if the node fires, the firing is triggered by the same set of (starting or ending) postsynaptic pulses.

Clearly, as the input nodes have no parameters, we may set $\mathcal{G}_0 = \emptyset$. Hence, there is only one class of parameters, which is the entire $\mathbb{R}^W$. This constitutes the base of the induction.

Suppose that $\mathcal{G}_{\lambda-1}, \lambda \geq 1$ has been obtained with the properties stated above. Let $(s, u) \in S$ be given. Assume that v is some node of level λ that receives spikes through its i-th and j-th synapse with corresponding delay parameters $d_{v,i}$ and $d_{v,j}$. To specify the temporal order of starting and ending points of postsynaptic pulses caused by the presynaptic spikes, we introduce the expressions

$$r_{v,i} + d_{v,i} + b_i - r_{v,j} - d_{v,j} - b_j. \tag{2}$$

Here $r_{v,i}$ and $r_{v,j}$ are the firing times of the corresponding presynaptic nodes. (If $\lambda = 1$ then $r_{v,i}$ and $r_{v,j}$ are firing times of input nodes.) Further, b_i and b_j are binary values depending on whether we refer to the starting or ending point of the corresponding postsynaptic pulse, respectively. Thus, the sign of the expression (2) indicates the temporal order of these points. As a function of $d_{v,i}, d_{v,j}$, the expression is affine in these delay variables.

Let C_λ denote the total number of synapses of nodes of level λ. Varying over i, j and the possible assignments of values to b_i, b_j, there are at most $(2C_\lambda)^2$ functions of the form (2) for the given input vector (s, u). The union of these functions partitions the parameter domain of the delays of the nodes of level λ into equivalence classes such that for parameters within the same class, the temporal order of postsynaptic pulses for each of these nodes does not change. This implies that for input vector (s, u) and each node of level λ, the sequence of subsets of simultaneously active synapses remains unchanged within the same class of delay parameters.

Having taken into account the delays of level λ, we next consider its weights and thresholds. We partition the space of these parameters into equivalence classes such that for each node on level λ the same set of postsynaptic pulses (some starting, some ending) is responsible for the firing of the node when choosing weights and thresholds within the same class. (This also includes the possible non-firing of the node.) This then completes the induction step.

For given input vector $(s, u) \in S$, node v of level λ, and time $t \in [t_0, t_1]$, consider the function

$$\sum_{j=1}^{k} M_{I_{v,j}}(t) - \theta_v, \tag{3}$$

where $I_{v,1}, \ldots, I_{v,k}$ are the synaptic clusters of v and θ_v is its threshold. Let n be the number of synapses of v. If the delay values for the nodes of level at most λ remain within the same class, the firing of v on input (s, u) depends on at most $2n$ functions of the form (3). This follows from the fact that for fixed delays, the subset of synapses that are active at a certain point in time can change only when some postsynaptic pulse starts or ends. Thus, the $2n$ starting and ending points define no more than $2n-1$ intervals during each of which a particular subset of synapses is active. (We use the weaker bound $2n$ for simplicity.) Considering all nodes of level λ, which have C_λ synapses altogether, there are at most $2C_\lambda$ functions that partition the parameter space of the weights and thresholds into equivalence classes such that for each node the same set of postsynaptic pulses triggers its firing. The partitioning of the parameter domain arising from these is such that for the given input vector (s, u), the firing of the nodes on level λ remains unchanged. Each function is a sum of at most k rational functions in the weights and thresholds of level λ, where each term in the sum has degree at most p.

Using solution set components bounds, we are now able to estimate the number of equivalence classes. Clearly, the class of affine functions has solution set components bound $B = 1$. According to Lemma 2, the class of sums of rational functions in d variables with degree at most p has solution set components bound $B = 2(2p+2)^d$. Both function classes are closed under addition of constants. Let W_λ denote the total number of parameters of nodes of level λ. For the given set of m input vectors, we apply Lemma 3 to the two sets of functions specified above. The first set consists of at most $(2C_\lambda)^2$ affine functions, the second set has no more than $2C_\lambda$ rational functions of degree at most p. Thus, we obtain for the number of equivalence classes generated by the functions defined in (2) and (3) the upper bound

$$(em(2C_\lambda)^2/W_\lambda)^{W_\lambda} \cdot 2(2p+2)^{W_\lambda} \cdot (2emC_\lambda/W_\lambda)^{W_\lambda}, \tag{4}$$

where the first factor is due to the affine functions and the last two factors take account of the rational functions.

Bound (4) is derived assuming a given class in the partitioning of the parameter domain that arises from the function class $\mathcal{G}_{\lambda-1}$. Defining $\mathcal{G}_\lambda$ to include $\mathcal{G}_{\lambda-1}$ and the functions from (2) and (3), it follows by the induction hypothesis that the product of expression (4), for $\lambda = 1, \ldots, l$, provides an upper bound on the number of equivalence classes for the parameter domain of all nodes of level at most l. At level D, we additionally have to account for the threshold values $u_1, \ldots, u_m$. For these we include $2mC_D$ additional affine functions

$$r_{v,i} + d_{v,i} + b_i - u,$$

with $b_i \in \{0, 1\}$ and $u \in \{u_1, \ldots, u_m\}$, to describe the starting or ending of postsynaptic pulses relative to u. By Lemma 3, they give rise to at most

$(2emC_D/W_D)^{W_D}$ equivalence classes. Combining all levels, we eventually obtain for the number of equivalence classes of the parameter domain of the entire network the upper bound

$$\left(\frac{2emC_D}{W_D}\right)^{W_D} \cdot \prod_{\lambda=1}^{D} \left[\left(\frac{em(2C_\lambda)^2}{W_\lambda}\right)^{W_\lambda} \cdot 2(2p+2)^{W_\lambda} \cdot \left(\frac{2emC_\lambda}{W_\lambda}\right)^{W_\lambda} \right],$$

such that within each class the same dichotomy is induced on S via the function $(x, y) \mapsto \operatorname{sgn}(f(x)-y)$, where f is computed by $\mathcal{N}$. Using $C_\lambda \leq W_\lambda$ and $W_\lambda \leq W$ for $\lambda = 1, \ldots, D$ yields the upper bound

$$2^D(2em)^W(16e^2m^2W(p+1))^{WD}.$$

If S is shattered, 2^m dichotomies must be induced and hence at least this many equivalence classes for the parameter domain of the network must exist. This implies that

$$m \leq (2WD+W)\log m + WD\log(16e^2W(p+1)) + W\log(2e) + D. \quad (5)$$

Now we use the well known fact that for every $\alpha, \beta > 0$, the inequality $\ln\alpha \leq \alpha\beta + \ln(1/\beta) - 1$ holds [1, Appendix A.1.1]. This yields for $\alpha = m$ and $\beta = (\ln 2)/(2(2WD+W))$,

$$(2WD+W)\log m \leq \frac{m}{2} + (2WD+W)\log\left(\frac{2(2WD+W)}{e\ln 2}\right).$$

Substituting this in the right-hand side of (5), we obtain

$$\begin{aligned} m \leq\ & 2(2WD+W)\log\left(\frac{2(2WD+W)}{e\ln 2}\right) \\ & + 2WD\log(16e^2W(p+1)) + 2(W\log(2e)+D), \end{aligned}$$

which is the claimed bound on the pseudo-dimension. □

If the depth and the degree of synaptic interactions are fixed, the previous theorem yields an asymptotically tight bound as provided by the following statement. The result also improves the quadratic bound for networks with linear interactions given in [21].

Corollary 4. *Consider the class of networks of spiking neurons where the depth and the degree of the rational synaptic interactions are bounded by constants. Suppose that $\mathcal{N}$ is a network from this class and has W parameters. Then the pseudo-dimension of $\mathcal{N}$ is $\Theta(W\log W)$.*

Proof. The upper bound is due to Theorem 1, the lower bound is given by Theorem 2.2 in [21], where it was shown that a single spiking neuron with W delay parameters has VC dimension $\Omega(W\log W)$. □

5 Networks of Arbitrary Depth

As the depth of a network is not larger than the number of parameters, Theorem 1 can be used to obtain for networks of arbitrary depth a bound that does not depend on the depth. A direct method yields a better bound that is even asymptotically tight for networks with rational synaptic interactions of fixed degree.

Theorem 5. *Let $\mathcal{N}$ be a network of spiking neurons with W parameters and rational synaptic interactions of degree at most p. Then the pseudo-dimension of $\mathcal{N}$ is at most*

$$4W^2 + 2W \log\left(\frac{64W(p+1)}{(\ln 2)^2}\right) + 2,$$

and hence bounded by $O(W^2 \log p)$.

Proof. The reasoning is similar to the proof of Theorem 1. Given a set $S = \{(s_1, u_1), \ldots, (s_m, u_m)\}$, we introduce functions that partition the parameter domain $\mathbb{R}^W$ of the network. First, we specify the possible firing times of the output node in response to an input vector s_i and relative to the output threshold u_i. This is accomplished by the functions

$$s_{i,\mu} + \sum_{(v,j)\in P} (d_{v,j} + b_{v,j,l}) - u_i, \tag{6}$$

where $s_{i,\mu}$ is the firing time of the μ-th input node, $d_{v,j}$ is the j-th delay parameter of node v, and $b_{v,j,l} \in \{0,1\}$, for $l \in \{0,1\}$, refers to the starting or ending of a postsynaptic pulse. Further, P is a set of synapses through which the signal travels from the input node to the output node. For a given input vector, the number of functions can be upper bounded by $2^W \cdot 2^W$, where one factor 2^W is a bound on the number of paths and the other factor limits the number of possible assignments of values to the $b_{v,j,l}$.

Next, we include functions that specify the firing conditions of the nodes. For a node v with synaptic clusters $I_{v,1}, \ldots, I_{v,k}$ and threshold θ_v, they are given by

$$\sum_{j=1}^{k} M_{I_{v,j}}(t) - \theta_v. \tag{7}$$

It was argued in the proof of Theorem 1 that for a given input vector and a fixed class of delay values, there need no more than $2n$ functions be taken into account for each v, where n is the number of synapses of v. For the entire network, this gives rise to at most $2W$ functions of this form.

The functions of (6) are affine in the network parameters, whereas the functions of (7) are sums of rational functions with degree at most p. These classes have solution set components bound $B = 1$ and $B = 2(2p+2)^W$, respectively. The set S has m elements, the affine and rational functions specified above are bounded in number by 2^{2W} and $2W$, respectively. According to Lemma 3, this yields the upper bound

$$\left(\frac{em2^{2W}}{W}\right)^W \cdot 2(2p+2)^W \cdot \left(\frac{2emW}{W}\right)^W$$

for the number of equivalence classes. If S is shattered, this implies that

$$m \;\leq\; 2W^2 + 2W\log m + W\log\left(\frac{4e^2(p+1)}{W}\right) + 1. \tag{8}$$

Similar as in the proof of Theorem 1, the inequality $\ln\alpha \leq \alpha\beta + \ln(1/\beta) - 1$ for $\alpha = m$ and $\beta = (\ln 2)/(2W)$ yields

$$2W\log m \;\leq\; \frac{m}{2} + 2W\log\left(\frac{4W}{e\ln 2}\right)$$

which, used in (8), implies that

$$\begin{aligned} m &\leq 4W^2 + 4W\log\left(\frac{4W}{e\ln 2}\right) + 2W\log\left(\frac{4e^2(p+1)}{W}\right) + 2 \\ &\leq 4W^2 + 2W\log\left(\frac{64W(p+1)}{(\ln 2)^2}\right) + 2. \end{aligned}$$

□

In analogy to Corollary 4, we obtain a tight bound where the lower bound is provided by Theorem 3.1 in [21].

Corollary 6. *Consider the class of networks of spiking neurons with rational synaptic interactions where the degree is bounded by a constant. Let $\mathcal{N}$ be a network from this class having W parameters. Then the pseudo-dimension of $\mathcal{N}$ is $\Theta(W^2)$.*

6 Discussion

Models of neural networks have given rise to some of the most popular techniques in pattern recognition. What makes these networks so attractive is that they need not be programmed explicitly but can learn from examples. A respectable number of learning methods is now available not only for traditional neural network models, such as the multi-layer perceptron, but with increasing importance also for models that are much closer to real nervous systems. Since one of the goals in designing neural models is to reach the performance of living systems, an increasing use of biologically more plausible network models in adaptive pattern recognition is to be expected in the future.

A major problem that one encounters in practical learning situations is the question of how to assess the generalization capability of the system being used. While heuristics such as cross-validation can provide some clues by simulations, theoreticians have found ways to obtain generalization performance guarantees in advance by mathematical means. The so-called sample complexity is a measure that represents the number of training examples that is required by a network to perform well on patterns that have not been part of the training set. One of the fundamental results of this theory of learning states that the sample complexity of a given neural network can be expressed in terms of its pseudo-dimension (see, e.g., [1]). This pseudo-dimension establishes a quantitative relationship between the sample complexity and the number of parameters that characterize the network. Specifically, the sample complexity is proportional to the pseudo-dimension. This is not only valid for a particular learning algorithm but holds for a wide class of adaptation methods.

The pseudo-dimension therefore provides valuable information when designing a learning system. Unfortunately, due to the complexity of the models the pseudo-dimension of neural networks is quite hard to calculate and often good bounds are not known. All the more, this seems to be the case with biologically close network models that comprise sophisticated coding mechanisms, powerful computation units, and many parameters describing the detailed functioning of the cells. Nevertheless, in this article we have shown that tight bounds can be obtained for networks of spiking neurons that use temporal coding and interact nonlinearly.

The two main results in this article state that the pseudo-dimension of feedforward networks with W adjustable parameters and fixed depth is proportional to $W \log W$; it is proportional to W^2 if the depth of the network is unrestricted. Thus, we have obtained asymptotically tight bounds for these architectures. Moreover, we have separated depth-restricted from general architectures in terms of the pseudo-dimension. Significant in these results is the fact that smaller pseudo-dimension bounds can be obtained when taking the depth of the networks into account. Therefore, the almost linear bound $\Theta(W \log W)$ for networks of fixed depth appears to be the result that is more relevant for practical applications of these networks where one or two hidden layers are often sufficient. In particular, in comparison with previous results it follows that the sample sizes required for these networks are not significantly larger than for classical artificial neural network models and, in some cases, even smaller. Thus, with regard to the generalization performance in pattern recognition applications the models studied here can be expected to be at least as good as traditional network models.

7 Conclusion

We have derived upper bounds on the pseudo-dimension of networks of spiking neurons that use temporal coding and have nonlinear synaptic interactions. The bounds are almost linear for networks with depth restrictions, whereas for general networks they are quadratic. The fact that the bounds are rather small is astonishing, since due to the powerful network model with adaptive delays and nonlinearities in interactive computations a much larger variety of network functions is available. The results lead to tight bounds when the network depth or the degree of the rational interactions is considered as a constant. Thus, depth-restricted networks have a significantly smaller pseudo-dimension than arbitrary architectures.

The results presented here show that even when the computational power of networks of spiking neurons increases considerably due to nonlinear interactions, the sample complexity remains close to that of linear interactions. As an issue for further theoretical studies, the question arises whether this is the most general model of spiking neuron networks for which such small and tight bounds can be obtained.

By introducing nonlinear synaptic interactions in spiking neurons that use temporal coding we have enhanced previous models with an essential realistic element. Clearly, there is still a gap between this model and biological neurons, but this might now be a gap that is to be bridged more easily. The results in this article may encourage the use of nonlinear synaptic interactions in implementations and hardware designs of pulse-based computations to obtain more powerful spiking neuron networks.

References

[1] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, Cambridge, 1999.

[2] T. Asai and Y. Amemiya. Frequency- and temporal-domain neural competition in analog integrate-and-fire neurochips. In *Proceedings of the International Joint Conference on Neural Networks IJCNN 2002.* IEEE Computer Society, 2002.

[3] P. L. Bartlett and W. Maass. Vapnik-Chervonenkis dimension of neural nets. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 1188–1192. MIT Press, Cambridge, MA, second edition, 2003.

[4] P. L. Bartlett, V. Maiorov, and R. Meir. Almost linear VC-dimension bounds for piecewise polynomial networks. *Neural Computation*, 10:2159–2173, 1998.

[5] S. Blomfield. Arithmetical operations performed by nerve cells. *Brain Research*, 69:115–124, 1974.

[6] A. Bofill, A. F. Murray, and D. P. Thompson. Circuits for VLSI implementation of temporally-asymmetric Hebbian learning. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 1091–1098, MIT Press, Cambridge, Mass., 2002.

[7] S. M. Bohte, J. N. Kok, and H. L. Poutré. Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing*, 48:17–37, 2002.

[8] M. Carandini and D. J. Heeger. Summation and division by neurons in primary visual cortex. *Science*, 264:1333–1336, 1994.

[9] S. Furukawa and J. C. Middlebrooks. Cortical representation of auditory space: Information-bearing features of spike patterns. *Journal of Neurophysiology*, 87:1749–1762, 2001.

[10] F. Gabbiani, H. G. Krapp, C. Koch, and G. Laurent. Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420:320–324, 2002.

[11] W. Gerstner and W. M. Kistler. *Spiking Neuron Models : Single Neurons, Populations, Plasticity.* Cambridge University Press, Cambridge, 2002.

[12] P. W. Goldberg and M. R. Jerrum. Bounding the Vapnik-Chervonenkis dimension of concept classes parameterized by real numbers. *Machine Learning*, 18:131–148, 1995.

[13] P. Heil and D. R. F. Irvine. First-spike timing of auditory-nerve fibers and comparison with auditory cortex. *Journal of Neurophysiology*, 78:2438–2454, 1997.

[14] J. J. Hopfield, C. D. Brody, and S. Roweis. Computing with action potentials. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 10*, pages 166–172, MIT Press, Cambridge, Mass., 1998.

[15] H. Hüning, H. Glünder, and G. Palm. Synaptic delay learning in pulse-coupled neurons. *Neural Computation*, 10:555–565, 1998.

[16] C. Koch. *Biophysics of Computation.* Oxford University Press, New York, 1999.

[17] C. Koch, T. Poggio, and V. Torre. Nonlinear interactions in a dendritic tree: Localization, timing, and role in information processing. *Proceedings of the National Academy of Sciences USA*, 80:2799–2802, 1983.

[18] M. Kuno and J. T. Miyahara. Non-linear summation of unit synaptic potentials in spinal motoneurones of the cat. *Journal of Physiology*, 201:465–477, 1969.

[19] T. Lehmann and R. Woodburn. Biologically-inspired on-chip learning in pulsed neural networks. *Analog Integrated Circuits and Signal Processing*, 18:117–131, 1999.

[20] W. Maass and C. M. Bishop, editors. *Pulsed Neural Networks.* MIT Press, Cambridge, Mass., 1999.

[21] W. Maass and M. Schmitt. On the complexity of learning for spiking neurons with temporal coding. *Information and Computation*, 153:26–46, 1999.

[22] B. W. Mel. The clusteron: Toward a simple abstraction for a complex neuron. In J. Moody, S. Hanson, and R. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, pages 35–42, Morgan Kaufmann, San Mateo, CA, 1992.

[23] A. F. Murray. Pulse-based computation in VLSI neural networks. In W. Maass and C. M. Bishop, editors, *Pulsed Neural Networks*, chapter 3, pages 87–109. MIT Press, Cambridge, Mass., 1999.

[24] S. Panzeri, R. S. Petersen, S. R. Schultz, M. Lebedev, and M. E. Diamond. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29:769–777, 2001.

[25] P. Reinagel and R. C. Reid. Precise firing events are conserved across neurons. *The Journal of Neuroscience*, 22:6837–6841, 2002.

[26] L. M. Reyneri. Theoretical and implementation aspects of pulse streams: An overview. In *Proceedings of the Seventh International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*, pages 78–89. IEEE Computer Society, 1999.

[27] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code.* MIT Press, Cambridge, Mass., 1999.

[28] B. Ruf and M. Schmitt. Self-organization of spiking neurons using action potential timing. *IEEE Transactions on Neural Networks*, 9:575–578, 1998.

[29] M. Schmitt. Complexity of learning for networks of spiking neurons with nonlinear synaptic interactions. In G. Dorffner, H. Bischof, and K. Hornik, editors, *Proceedings of the International Conference on Artificial Neural Networks ICANN 2001*, volume 2130 of *Lecture Notes in Computer Science*, pages 247–252, Springer, Berlin, 2001.

[30] M. Schmitt. On the complexity of computing and learning with multiplicative neural networks. *Neural Computation*, 14:241–301, 2002.

[31] E. D. Sontag. VC dimension of neural networks. In C. M. Bishop, editor, *Neural Networks and Machine Learning*, pages 69–95. Springer, Berlin, 1998.