Neural Systems as Nonlinear Filters

Wolfgang Maass*
Inst. for Theoretical Computer Science
Technische Universität Graz
Klosterwiesgasse 32/2,
A-8010 Graz, Austria
email: maass@igi.tu-graz.ac.at

Eduardo D. Sontag
Dept. of Mathematics
Rutgers University
New Brunswick, NJ 08903, USA
email: sontag@hilbert.rutgers.edu

Abstract

Experimental data show that biological synapses behave quite differently from the symbolic synapses in all common artificial neural network models. Biological synapses are dynamic, i.e., their "weight" changes on a short time scale by several hundred percent in dependence of the past input to the synapse. In this article we address the question how this inherent synaptic dynamics – which should not be confused with long term "learning" – affects the computational power of a neural network. In particular we analyze computations on temporal and spatio-temporal patterns, and we give a complete mathematical characterization of all filters that can be approximated by feedforward neural networks with dynamic synapses. It turns out that even with just a single hidden layer such networks can approximate a very rich class of nonlinear filters: all filters that can be characterized by Volterra series. This result is robust with regard to various changes in the model for synaptic dynamics. Our characterization result provides for all nonlinear filters that are approximable by Volterra series a new complexity hierarchy which is related to the cost of implementing such filters in neural systems.

1 Introduction

Synapses in common artificial neural network models are static: the value w_i of a synaptic weight is assumed to change only during "learning". In contrast to that, the "weight"

*Wolfgang Maass would like to thank the Sloan Foundation (USA), the Fonds zur Förderung der wissenschaftlichen Forschung (FWF) , Austria, project P12153, and the NeuroCOLT project of the EC for partial support, and the Computational Neurobiology Lab at the Salk-Institute for its hospitality.

$w_i(t)$ of a biological synapse at time t is known to be strongly dependent on the inputs $x_i(t-\tau)$ that this synapse has received from the presynaptic neuron i at previous time steps $t-\tau$. (Varela et al., 1997) have shown that a model of the form

$$w_i(t) = w_i \cdot D(t) \cdot (1 + F(t)) \tag{1}$$

with a constant w_i, a depression term $D(t)$ with values in $(0, 1]$, and a facilitation term $F(t) \geq 0$, can be fitted remarkably well to experimental data for synaptic dynamics. The facilitation term $F(t)$ is usually modeled as a linear filter with exponential decay: If $x_i(t-\tau)$ is the output of the presynaptic neuron (typically modeled by a sum of δ-functions), then the current value of this facilitation term is of the form

$$F(t) \ = \ \rho \int_0^\infty x_i(t-\tau) \cdot e^{-\tau/\gamma} d\tau \tag{2}$$

for certain parameters $\rho, \gamma > 0$ that vary from synapse to synapse. A few other models have been proposed for synaptic dynamics (see e.g. [Dobrunz and Stevens, 1997], (Murthy et al., 1997), (Tsodyks et al., 1998), (Koch, 1999), (Maass and Zador, 1998), (Maass and Zador, 1999)) that are all quite similar. Closely related models had already been proposed and investigated in (Grossberg, 1969), (Grossberg, 1972), (Grossberg, 1984), (Francis et al., 1994). Our analysis in this article is primarily based on the model of (Varela et al., 1997). However we will prove that our results also hold for the somewhat more complex model for synaptic dynamics in a mean-field context of (Tsodyks et al., 1998).

We show in this article that such inherent synaptic dynamics empower neural networks with a remarkable capability for carrying out computations on temporal patterns (i.e., time series) and spatio-temporal patterns. This computational mode, where inputs and outputs consist of temporal patterns or spatio-temporal patterns – rather than static vectors of numbers – appears to provide a more adequate framework for analyzing computations in biological neural systems. Furthermore their capability for processing temporal and spatio-temporal patterns in a very efficient manner may be linked to their superior capabilities for real-time processing of sensory input, hence our analysis may provide new ideas for designing artificial neural systems with similar capabilities.

We consider not just computations of neural systems with a *single* temporal pattern as input, but also characterize their computational power for the case where *several* different temporal patterns $u_1(t), \ldots, u_n(t)$ are presented in parallel as input to the neural system. Hence we also provide a complete characterization of the computational power of feedforward neural systems for the case where salient information is encoded in temporal correlations of firing activity in different pools of neurons (represented by correlations among the corresponding continuous functions $u_1(t), \ldots, u_n(t)$). Therefore various informal suggestions for computational uses of such code can be placed on a rigorous mathematical foundation: It is easy to see that a large variety of computational operations that respond in a particular manner to correlations in temporal input patterns define time invariant filters with fading memory, hence they can in principle be implemented on each of the various kinds of dynamic networks considered in this article.

Previous standard models for computations on temporal patterns in artificial neural networks are time-delay neural networks (where temporal structure is transformed into spatial structure) and recurrent neural networks, both being based on standard "static" synapses (Hertz et al., 1991). Such transformation makes it impossible to let "time represent itself" (Mead, 1989) in subsequent computations, which tends to result in a loss of computational efficiency. The results of this article suggest that feedforward neural networks with simple dynamic synapses provide an attractive alternative.

Various questions regarding artificial neural networks with more general recurrent structure, in which the time-series character of the data plays a central role, were answered, within the framework of computational learning theory, in the papers (Dasgupta and Sontag, 1996) (studied hard-threshold filters with a discrete time scale), (Koiran and Sontag, 1998) (discrete-time recurrent networks), and (Sontag, 1998) (continuous-time recurrent networks). The paper (Sontag, 1997) summarizes some of the approximation capabilities and other properties of these classes of recurrent networks.

In section 2 of this article we introduce the formal notion of a dynamic network, which combines biologically realistic synaptic dynamics according to (Varela et al., 1997) with standard sigmoidal neurons (modeling firing activity in a population of neurons), and we review some basic concepts regarding filters. In section 3 we characterize the computational power of feedforward dynamic networks for computations on temporal patterns (i.e., functions of time), and we show that our result can be extended to the model of (Tsodyks et al., 1998) for synaptic dynamics. The formal proofs of the characterization results in this article rely on standard techniques from mathematical analysis. In section 4 we extend our investigation to computations on spatio-temporal patterns. Section 5 discusses some conclusions.

2 Basic Concepts

In contrast to the static output of gates in feedforward artificial neural networks the output of biological neurons consists of action potentials ("spikes"), i.e., stereotyped events that mark certain points in time. These spikes are transmitted by synapses to other neurons, where they cause changes in the membrane potential that affect the times when these other neurons fire and thereby emit a spike. We will focus in this article on the implications of one type of temporal dynamics provided by the components of such neural computations: the inherent temporal dynamics of synapses.

The empirical data of (Varela et al., 1997) describe the amplitudes of EPSC's (excitatory postsynaptic currents) in a neuron in response to a spike train from a presynaptic neuron. These two neurons are likely to be connected by multiple synapses, and the resulting EPSC amplitude can be understood as a population response of these multiple synapses. Therefore it is justified to employ a deterministic model for synaptic dynamics in spite of the stochastic nature of synaptic transmission at a single release sit (Dobrunz and Stevens, 1997). The EPSC amplitude in response to a spike is modeled in (Varela et al., 1997) by terms of the form $w \cdot (1 + \mathcal{F})$ and $w \cdot \mathcal{D} \cdot (1 + \mathcal{F})$, where $\mathcal{F}$ is a linear filter with impulse response $\rho \cdot e^{-\tau/\gamma}$ modeling facilitation and $\mathcal{D}$ is some nonlinear filter modeling depression at synapses. In some versions of the model considered in (Varela et al., 1997) this filter $\mathcal{D}$ consists of several depression terms. However it only assumes values > 0 and is always time invariant and has fading memory.

We analyze the impact of this synaptic dynamics in the context of common models for computations in populations of neurons where one can ignore the stochastic aspects of computation in individual neurons in favor of the deterministic response of pools of neurons that receive similar input ("population coding" or "space rate coding"), see (Georgopoulos et al., 1986), (Abbott, 1994), (Gerstner, 1999). More precisely, our subsequent neural network model is based on a mean-field analysis of networks of biological neurons, where pools P of neurons serve as computational units, whose time-varying firing activity (measured as the number of neurons in P that fire during a short time interval $[t, t + \Delta]$) is represented by a continuous bounded function $y(t)$. In case that pool P receives inputs from m other pools of neurons $P_1, \ldots, P_m$, we assume that $y(t) = \sigma(\sum_{i=1}^{m} w_i(t)x_i(t) + w_0)$, where $x_i(t)$ represents the time-varying firing activity in pool P_i and $w_i(t)$ represents the time-varying average "weight" of the synapses from neurons in pool P_i to neurons in pool P.[1] In the context of neural computation with population coding considered in this article we have to expand the model of (Varela et al., 1997) to populations of synapses that connect two *pools* of neurons, where presynaptic activity is described not by spike trains but by continuous functions $x_i(t)$ ranging over some bounded interval $[B_0, B_1]$ with $0 < B_0 < B_1$. Therefore we generalize their model for the dynamics of synapses from a nonlinear filter applied to a sequence of δ-functions (i.e., to a spike train) to a corresponding nonlinear filter applied to a continuous input function $x_i(t)$.[2] Thus if $x_i(t)$ is a continuous function describing the firing activity in the ith presynaptic pool P_i of neurons we model the size of the resulting synaptic input to a subsequent pool P of neurons by terms of the form $w_i(t) \cdot x_i(t)$ with $w_i(t) := w_i \cdot (1 + \mathcal{F}x_i(t))$ or $w_i(t) := w_i \cdot \mathcal{D}x_i(t) \cdot (1 + \mathcal{F}x_i(t))$, where the filters $\mathcal{F}$ and $\mathcal{D}$ are defined as in (Varela et al., 1997). The first equation that just models facilitation gives rise to the definition of the class DN of dynamic networks in Definition 2.1, and the second equation, that models the more common co-occurrence of facilitation and depression, gives rise to the definition of the class DN*.

[1] The function $\sigma : \mathbb{R} \to \mathbb{R}$ is some "activation function", for example $\sigma(x) = 1/(1 + e^{-x})$. For the following it suffices to assume that σ is continuous and not a polynomial. In sections 3.2 and 3.3 we have to assume in addition that σ assumes nonnegative values only. We refer to (Maass and Natschläger, 1999) for theoretical arguments and computer simulations that support the use of a sigmoidal activation function in this context.

[2] So far no empirical data are available for the temporal dynamics of a population of synapses (that connects two pools of neurons in a feedforward direction) in dependence of the pool-activity of the presynaptic pool of neurons. It is not completely unproblematic to assume that synaptic dynamics can be modeled on the level of pool-activity in the same way as for spiking neurons, although this is commonly done. The exact formula for the firing activity $y(t)$ in the postsynaptic pool P of neurons requires to multiply for each presynaptic pool P_i of neurons the product of the vector of spike activity of individual neurons $\nu_{i,k}$ in pool P_i with the matrix of current synaptic coupling strengths $w_{i,k,j}(t)$ for neurons ν_j in pool P. The resulting firing activity $y(t)$ of pool P is the average of the current firing activities of neurons ν_j in pool P. In our mean-field model we assume that this average over j can be expressed in terms of products of the average $w_i(t)$ of the synaptic weights $w_{i,k,j}(t)$ over j and k with the average firing activity $x_i(t)$ in the presynaptic pool P_i. In particular this mean-field model ignores that the value of $w_{i,k,j}(t)$ will in general depend on the specific firing history of the specific presynaptic neuron $\nu_{i,k}$.

We refer to (Tsodyks et al., 1998) for a detailed mathematical analysis of this problem. It is shown in that article through computer simulations and theoretical arguments that for the slightly different model for synaptic dynamics considered there the error resulting from generalizing the model from presynaptic individual neurons to presynaptic pools is benign. We will discuss the model from (Tsodyks et al., 1998) in sections 3.2 and 3.3, and we will show in Theorems 3.4 and 3.6 that our results can be extended to their model.

Definition 2.1. *We define the class DN of* dynamic networks *(see Fig. 1) as the class of arbitrary feedforward networks consisting of sigmoidal gates that map input functions* $x_1(t), \ldots, x_m(t)$ *to a function*

$$y(t) = \sigma(\sum_{i=1}^{m} w_i(t)x_i(t) + w_0),$$

with

$$w_i(t) = w_i \cdot (1 + \rho \int_0^\infty x_i(t-\tau)e^{-\tau/\gamma}d\tau)$$

for parameters $w_i \in \mathbb{R}$ *and* $\rho, \gamma > 0$. σ *is some "activation function" from* $\mathbb{R}$ *into* $\mathbb{R}$ *, for example the logistic sigmoid function defined by* $\sigma(x) = 1/(1 + e^{-x})$*. We will assume in the following only that* σ *is continuous and not a polynomial* [3].

The slightly different class DN is defined in the same way, except that* $w_i(t)$ *is of the form*

$$w_i(t) = w_i \cdot \mathcal{D}x_i(t) \cdot (1 + \rho \int_0^\infty x_i(t-\tau)e^{-\tau/\gamma}d\tau),$$

where $\mathcal{D}$ *is some arbitrary* given *time invariant fading memory filter*[4] *with values* $\mathcal{D}x_i(t) \in (0, 1]$.[5]

Thus dynamic networks in DN or DN are simply feedforward neural networks consisting of sigmoidal neurons, where static weights* w_i *are replaced by biologically realistic history-dependent functions* $w_i(t)$*. The input to a dynamic network consists of an arbitrary vector of functions* $u_1(\cdot), \ldots, u_n(\cdot)$*. The output of a dynamic network is defined as a weighted sum*

$$z(t) = \sum_{i=1}^{k} \alpha_i y_i(t) + \alpha_0$$

of the time-varying outputs $y_1(t), \ldots, y_k(t)$ *of certain sigmoidal neurons in the network, where the "weights"* $\alpha_0, \ldots, \alpha_k$ *can be assumed to be static. Thus a dynamic network with* n *inputs maps* n *input functions* $u_1(\cdot), \ldots, u_n(\cdot)$ *onto some output function* $z(\cdot)$. [6]

A somewhat related network model has been investigated in (Back and Tsoi, 1991). They exhibited a learning algorithm for this model, but no characterization of the computational power of such networks was given there.

Temporal patterns are modeled in mathematics as functions of time. Hence networks that operate on temporal patterns map functions of time onto functions of time. We will

[3]According to (Leshno et al., 1993) the subsequent theorems would hold under even weaker conditions on σ.

[4]See the remainder of this section for a review of these notions.

[5]This filter $\mathcal{D}$ models synaptic depression, and can for example be defined as in (Varela et al., 1997). Our subsequent results are independent of the specific definition of $\mathcal{D}$.

[6]In principle one is also interested in a more general type of operators that map vectors $\underline{u}$ of real valued functions on *vectors* $\underline{y}$ of m real valued functions, where m is larger than 1. However in order to answer the questions that are addressed in this paper for the case $m > 1$ it suffices to focus on the case $m = 1$. The reason is that operators which output vectors of m real valued functions can be viewed as vectors of m operators that output one real valued function each. In this way our results for the case $m = 1$ will imply a complete characterization of all operators that can be approximated by a more generalized type of dynamic networks that output m real valued functions instead of just one.

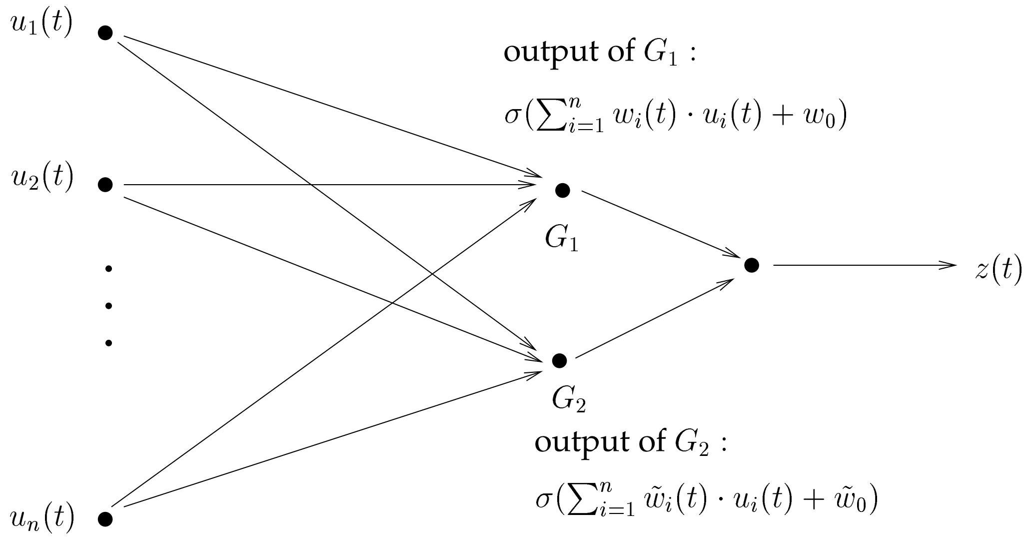

Figure 1: *A dynamic network with one hidden layer consisting of two hidden neurons G_1 and G_2. The synapse from the ith input to G_1 computes the filter $u_i(\cdot) \mapsto w_i(\cdot) \cdot u_i(\cdot)$, the synapse from the ith input to G_2 computes the filter $x_i(\cdot) \mapsto \tilde{w}_i(\cdot) \cdot u_i(\cdot)$. The output of the network is of the form $z(t) = \alpha_1 \cdot \sigma(\sum\limits_{i=1}^{n} w_i(t) \cdot u_i(t) + w_0) + \alpha_2 \cdot \sigma(\sum\limits_{i=1}^{n} \tilde{w}_i(t) \cdot u_i(t) + \tilde{w}_0) + \alpha_0$ with $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$. Thus the network computes a filter that maps the input functions $u_1(\cdot), \ldots, u_n(\cdot)$ onto the output function $z(\cdot)$.*

refer to such maps from functions to functions (or from vectors of functions to functions) as *filters* (in mathematics they are usually referred to as operators). We will reserve the letters $\mathcal{F}, \mathcal{H}, \mathcal{S}$ for filters, and we write $\mathcal{F}\underline{u}$ for the function resulting from an application of the filter $\mathcal{F}$ to a vector $\underline{u}$ of functions. Notice that when we write $\mathcal{F}\underline{u}(t)$ we mean, of course, $(\mathcal{F}\underline{u})(t)$ (that is, the function $\mathcal{F}\underline{u}$ evaluated at time t). We write $C(A, B)$ for the class of all continuous functions $f : A \to B$. We will consider suitable subclasses $U \subseteq C(A, B)$ for $A \subseteq \mathbb{R}^k$ and $B \subseteq \mathbb{R}$, and study filters that map U^n into $\mathbb{R}^{\mathbb{R}}$ (where $\mathbb{R}^{\mathbb{R}}$ is the class of all functions from $\mathbb{R}$ into $\mathbb{R}$), i.e. filters that map n functions $u(\cdot), \ldots, u_n(\cdot)$ onto another function $z(\cdot)$. In this section and in section 3 we will focus on the case $k = 1$, i.e. the case where the input functions $u_1(\cdot), \ldots, u_n(\cdot)$ are functions of a single variable – which we will interpret as time. The case $k > 1$ will be considered in section 4.

A trivial special case of a filter is the shifting filter $\mathcal{S}_{t_0}$ with $\mathcal{S}_{t_0}u(t) = u(t - t_0)$. An arbitrary filter $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$ is called *time invariant* if a shift of the input functions by a constant t_0 just causes a shift of the output function by the same constant t_0, i.e., if for any $t_0 \in \mathbb{R}$ and any $\underline{u} = \langle u_1, \ldots, u_n \rangle \in U^n$ one has that $\mathcal{F}\underline{u}_{t_0}(t) = \mathcal{F}\underline{u}(t - t_0)$ where $\underline{u}_{t_0} = \langle \mathcal{S}_{t_0}u_1, \ldots, \mathcal{S}_{t_0}u_n \rangle$. All filters considered in this article will be time invariant. Note that if U is closed under $\mathcal{S}_{t_0}$ for all $t_0 \in \mathbb{R}$ then a time invariant filter $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$ is fully characterized by the values $\mathcal{F}\underline{u}(0)$ for $\underline{u} \in U^n$.

Another essential property of filters considered in this article is *fading memory*. If a filter $\mathcal{F}$ has fading memory then the value of $\mathcal{F}\underline{v}(0)$ can be approximated arbitrarily closely by the value of $\mathcal{F}\underline{u}(0)$ for functions $\underline{u}$ that approximate the functions $\underline{v}$ for sufficiently long bounded intervals $[-T, 0]$. The formal definition is as follows:

Definition 2.2. *We say that a filter $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$ has* fading memory *if for every $\underline{v} = \langle v_1, \ldots, v_n\rangle \in U^n$ and every $\varepsilon > 0$ there exist $\delta > 0$ and $T > 0$ so that $|\mathcal{F}\underline{v}(0) - \mathcal{F}\underline{u}(0)| < \varepsilon$ for all $\underline{u} = \langle u_1, \ldots, u_n\rangle \in U^n$ with the property that $\|\underline{v}(t) - \underline{u}(t)\| < \delta$ for all $t \in [-T, 0]$.*[7]

Remark 2.3. Interesting examples of linear and nonlinear filters $\mathcal{F} : U \to \mathbb{R}^{\mathbb{R}}$ can be generated with the help of representations of the form

$$\mathcal{F}u(t) = \int_0^\infty \ldots \int_0^\infty u(t-\tau_1) \cdot \ldots \cdot u(t-\tau_k) h(\tau_1, \ldots, \tau_k) d\tau_1 \ldots d\tau_k$$

for measurable and essentially bounded functions $u : \mathbb{R} \to \mathbb{R}$. We will always assume in this article that $h \in L^1$. One refers to such integral as a *Volterra term of order k*. Note that for $k = 1$ it yields the usual representation for a *linear* time invariant filter. The class of filters that can be represented by Volterra series, i.e., by finite or infinite sums of Volterra terms of arbitrary order, has been investigated for quite some time in neurobiology and engineering (see for example (Palm and Poggio, 1977), (Palm, 1978), (Marmarelis and Marmarelis, 1978), (Schetzen, 1980), (Poggio and Reichardt, 1980), (Rugh, 1981), (Rieke et al., 1997)).

It is obvious that any filter $\mathcal{F}$ which can be represented by a sum of finitely many Volterra terms of any order (i.e., by a Volterra polynomial or finite Volterra series) is time invariant and has fading memory. This holds for any class U of uniformly bounded input functions u. According to the subsequent Lemma 2.5 both of these properties are inherited by filters $\mathcal{F}$ that can be approximated by some arbitrary infinite sequence of such filters. This implies that any filter that can be approximated by finite or infinite Volterra series (which converge in the sense used here) is time invariant and has fading memory (over any class U of uniformly bounded functions u). (Boyd and Chua, 1985) have shown that under some additional assumptions about U (for example the assumptions in Theorem 3.1 below) the converse also holds: any time invariant filter $\mathcal{F} : U \to \mathbb{R}^{\mathbb{R}}$ with fading memory can be approximated arbitrarily closely by Volterra polynomials.

Remarks 2.4.

1. It is easy to see that for classes U of functions that are uniformly bounded (i.e., $U \subseteq C(A, B)$ for some bounded set $B \subseteq \mathbb{R}$) our definition of fading memory agrees with that considered in (Boyd and Chua, 1985). All classes U considered in this article are uniformly bounded.

2. It is obvious that any time invariant filter $\mathcal{F}$ that has fading memory is causal, i.e., $\underline{u}(t) = \underline{v}(t)$ for all $t \leq t_0$ implies that $\mathcal{F}\underline{u}(t_0) = \mathcal{F}\underline{v}(t_0)$ for all $t_0 \in \mathbb{R}$.

3. All dynamic synapses considered in this article are modeled as filters that map an input function $x_i(\cdot)$ onto an output function $w_i(\cdot) \cdot x_i(\cdot)$. Furthermore all these filters turn out to be time invariant with fading memory. This has the consequence that all models for dynamic networks considered in this article compute time invariant filters with fading memory.

[7] We will reserve $\|\cdot\|$ for the max-norm on $\mathbb{R}^n$, i.e., for $\underline{x} = \langle x_1, \ldots, x_n\rangle \in \mathbb{R}^n$ we write $\|\underline{x}\|$ for $\max\{|x_i| : i = 1, \ldots, n\}$.

4. If one considers recurrent versions of such networks, then in the absence of noise such networks can theoretically also compute filters without fading memory. Consider for example some filter $\mathcal{F}$ with $\mathcal{F}u(0) = 0$ if $u(t) = 0$ for all $t \leq 0$ and $\mathcal{F}u(0) = 1$ if there exists some $t_0 \leq 0$ so that $u(t_0) \geq 1$. It is obvious that such filter does not have fading memory. But a network where some "self exciting" recurrent subcircuit is turned on (and stays on permanently) whenever the input u reaches a value ≥ 1 for some $t_0 \in \mathbb{R}$ can compute such filter. Alternatively a feedforward network can of course also compute a non-fading-memory filter if any of its components (synapses or neurons) have some permanent memory feature.

5. A special case of time invariant filters $\mathcal{F}$ with fading memory are those defined by $\mathcal{F}\underline{u}(0) = f(\underline{u}(0))$ for arbitrary continuous functions $f : \mathbb{R}^n \to \mathbb{R}$. Therefore the "Universal Approximation Theorem for Filters" that follows from our subsequent Theorem 3.1 contains as a special case the familiar "Universal Approximation Theorem for Functions" from (Hornik et al., 1989).

6. It is obvious that a filter $\mathcal{F}$ on U^n has fading memory if and only if the functional $\tilde{\mathcal{F}} : U^n \to \mathbb{R}$ defined by $\tilde{\mathcal{F}}\underline{u} := \mathcal{F}\underline{u}(0)$ is continuous on U^n with regard to the topology $\mathcal{T}$ generated by the neighborhoods $\{\underline{u} \in U^n : \|\underline{v}(t) - \underline{u}(t)\| < \delta$ for all $t \in [-T, 0]\}$ for arbitrary $\underline{v} \in U^n$ and $\delta, T > 0$.

Lemma 2.5. *Assume that U is closed under $\mathcal{S}_{t_0}$ for all $t_0 \in \mathbb{R}$ and a sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of filters converges to a filter $\mathcal{F}$ in the sense that for every $\varepsilon > 0$ there exists an $n_0 \in \mathbb{N}$ so that $|\mathcal{F}_n\underline{u}(t) - \mathcal{F}\underline{u}(t)| < \varepsilon$ for all $n \geq n_0, \underline{u} \in U^n$, and $t \in \mathbb{R}$. Then the following holds:*

a) If all the filters $\mathcal{F}_n$ are time-invariant then $\mathcal{F}$ is time-invariant.

b) If all the filters $\mathcal{F}_n$ have fading memory then $\mathcal{F}$ has fading memory.

Proof. Claim a) follows immediately from the fact that $\mathcal{F}\underline{u}(t) = \lim_{n\to\infty} \mathcal{F}_n\underline{u}(t)$ for all $\underline{u} \in U^n$ and $t \in \mathbb{R}$. In order to prove b) we can assume that some $\varepsilon > 0$ and some $\underline{v} \in U^n$ have been given. We fix some $n_0 \in \mathbb{N}$ so that $|\mathcal{F}_{n_0}\underline{u}(t) - \mathcal{F}\underline{u}(t)| < \varepsilon/3$ for all $\underline{u} \in U^n, t \in \mathbb{R}$. Since $\mathcal{F}_{n_0}$ has fading memory there exists some $T > 0$ and some $\delta > 0$ so that $|\mathcal{F}_{n_0}\underline{u}(0) - \mathcal{F}_{n_0}\underline{v}(0)| < \varepsilon/3$ for all $\underline{u} \in U^n$ with the property that $\|\underline{u}(t) - \underline{v}(t)\| < \delta$ for all $t \in [-T, 0]$. By our choice of n_0 this implies that $|\mathcal{F}\underline{u}(0) - \mathcal{F}\underline{v}(0)| < \varepsilon$ for all $\underline{u} \in U^n$ with $\|\underline{u}(t) - \underline{v}(t)\| < \delta$ for all $t \in [-T, 0]$. Hence $\mathcal{F}$ has fading memory. ∎

3 Computations on Temporal Patterns

3.1 Characterizing the Computational Power of Neural Networks with Dynamic Synapses

Our subsequent Theorem 3.1 shows that simple filters that only model synaptic facilitation (as considered in the definition of DN) provide the networks already with sufficient dynamics to approximate arbitrary given time invariant filters with fading memory. We show that the simultaneous occurrence of depression (as in DN*) is not needed for that, but it also does not hurt. This appears to be of some interest for the analysis of computations in biological neural systems, since a fairly large variety of different functional roles have already been proposed for synaptic depression: explaining psychological data on conditioning and reinforcement (Grossberg, 1972), boundary formation in vision and visual persistence (Francis et al., 1994), switching between different neural codes (Tsodyks and Markram, 1997), and automatic gain control (Abbott et al., 1997). As a complement of these conjectured roles for synaptic depression our subsequent Theorem 3.1 points to a possible functional role for synaptic facilitation: it empowers even very shallow feed-forward neural systems with the capability to approximate basically any linear or non-linear filter that appears to be of interest in a biological context. Furthermore we show that this possible functional role for facilitation can co-exist with independent other functional roles for synaptic depression: Our result shows that one can first choose the parameters that control synaptic depression to serve some other purpose, and can then still choose the parameters that control synaptic facilitation so that the resulting neural system can approximate any given time invariant filter with fading memory.[8]

Theorem 3.1. *Assume that U is the class of functions from $\mathbb{R}$ into $[B_0, B_1])$ which satisfy $|u(t) - u(s)| \leq B_2 \cdot |t - s|$ for all $t, s \in \mathbb{R}$, where B_0, B_1, B_2 are arbitrary real-valued constants with $0 < B_0 < B_1$ and $0 < B_2$. Let $\mathcal{F}$ be an arbitrary filter that maps vectors $\underline{u} = \langle u_1, \ldots, u_n \rangle \in U^n$ into functions from $\mathbb{R}$ into $\mathbb{R}$.*

Then the following are equivalent:[9]

- *(a) $\mathcal{F}$ can be approximated by dynamic networks $\mathcal{S} \in DN$ (i.e., for any $\varepsilon > 0$ there exists some $\mathcal{S} \in DN$ such that $|\mathcal{F}\underline{u}(t) - \mathcal{S}\underline{u}(t)| < \varepsilon$ for all $\underline{u} \in U^n$ and all $t \in \mathbb{R}$)*
- *(b) $\mathcal{F}$ can be approximated by dynamic networks $\mathcal{S} \in DN$ with just a single layer of sigmoidal neurons*
- *(c) $\mathcal{F}$ is time invariant and has fading memory*
- *(d) $\mathcal{F}$ can be approximated by a sequence of (finite or infinite) Volterra series.*

These equivalences remain valid if DN is replaced by DN*.

The following result follows from the proof of Theorem 3.1. It shows that the class of filters that can be approximated by dynamic networks is very stable with regard to changes in the definition of a dynamic network.

Corollary 3.2. *Dynamic networks with just one layer of dynamic synapses and one subsequent layer of sigmoidal gates can approximate the same class of filters as dynamic networks with an arbitrary finite number of layers of dynamic synapses and sigmoidal gates. Even with a sequence of dynamic networks that have an unboundedly growing number of layers one cannot approximate more filters.*

Furthermore if one restricts the synaptic dynamics in the definition of dynamic networks to the simplest form $w_i(t) = w_i \cdot (1 + \rho \int_0^\infty x_i(t - \tau) e^{-\tau/\gamma} d\tau)$ with some arbitrarily fixed *$\rho > 0$ and time constants γ from some arbitrarily* fixed *interval $[a, b]$ with $0 < a < b$, the resulting class of dynamic networks can still approximate (with just one layer of sigmoidal neurons) any filter that can be approximated by a sequence of arbitrary dynamic networks considered in Definition*

[8]We will show in section 3.3 that alternatively one can employ just depressing synapses for approximating any such filter by a neural system.

[9]The implication "$(c) \Rightarrow (d)$" was already shown in (Boyd and Chua, 1985).

2.1. In the case of DN* one can either choose to fix $\rho > 0$ or one can arbitrarily fix the interval $[a, b]$ for the value of γ.

In addition we will show in section 3.2 that the claim of Theorem 3.1 remains valid if we replace the model from (Varela et al., 1997) for synaptic dynamics (that is employed in the definition of the classes DN and DN* of dynamic networks) by the model from (Tsodyks et al., 1998). Furthermore we will show in section 3.3 that the claim of Theorem 3.1 also holds for networks where synapses exhibit just depression, not facilitation.

Remark 3.3. The proof of Theorem 3.1 shows that its claim as well as the claims of Corollary 3.2 hold under much weaker conditions on the class U. Apart from the requirement that U is closed under translation it suffices to assume that U is some arbitrary class of uniformly bounded and equicontinuous[10] functions that is closed with regard to the topology defined in part 6 of Remarks 2.4, since this assumption is sufficient for the application of the Arzela-Ascoli Theorem (see (Dieudonne, 1969) or (Boyd and Chua, 1985)) in the proof.

Proof of Theorem 3.1: According to Lemma 2.5 any filter that can be approximated by finite or infinite Volterra series is time invariant and has fading memory. This implies "$(d) \Rightarrow (c)$". Furthermore it is shown in (Boyd and Chua, 1985) that for the classes U considered in this article any time invariant filter $\mathcal{F} : U \to \mathbb{R}^{\mathbb{R}}$ with fading memory can be approximated by a sequence of finite Volterra series (i.e., by Volterra polynomials). This argument can be trivially extended to filters $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$ with $n \geq 1$. This implies "$(c) \Rightarrow (d)$". Hence we have shown that $(c) \Leftrightarrow (d)$.

The implication "$(b) \Rightarrow (a)$" is obvious. In order to prove "$(a) \Rightarrow (c)$" we observe that all filters occurring at synapses of a dynamic network (see Definition 2.1) are time invariant and have fading memory. This implies that all filters $\mathcal{S}$ defined by dynamic networks (i.e., all $\mathcal{S} \in DN \cup DN^*$) are time invariant and have fading memory. According to Lemma 2.5 this implies that any filter $\mathcal{F}$ that can be approximated by such networks is time invariant and has fading memory.

For the proof of "$(c) \Rightarrow (b)$" we first consider the case $n = 1$. We assume that $\mathcal{F}$ is some arbitrary given filter that is time invariant and has fading memory. We will first show that $\mathcal{F}$ can be approximated by filters $\mathcal{S} \in DN$. The proof is based on an application of the Stone-Weierstrass Theorem (see for example (Dieudonne, 1969) or (Folland, 1984)) similarly as in (Boyd and Chua, 1985). That article extends earlier arguments by (Sussmann, 1975), (Fliess, 1975), and (Gallman and Narendra, 1976) from a bounded to an unbounded time interval. Furthermore our proof exploits the fact that any continuous function can be uniformly approximated on any compact set by weighted sums of sigmoidal gates (Hornik et al., 1989), (Sandberg, 1991), (Leshno et al., 1993). We will apply the Stone-Weierstrass Theorem to functionals from $U_- := \{u|_{(-\infty,0]} : u \in U\}$ into $\mathbb{R}$. For that purpose we have to show that the filters $\mathcal{H}$ of the form

$$\mathcal{H}u(t) = u(t) \cdot (1 + \rho \int_0^\infty u(t-\tau) e^{-\tau/\gamma} d\tau)$$

separate points in U_-, i.e., for any $u, v \in U_-$ with $u \neq v$ there exists a filter $\mathcal{H}$ of this form such that $\mathcal{H}u(0) \neq \mathcal{H}v(0)$. Thus we consider some arbitrary given $u, v \in U$ with

[10]U is equicontinuous if for any $\varepsilon > 0$ there exists a $\delta > 0$ so that $|t - s| \leq \delta$ implies $|u(t) - u(s)| \leq \varepsilon$ for all $t, s \in \mathbb{R}$ and all $u \in U$.

$u(t) \neq v(t)$ for some $t \leq 0$. Then the function $u(0) \cdot u(-\tau) - v(0) \cdot v(-\tau)$ assumes a value $\neq 0$ for some $\tau \geq 0$. This implies that

$$q(l) \;=\; \int_0^\infty (u(0) \cdot u(-\tau) - v(0) \cdot v(-\tau)) e^{-\tau/l} d\tau$$

does not assume a constant value for all arguments l in $[a, b]$. This follows because, if $q(l) = c$ for all such l, then q, being an analytic function of $l \in \mathbb{C}$ with real part > 0, would equal c, for all real $l > 0$. Since the limit of $q(l)$ as $l \to \infty$ is zero, this means $c = 0$. However the Laplace transform is one-to-one (this is a standard fact; one way to prove it is using that the Laplace transform of a bounded measurable function w, evaluated at any point of the form $s = 1 + i\omega$, coincides, as a function of ω, with the Fourier transform of $w(t)e^{-t}$, and the Fourier transform is one-to-one on integrable functions, cf. for instance (Hewitt and Stromberg, 1965), Corollary 21.47). Hence, $q(l)$ does not assume a constant value for all arguments l in $[a, b]$. Since q is analytic, it therefore assumes, in any interval $[a, b]$ with $0 < a < b$, infinitely many different values. This implies that for any fixed $\rho > 0$

$$u(0) + \rho \cdot \int_0^\infty u(0) \cdot u(-\tau) e^{-\tau/\gamma} d\tau \quad \neq \quad v(0) + \rho \cdot \int_0^\infty v(0) \cdot v(-\tau) e^{-\tau/\gamma} d\tau$$

for some $\gamma \in [a, b]$. Therefore we have $\mathcal{H}_\gamma u(0) \neq \mathcal{H}_\gamma v(0)$ for the filter $\mathcal{H}_\gamma$ defined by

$$\mathcal{H}_\gamma u(t) = u(t) \cdot (1 + \rho \cdot \int_0^\infty u(t - \tau) e^{-\tau/\gamma} d\tau) \; .$$

In order to apply the Stone-Weierstrass Theorem we also need to show that U_- is a compact metric space with regard to the topology $\mathcal{T}$ defined in part 6 of Remarks 2.4. Obviously this topology $\mathcal{T}$ coincides with the topology generated on U_- by the metric

$$d(u, v) := \sup_{t \leq 0} \frac{|u(t) - v(t)|}{1 + |t|}$$

(since all functions in U are assumed to be uniformly bounded). The compactness of U_- with regard to this metric follows by a routine argument, applying the Arzela-Ascoli Theorem successively to the sequence of restrictions $U|_{[-T,0]} := \{u|_{[-T,0]} : u \in U\}$ for $T \in \mathbb{N}$ and by diagonalizing over converging subsequences for these restrictions (see, for instance, Lemma 1 in (Boyd and Chua, 1985)).

The Stone-Weierstrass Theorem implies then that there exists for every given $\varepsilon > 0$ some $m \in \mathbb{N}$, filters $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$ as specified above, and a polynomial p such that

$$|\mathcal{F}u(0) - p(\mathcal{H}_{\gamma_1} u(0), \ldots, \mathcal{H}_{\gamma_m} u(0))| < \frac{\varepsilon}{2}$$

for all $u \in U_-$. Since the functionals $\tilde{\mathcal{H}}_{\gamma_i} : U_- \to \mathbb{R}$ defined by $\tilde{\mathcal{H}}_{\gamma_i} u := \mathcal{H}_{\gamma_i} u(0)$ are continuous over the compact space U_-, the values $\mathcal{H}_{\gamma_i} u(0)$ for $i \in \{1, \ldots, n\}$ and $u \in U_-$ are contained in some bounded interval $[-b, b]$. Furthermore according to (Hornik et al., 1989), (Leshno et al., 1993) there exist sigmoidal gates $G_1, \ldots, G_k$ and parameters $\alpha_0, \ldots, \alpha_k \in \mathbb{R}$ such that

$$|p(\underline{x}) - (\sum_{j=1}^{k} \alpha_j G_j(\underline{x}) + \alpha_0)| < \frac{\varepsilon}{2}$$

for all $\underline{x} \in [-b, b]^m$.[11] Note that $G_j(\mathcal{H}_{\gamma_1} u(0), \ldots, \mathcal{H}_{\gamma_m} u(0))$ is of the form $\sigma(\sum_{i=1}^m w_i(0) u(0) + w_0)$ with $w_0 \in \mathbb{R}$ and $w_i(t)$ as in Definition 2.1 (with $x_i(\cdot)$ replaced by $u(\cdot)$). Hence the previously constructed $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$ together with this layer of k sigmoidal gates $G_1, \ldots, G_k$ define a dynamic network $S \in$ DN. We then have $|\mathcal{F}u(0) - \mathcal{S}u(0)| < \varepsilon$ for all $u \in U_-$. Because of the time invariance of $\mathcal{F}$ and $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$ this implies that $|\mathcal{F}u(t) - \mathcal{S}u(t)| < \varepsilon$ for all $u \in U$ and all $t \in \mathbb{R}$. This completes the proof of "$(c) \Rightarrow (b)$" for the case of dynamic networks that define filters $\mathcal{S} \in DN$ and $n = 1$.

In order to show that for $u, v \in U_-$ with $u \neq v$ we have $\mathcal{H}u(0) \neq \mathcal{H}v(0)$ also for some filter $\mathcal{H}$ that reflects synaptic dynamics with some arbitrary *given* depression filter $\mathcal{D}$ as in the definition of DN* we consider two cases for the filter $\mathcal{H}_\gamma$ with $\mathcal{H}_\gamma u(0) \neq \mathcal{H}_\gamma v(0)$ that we have already constructed.

Case 1: $u(0) \cdot \mathcal{D}u(0) = v(0) \cdot \mathcal{D}v(0)$

Then the function $u(0) \cdot \mathcal{D}u(0) \cdot u(-\tau) - v(0) \cdot \mathcal{D}v(0) \cdot v(-\tau)$ assumes a value $\neq 0$ for some $\tau \geq 0$. Hence we can apply the same argument as before to the function

$$\tilde{q}(l) = \int_0^\infty (u(0) \cdot \mathcal{D}u(0) \cdot u(-\tau) - u(0) \cdot \mathcal{D}v(0) \cdot v(-\tau)) e^{-\tau/l} d\tau$$

to show that this function assumes infinitely many different values for $l \in [a, b]$, for any given interval $[a, b]$ with $0 < a < b$. This implies that there exists for every given $\rho > 0$ some $\gamma \in [a, b]$ so that $u(0) \cdot \mathcal{D}u(0) \cdot (1 + \rho \cdot \int_0^\infty u(-\tau) e^{-\tau/\gamma} d\tau) \neq v(0) \cdot \mathcal{D}v(0) \cdot (1 + \rho \cdot \int_0^\infty v(-\tau) e^{-\tau/\gamma} d\tau)$.

Case 2: $u(0) \cdot \mathcal{D}u(0) \neq v(0) \cdot \mathcal{D}v(0)$

Then the claim follows since $\rho \cdot \int_0^\infty u(-\tau) e^{-\tau/\gamma} d\tau - \rho \cdot \int_0^\infty v(-\tau) e^{-\tau/\gamma} d\tau$ converges to 0 if $\rho \to 0$ or $\gamma \to 0$.

The rest of the argument is exactly as in the preceding argument for filters $\mathcal{S} \in DN$. This completes the proof of "$(c) \Rightarrow (b)$" also for the case of dynamic networks that define filters $\mathcal{S} \in DN^*$ and $n = 1$.

In the claim of the theorem we had considered a slightly more general class of filters $\mathcal{F}$ that are defined over U^n for some given $n \geq 1$. In order to extend the preceding proof of "$(c) \Rightarrow (b)$" to the more general input space for $n \geq 1$ one just has to note that U^n is a compact metric space with regard to the product topology generated by the topology $\mathcal{T}$ over U as in part 6 of Remarks 2.4 , and that our preceding arguments imply that filters over U^n of the form $\langle u_1, \ldots, u_n \rangle \to \mathcal{H}u_i$ with $i \in \{1, \ldots, n\}$ (and $\mathcal{H}$ modeling synaptic dynamics according to Definition 2.1) separate points in U_-^n. ∎

3.2 Extension of the Result to the Model for Synaptic Dynamics by Tsodyks, Pawelzik, and Markram

In (Tsodyks et al., 1998) a slightly different temporal dynamics for depression and facilitation in populations of synapses has been proposed. In contrast to the model from

[11] These approximation results were previously applied in this context by (Sandberg, 1991).

(Varela et al., 1997) that underlies our Definition 2.1, this model has been explicitly formulated for a mean-field analysis, where the input to a population of synapses consists of a continuous function $x_i(t)$ that models firing activity in a presynaptic pool P_i of neurons, rather than a spike train from a single presynaptic neuron. We will show in this section that our characterization result from the preceding section also holds for this model for synaptic dynamics.

The first difference to the synapse model from (Varela et al., 1997) is a use-dependent discount factor $e^{-\rho\int_{-\tau}^{0} x_i(\tau')d\tau'} \cdot e^{-\tau/\gamma}$ instead of just $e^{-\tau/\gamma}$ in the model for facilitation, that reduces the facilitating impact of preceding large input $x_i(-\tau)$ on the value of the synaptic weight at time 0. In other words: facilitation is no longer modeled by a linear filter, instead one assumes that facilitation has less impact on a synapse that has already been facilitated by preceding inputs.

For a precise definition of the resulting variation DN$^+$ of our definition of dynamic networks from Definition 2.1 we replace $w_i(t) = w_i \cdot (1 + \rho\int_0^\infty x_i(t-\tau)\ e^{-\tau/\gamma}d\tau)$ by $w_i(t) = w_i \cdot \hat{w}_i(t)$, where

$$\hat{w}_i(t) = \rho \cdot \int_0^\infty x_i(t-\tau) \cdot e^{-\rho\int_{t-\tau}^{t} x_i(\tau')d\tau'} \cdot e^{-\tau/\gamma}d\tau\ . \tag{3}$$

This is the model for facilitation proposed in equation (3.5) of (Tsodyks et al., 1998) for a mean-field setting, where $x_i(t-\tau)$ models firing activity at time $t-\tau$ in a presynaptic pool P_i of neurons ($\hat{w}_i(t)$ is denoted by U_{SE}^1, ρ is denoted by U_{SE}, γ is denoted by τ_{facil}, and w_i is denoted by A_{SE} in (Tsodyks et al., 1998)).

We will show in the subsequent result that for any given value of the parameter ρ (which models the normal utilization of synaptic resources caused by input to a "rested" synapse) and for any given interval $[a, b]$ one can choose the values $w_i \in \mathbb{R}$ and time constants γ from $[a, b]$ so that a network consisting of facilitating synapses in combination with one layer of sigmoidal neurons can approximate any time invariant filter with fading memory.

(Tsodyks et al., 1998) also propose a model for populations of synapses that exhibit both depression and facilitation (one substitutes equation (3.5) for U_{SE} in (3.3) of (Tsodyks et al., 1998)). A new feature of this model is that one can no longer express the current synaptic weight $w_i(t)$ as a product of the outputs of two separate filters, one for depression and one for facilitation. Rather, the output of the filter for facilitation (see our equation (3)) enters the computation of the current output of the filter for depression. This is biologically plausible, since a facilitated synapse spends its resources more quickly – and hence is subject to stronger depression. In our notation the model from (Tsodyks et al., 1998) for depression and facilitation in a mean-field setting (equations (3.3) and (3.5) in (Tsodyks et al., 1998)) yields the following formula for the value $w_i(t)$ of the current weight of a population of synapses (with $\hat{w}_i(t)$ defined according to equation (3)):

$$w_i(t) := w_i \cdot \hat{w}_i(t) \cdot \int_0^\infty e^{-\tau/\tau_{rec}} \cdot e^{-\int_{t-\tau}^{t} \hat{w}_i(\tau')x_i(\tau')d\tau'}d\tau\ . \tag{4}$$

This formula involves another parameter τ_{rec}: the time constant for the *recovery* from utilizing synaptic resources. We will write DN^{++} for the class of feedforward networks consisting of sigmoidal neurons with dynamic weights $w_i(t)$ according to equation (4).

In order to make sure that the integrals in equation (3) and (4) assume a finite value for bounded synaptic inputs $x_i(\cdot)$, one has to make sure that not only the network inputs, but also the outputs of sigmoidal units in networks from the classes DN$^+$ and DN^{++} are always nonnegative. For that purpose we assume in this section and the next section that the sigmoidal activation function σ assumes nonnegative values only. Note that this is no real restriction since the output of a sigmoidal unit models the current firing activity in a pool of neurons.

Note that any filter that maps $x_i(\cdot)$ onto $w_i(\cdot) \cdot x_i(\cdot)$ with $w_i(\cdot)$ defined according to equation (3) or (4) is time invariant and has fading memory. Hence every network in DN$^+$ and DN^{++} computes a time invariant filter with fading memory.

Theorem 3.4. *Assume that U is the class of functions from $\mathbb{R}$ into $[B_0, B_1]$) which satisfy $|u(t) - u(s)| \leq B_2 \cdot |t - s|$ for all $t, s \in \mathbb{R}$, where B_0, B_1, B_2 are arbitrary real-valued constants with $0 < B_0 < B_1$ and $0 < B_2$. Let $\mathcal{F}$ be an arbitrary filter that maps vectors $\underline{u} = \langle u_1, \ldots, u_n \rangle \in U^n$ into functions from $\mathbb{R}$ into $\mathbb{R}$.*

Then the following are equivalent:

- *(a) $\mathcal{F}$ can be approximated by dynamic networks $\mathcal{S} \in DN^+$ (i.e., for any $\varepsilon > 0$ there exists some $\mathcal{S} \in DN^+$ such that $|\mathcal{F}\underline{u}(t) - \mathcal{S}\underline{u}(t)| < \varepsilon$ for all $\underline{u} \in U^n$ and all $t \in \mathbb{R}$)*
- *(b) $\mathcal{F}$ can be approximated by dynamic networks $\mathcal{S} \in DN^+$ with just a single layer of sigmoidal neurons*
- *(c) $\mathcal{F}$ is time invariant and has fading memory*
- *(d) $\mathcal{F}$ can be approximated by a sequence of (finite or infinite) Volterra series.*

These equivalences remain valid if DN$^+$ is replaced by DN^{++}.

It will be obvious from the proof of Theorem 3.4 that in principle quite small ranges suffice for the "free" parameters γ and τ_{rec} that control the synaptic dynamics according to (3) and (4):

Corollary 3.5. *In order to approximate an arbitrary given time invariant fading memory filter $\mathcal{F}$ by dynamic networks S from DN$^+$, one can choose for any given $\rho > 0$ the parameters γ of the synapses in S (defined according to (3)) from some arbitrarily given interval $[a, b]$. In order to approximate $\mathcal{F}$ by networks S from DN^{++} one can choose for any given $\rho > 0$ the parameters γ from some arbitrarily given interval $[a, b]$ and the parameters τ_{rec} according to equation (4) from some arbitrarily given interval $[a', b']$.*

Proof of Theorem 3.4. It suffices to describe how the proof of "$(c) \Rightarrow (b)$" from Theorem 3.1 has to be changed. For the case of networks from the class DN$^+$ we have to show that the filters $\mathcal{H}_\gamma^+$ of the form

$$\mathcal{H}_\gamma^+ u(t) = u(t) \cdot \rho \int_0^\infty u(t-\tau) \cdot e^{-\rho \int_{t-\tau}^t u(\tau')d\tau'} \cdot e^{-\tau/\gamma} d\tau$$

separate points in U_-. We show that for any given $\rho > 0, a, b \in \mathbb{R}$ with $0 < a < b$, and any $u, v \in U_-$ with $u \neq v$ there exists some $\gamma \in [a, b]$ such that $\mathcal{H}_\gamma^+ u(0) \neq \mathcal{H}_\gamma^+ v(0)$. Thus we consider some arbitrary given $u, v \in U$ with $u(t) \neq v(t)$ for some $t \leq 0$. According to our argument in the proof of Theorem 3.1 it suffices for that to show that

$$u(0) \cdot u(-\tau) \cdot e^{-\rho \int_{-\tau}^{0} u(\tau')d\tau'} \quad \neq \quad v(0) \cdot v(-\tau) \cdot e^{-\rho \int_{-\tau}^{0} v(\tau')d\tau'} \text{ for some } \tau \geq 0 \ , \qquad (5)$$

because this implies that the function q^+ defined by

$$q^+(\ell) := \int_0^\infty (u(0) \cdot u(-\tau) \cdot e^{-\rho \int_{-\tau}^{0} u(\tau')d\tau'} - v(0) \cdot v(-\tau) \cdot e^{-\rho \int_{-\tau}^{0} v(\tau')d\tau'}) e^{-\tau/\ell} d\tau$$

assumes infinitely many different values for $\ell \in [a, b]$.

Assume for a contradiction that (5) does not hold. This implies that $u(0) = v(0)$. Consider some $t_0 < 0$ with $u(t_0) \neq v(t_0)$. We assume without loss of generality that $u(t_0) > v(t_0)$. Set

$$t_0^+ := \inf\{t > t_0 : u(t) \leq v(t)\}$$
$$t_0^- := \sup\{t < t_0 : u(t) \leq v(t)\} \ .$$

We have $t_0^+ \leq 0$ since $u(0) = v(0)$.

Case 1: $t_0^- > -\infty$ (i.e., $\exists\, t < t_0 : u(t) \leq v(t)$)

Then $t_0^- < t_0 < t_0^+, u(t_0^-) = v(t_0^-), u(t_0^+) = v(t_0^+)$ and $u(t) > v(t)$ for all $t \in (t_0^-, t_0^+)$. According to our assumption this implies that $\int_{t_0^+}^0 u(t)dt = \int_{t_0^+}^0 v(t)dt$ and $\int_{t_0^-}^0 u(t)dt = \int_{t_0^-}^0 v(t)dt$, although $\int_{t_0^-}^{t_0^+} u(t)dt > \int_{t_0^-}^{t_0^+} v(t)dt$. This yields a contradiction.

Case 2: $u(t) > v(t)$ for all $t < t_0$

Our assumptions imply then that $\int_{t_0^+}^0 u(t)dt = \int_{t_0^+}^0 v(t)dt$ and $u(t) > v(t)$ for all $t < t_0^+$. Therefore there exists some $\varepsilon > 0$ such that $e^{\rho \int_t^0 (u(\tau') - v(\tau'))d\tau'} \geq 1 + \varepsilon$ for all $t \leq t_0$. Hence we can conclude from our assumption that $\frac{u(t)}{v(t)} \geq 1 + \varepsilon$ for all $t \leq t_0$. This implies that $e^{\rho \int_t^0 (u(\tau') - v(\tau'))d\tau'} \to \infty$ for $t \to -\infty$, hence $\frac{u(t)}{v(t)} \to \infty$ for $t \to -\infty$. This provides a contradiction to our definition of the class U of functions to which u and v belong, since all functions in U have values in $[B_0, B_1]$ for $0 < B_0 < B_1$.

This completes our proof of the direction "$(c) \Rightarrow (b)$". The remainder of the proof for the case of dynamic networks from the class DN$^+$ is the same as for Theorem 3.1.

In order to prove "$(c) \Rightarrow (b)$ for networks from the class DN^{++}, we have to show that the filters that map an input function $x_i(\cdot)$ onto the output function $w_i(\cdot) \cdot x_i(\cdot)$ with $w_i(t)$ defined according to equation (4), separate points in U^-. Thus we fix some $u, v \in U$ with $u(t) \neq v(t)$ for some $t \leq 0$. According to the preceding proof for DN$^+$ there exists some $\gamma \in [a, b]$ such that $\mathcal{H}_\gamma^+ u(0) \neq \mathcal{H}_\gamma^+ v(0)$. We want to show for this γ that there exists for any given a', b' with $0 < a' < b'$ some $\tau_{rec} \in [a', b']$ so that the resulting filter defined by the synapse according to equation (4) can separate u and v. More precisely, we show for the filter $\mathcal{G}_\gamma$ that is defined in analogy to equation (3) by

$$\mathcal{G}_\gamma u(t) := \rho \cdot \int_0^\infty u(t - \tau) \cdot e^{-\rho \int_{t-\tau}^{\tau} u(\tau')d\tau'} \cdot e^{-\tau/\gamma} d\tau$$

(thus $\mathcal{H}_\gamma^+ u(t) = u(t) \cdot \mathcal{G}_\gamma u(t)$) that

$$u(0) \cdot \mathcal{G}_\gamma u(0) \cdot \int_0^\infty e^{-\tau/\tau_{rec}} \cdot e^{-\int_{-\tau}^0 \mathcal{G}_\gamma u(\tau')\cdot u(\tau')d\tau'} d\tau \neq \tag{6}$$
$$v(0) \cdot \mathcal{G}_\gamma v(0) \cdot \int_0^\infty e^{-\tau/\tau_{rec}} \cdot e^{-\int_{-\tau}^0 \mathcal{G}_\gamma v(\tau')\cdot v(\tau')d\tau'} d\tau \; .$$

It is obvious that the function $h : \mathbb{R} \to \mathbb{R}$ defined by

$$h(\tau) := u(0) \cdot \mathcal{G}_\gamma u(0) \cdot e^{\int_{-\tau}^0 \mathcal{G}_\gamma u(\tau')\cdot u(\tau')d\tau'} - v(0) \cdot \mathcal{G}_\gamma v(0) \cdot e^{-\int_{-\tau}^0 \mathcal{G}_\gamma v(\tau')\cdot v(\tau')d\tau'}$$

assumes a value $\neq 0$ for some $\tau \geq 0$, since $h(0) \neq 0$ by our choice of γ. Hence by the argument via the Laplace transform from the proof of Theorem 3.1 there exists some $\tau_{rec} \in [a', b']$ (for any given $a', b' \in \mathbb{R}$ with $0 < a' < b'$) so that

$$\int_0^\infty e^{-\tau/\tau_{rec}} \cdot h(\tau) d\tau \neq 0 \; ,$$

which is equivalent to the desired inequality (6).

Thus we have shown that the filters defined by the temporal dynamics of synapses in dynamic networks from the class DN^{++} separate points in U^-. The rest of the proof is the same as for Theorem 3.1. ∎

3.3 Universal Approximation of Filters with Depressing Synapses only

We will show in this section that the computational power of feedforward neural networks with dynamic synapses remains the same if the synapses just exhibit depression, not facilitation. This holds provided that the time constants τ_{rec} for their recovery from depression can be chosen individually from some interval $[a, b]$ (this holds for *any* values of a, b with $0 < a < b$). This result is of interest since according to (Tsodyks et al., 1998) all synapses between pyramidal neurons just exhibit depression, and not facilitation. We will employ the model from (Tsodyks et al., 1998) for synaptic depression in a mean-field setting, which is specified in equation (3.3) of (Tsodyks et al., 1998).

We write DN$^-$ for the class of feedforward neural networks consisting of sigmoidal neurons (whose activation function σ assumes nonnegative values only) with weights $w_i(t)$ evolving according to

$$w_i(t) = w_i \cdot U_{SE} \cdot \int_0^\infty e^{-\tau/\tau_{rec}} \cdot e^{-\int_{t-\tau}^t U_{SE}\cdot x_i(\tau')d\tau'} d\tau \tag{7}$$

in dependence of the presynaptic pool activity $x_i(\tau)$, where $U_{SE} > 0$ is some given constant. Note that $w_i(t) \cdot x_i(t)$ agrees with the term $A_{SE} \cdot \langle y(t) \rangle$ with $\langle y(t) \rangle$ defined by equation (3.3) in (Tsodyks et al., 1998), which models the average value of the postsynaptic current caused in pool P by the firing activity $x_i(t)$ in the pool P_i in the case of depressing synapses between pools P_i and P (the parameter A_{SE} is denoted by w_i in our notation).

Theorem 3.6. *Assume that U is the class of functions from $\mathbb{R}$ into $[B_0, B_1]$) which satisfy $|u(t) - u(s)| \leq B_2 \cdot |t - s|$ for all $t, s \in \mathbb{R}$, where B_0, B_1, B_2 are arbitrary real-valued constants with $0 < B_0 < B_1$ and $0 < B_2$. Let $\mathcal{F}$ be an arbitrary filter that maps vectors $\underline{u} = \langle u_1, \ldots, u_n \rangle \in U^n$ into functions from $\mathbb{R}$ into $\mathbb{R}$.*

Then the following are equivalent:

(a) $\mathcal{F}$ can be approximated by dynamic networks $\mathcal{S} \in DN^-$ (i.e., for any $\varepsilon > 0$ there exists some $\mathcal{S} \in DN^-$ such that $|\mathcal{F}\underline{u}(t) - \mathcal{S}\underline{u}(t)| < \varepsilon$ for all $\underline{u} \in U^n$ and all $t \in \mathbb{R}$)

(b) $\mathcal{F}$ can be approximated by dynamic networks $\mathcal{S} \in DN^-$ with just a single layer of sigmoidal neurons

(c) $\mathcal{F}$ is time invariant and has fading memory

(d) $\mathcal{F}$ can be approximated by a sequence of (finite or infinite) Volterra series.

Proof. It is obvious that all filters defined by dynamic networks from the class DN$^-$ are time invariant and have fading memory. Hence it suffices to show how the proof of "$(c) \Rightarrow (b)$" has to be changed in comparison with the proof of Theorem 3.1. Assume that parameters $a, b, U_{SE} \in \mathbb{R}$ with $0 < a < b$ and $0 < U_{SE}$ have been fixed in some arbitrary manner. We have to show that for any two functions $u, v \in U$ with $u(t_0) > v(t_0)$ for some $t_0 \leq 0$ there exists some $\tau_{rec} \in [a, b]$ so that the filter that models synaptic dynamics according to (7) differs at time 0 for the two input functions u, v (instead of x_i), i.e., we have to show that

$$u(0) \cdot \int_0^\infty e^{-\tau/\tau_{rec}} \cdot e^{-\int_{-\tau}^0 U_{SE} \cdot u(\tau')d\tau'} d\tau \neq v(0) \cdot \int_0^\infty e^{-\tau/\tau_{rec}} \cdot e^{-\int_{-\tau}^0 U_{SE} \cdot v(\tau')d\tau'} d\tau \ .$$

According to our argument with the Laplace-transform in the proof of Theorem 3.1 it suffices for that to show that $h(\tau) \neq 0$ for some $\tau \geq 0$, where $h : \mathbb{R} \rightarrow \mathbb{R}$ is the function defined by

$$h(\tau) := u(0) \cdot e^{-\int_{-\tau}^0 U_{SE} \cdot u(\tau')d\tau'} - v(0) \cdot e^{-\int_{-\tau}^0 U_{SE} \cdot v(\tau')d\tau'} \ .$$

If $u(0) \neq v(0)$ this is obvious, since then $h(0) \neq 0$. Hence we assume that $u(0) = v(0)$. Furthermore we assume for a contradiction that $h(\tau) = 0$ for all $\tau \geq 0$. Set

$$t_0^+ := \ \inf \{t > t_0 : u(t) \leq v(t)\} \ .$$

Then we have $t_0 < t_0^+ \leq 0$, $\int_{t_0^+}^0 u(\tau')d\tau' = \int_{t_0^+}^0 v(\tau')d\tau'$, and $\int_{t_0}^0 u(\tau')d\tau' = \int_{t_0}^0 v(\tau')d\tau'$. This yields a contradiction to the fact that $u(\tau') > v(\tau')$ for all $\tau' \in [t_0, t_0^+]$, and hence $\int_{t_0}^{t_0^+} u(\tau')d\tau' > \int_{t_0}^{t_0^+} v(\tau')d\tau'$. ∎

3.4 Focusing on Excitatory Synapses

In the preceding dynamic network models we had assumed that the constant factors w_i could be chosen to be positive or negative, thus yielding excitatory or inhibitory synapses in a biological interpretation. This formal symmetry between excitatory and inhibitory synapses is not adequate for most biological neural systems, for example the cortex of primates, where just 15% of the synapses are inhibitory. We would like to point out in this section that according to (Maass, 1999a) one can replace the dynamic networks considered in the preceding sections by an alternative type of network where just the dynamics of excitatory synapses matters – which can be just depressing, just facilitating, or depressing and facilitating, like in the preceding sections.

The key observation is that instead of approximating the given polynomial p by a weighted sum of sigmoidal neurons in the proof of Theorem 3.1 (and analogously in the subsequent Theorems), one can approximate p by a single soft winner-take-all module applied to several weighted sums of the filters $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$ with *non-negative* weights w_i only.[12] The resulting network structure corresponds to a biological neural system where the filters $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$ are realized exclusively by *excitatory* dynamic synapses, and the role of inhibitory synapses is restricted to the realization of the subsequent soft winner-take-all module. We refer to (Maass, 1999a) for details of this alternative style of network construction.

3.5 Allowing the Input Functions to Vanish

We had assumed in Theorem 3.1 that all input functions u_i satisfy $u_i(t) \geq B_0$ for all $t \in \mathbb{R}$, where $B_0 > 0$ is some arbitrary constant. This assumption is usually met in the sketched application to biological neural systems, because the minimum firing rate of neurons is larger than 0 (typically in the range of 5 Hz). Alternatively one can assume that all input functions u_i are superimposed with some positive constant input (that could be interpreted as background firing activity).

The following result shows that from a mathematical point of view the assumption $B_0 > 0$ is *necessary* at least in the case of single-layer networks, since in the case $B_0 = 0$ a strictly smaller class of filters is approximated by dynamic networks with a single layer of sigmoidal neurons. Theorem 3.7 gives a precise characterization of this smaller class of filters.

Theorem 3.7. *Assume that U is the class of functions u in $C(\mathbb{R}, [0, B_1])$ which satisfy $|u(t) - u(s)| \leq B_2 \cdot |t - s|$ for all $t, s \in \mathbb{R}$, where B_1 and B_2 are arbitrary real-valued constants with $0 < B_1$ and $0 < B_2$. Let $\mathcal{F}$ be an arbitrary filter that maps functions from U into $\mathbb{R}^{\mathbb{R}}$. We consider here only the version of dynamic networks giving rise to filters in DN.*

Then $\mathcal{F}$ can be approximated by dynamic networks with a single layer of sigmoidal gates if and only if *$\mathcal{F}$ is time invariant, has fading memory, and there exists a constant $c_{\mathcal{F}}$ such that $\mathcal{F}u(t) = c_{\mathcal{F}}$ for all $u \in U$ and $t \in \mathbb{R}$ with $u(t) = 0$.*

Proof of Theorem 3.7: Notice that the form of the filters defining the class DN implies that, when $u(t) = 0$, all filters output the value 0 at the given time t, and hence the sigmoidal gate outputs the value $G(t) = \sigma(w_0)$, irrespective of the values of u at other times. It is easy to see from here that all filters approximated by such networks must also have the same property. The converse implication is established in almost exactly the same manner as in the proof of Theorem 3.1. The only difference is as follows.

It could be the case that $u(0) = v(0) = 0$, in which case our argument fails to provide a separating filter. However, this separation is not needed if we only need to approximate filters which are constant on the set of inputs which have zero value at $t = 0$.

[12]If one prefers, one can replace the non-negative weighted sums of the filters $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$ in this alternative approximation result by sigmoidal neurons applied to non-negative weighted sums of the filters $\mathcal{H}_{\gamma_1}, \ldots, \mathcal{H}_{\gamma_m}$.

This follows from the following lemma, which is, in turn, a small variation of the Stone-Weierstrass Theorem. Given a compact Hausdorff topological space U, and a closed subset S of U, we say that a function $f : U \to \mathbb{R}$ is *S-constant* if the restriction of f to S is a constant function. We say that a class of real-valued functions on U is *S-separating* if, for each $u, v \in U$, $u \neq v$, such that not both of u and v are in S, there is some $f \in F$ such that $f(u) \neq f(v)$.

Lemma 3.8. *Suppose that F is a class consisting of continuous and S-constant functions which S-separates. Then, polynomials in elements of F approximate every S-constant continuous function $U \to \mathbb{R}$.*

This Lemma is proved as follows. We consider the quotient space $U_S := U/S$, where we collapse all points of S to one point, endowed with the quotient topology (its open sets are those open sets V in U for which $S \cap V \neq \emptyset$ implies $S \subseteq V$). The topological space U_S is compact, because the canonical projection onto the quotient is continuous, and U was assumed to be compact. In addition, U_S is a Hausdorff space, as follows from the fact that for each $x \notin S$ there are disjoint neighborhoods of x and S (since S is compact). The Lemma is established by noticing that continuous S-constant functions induce continuous functions on U_S, so that we may apply the standard Stone-Weierstrass Theorem to this quotient space.

Now the Theorem follows, using as S the set consisting of all inputs so that $u(0) = 0$. The S-separation property is established just as in the proof of Theorem 3.1; we omit the routine details. ∎

3.6 Combining Synaptic Dynamics with Membrane Dynamics

One other source of temporal dynamics in biological neural systems is the dynamics of the membrane potential of neurons. Hence it is of interest to consider a variation of our notion of a dynamic network where a function $x(t)$ is processed at a connection of the network first by a filter $\mathcal{H}$ that maps $x_i(\cdot)$ onto $w_i(\cdot) \cdot x_i(\cdot)$ (modeling synapse dynamics) and then by another filter $\mathcal{G}$ modeling membrane dynamics of the receiving neurons. Since each single EPSP and IPSP (i.e., each excitatory and inhibitory postsynaptic potential) can be fitted very well by a function of the form $\beta_1 e^{-\tau/\gamma_1} - \beta_2 e^{-\tau/\gamma_2}$, it appears to be justified to model membrane dynamics in the context of our model for population coding with pools of neurons by first order linear filters $\mathcal{G}$ with an impulse response $g(\tau)$ consisting of a weighted sum of functions of the form $e^{-\tau/\gamma}$. In the resulting variation of our notion of a dynamic network one replaces the filters $\mathcal{H}$ that model synaptic dynamics according to Definition 2.1 at the connections of the network by compositions $\mathcal{G} \circ \mathcal{H}$ with such linear filters $\mathcal{G}$. All preceding results remain valid for this variation of the network model. In order to approximate arbitrary given time invariant filters with fading memory by such networks one just has to show that for any two functions $u, v \in U$ with $u(t) \neq v(t)$ for some $t \leq 0$ there exist filters $\mathcal{H}, \mathcal{G}$ of the desired type such that $(\mathcal{G} \circ \mathcal{H})u(0) \neq (\mathcal{G} \circ \mathcal{H})v(0)$. This holds even for $u, v \in C(\mathbb{R}, [B_0, B_1])$ with $B_0 = 0$, since we just have to find a filter $\mathcal{H}$ modeling synaptic dynamics according to Definition 2.1 so that $\mathcal{H}u(t) \neq \mathcal{H}v(t)$ for some $t \leq 0$ (hence we can allow that $u(0) = v(0) = 0$). We then can apply to the functions $\mathcal{H}u(t)$ and $\mathcal{H}v(t)$ for $t \leq 0$ the argument from the proof of Theorem 3.1 to find a linear filter $\mathcal{G}$ with impulse response $e^{-\tau/\gamma}$ so that $(\mathcal{G} \circ \mathcal{H})u(0) \neq (\mathcal{G} \circ \mathcal{H})v(0)$. The same argument shows that theoretically the same class of filters can be approximated by dynamic networks if one only relies on linear filters $\mathcal{G}$ modeling membrane dynamics.

4 Computations on Spatio-Temporal Patterns

A closer look shows that many temporal patterns that are relevant for biological neural systems are not just temporal but spatio-temporal patterns. For example in auditory systems the additional spatial dimension parameterizes different frequency bands. These are represented by spatial locations of the inner hair cells on the cochlea , and corresponding spatial maps in higher parts of the auditory system. In visual systems it is obvious that the analysis of moving objects (and/or the stabilization of visual input in spite of body movements of the receiving organism) requires the processing of complex spatio-temporal patterns. In this context two spatial dimensions correspond directly to retina locations. But one should note that other “spatial” dimensions in the subsequent Definition 4.1 need not correspond to spatial locations in the outer world (or on the retina), but can also correspond to scales in a more abstract feature space, for example to spatial frequency or phase. Therefore we will consider in the following spatio-temporal patterns with an arbitrary finite number $d \in \mathbb{N}$ of spatial dimensions.

The transformation and classification of complex spatio-temporal patterns appears to be relevant also for higher cortical areas, since recordings from larger populations of neurons via voltage-sensitive dyes or MEG, EEG etc. suggest that sensory input is encoded in spatio-temporal activation patterns of associated cortical neural systems. These spatio-temporal activation patterns provide the input to higher cortical areas. Also the output of various systems in the motor cortex can be viewed as spatio-temporal patterns (Georgopoulos, 1995). Hence one may argue that also higher cortical areas carry out computations on spatio-temporal patterns.

We will show in this section that one can extend our analysis of computations on temporal patterns to an analysis of computations on spatio-temporal patterns. For that purpose we introduce a suitable extension of our definition of a dynamic network that allows for d spatial dimensions in the input functions u.

Definition 4.1. *We define a* spatial dynamic network *and the corresponding classes SDN and SDN* of filters as a variation of the preceding Definition 2.2 of a dynamic network. Fix some arbitrary $d \in \mathbb{N}$. A spatial dynamic network with d spatial input dimensions (in addition to the time dimension) assigns to vectors $\underline{u}$ of n input functions $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ an output function $z : \mathbb{R} \to \mathbb{R}$. The only difference to the preceding definition of a dynamic network is that now there exists for each network a finite set X of vectors $\underline{x} \in \mathbb{R}^d$ so that the actual input to the network consists of functions of the form $u(\underline{x}, \cdot)$ for $\underline{x} \in X$.*

According to this definition any spatial dynamic network samples the input functions $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ just at a fixed finite set X of points $\underline{x}$. Nevertheless we will show in Theorem 4.2 that these networks can approximate a very large class of filters on functions $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$.

The notion of a Volterra series (see Remark 2.3) can be readily extended to input functions $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ (again we assume that u is measurable and essentially bounded).

In this case a k-th order Volterra term is of the form

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \int_{0}^{\infty} \ldots \int_{0}^{\infty} u(x_1, \ldots, x_d, t-\tau_1) \cdot \ldots \cdot u(x_1, \ldots, x_d, t-\tau_k) \cdot h(x_1, \ldots, x_d, \tau_1, \ldots, \tau_k)\, dx_1 \ldots dx_d\, d\tau_1 \ldots d\tau_k \tag{8}$$

for some function $h \in L^1$. Analogously as before we refer to a series consisting of finitely many such terms as a Volterra polynomial or finite Volterra series.

Theorem 4.2. *Let U be the class of functions in $C(\mathbb{R}^d \times \mathbb{R}, [B_0, B_1])$ that satisfy $|u(\underline{x}, t) - u(\underline{\tilde{x}}, \tilde{t})| \leq B_2 \parallel \langle \underline{x}, t\rangle - \langle \underline{\tilde{x}}, \tilde{t}\rangle \parallel$ for all $\langle \underline{x}, t\rangle$, $\langle \underline{\tilde{x}}, \tilde{t}\rangle \in \mathbb{R}^d \times \mathbb{R}$, where $d \in \mathbb{N}$ and B_0, B_1, B_2 are arbitrary real-valued constants with $0 < B_0 < B_1$ and $0 < B_2$. Then the following holds for any filter $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$:*

$\mathcal{F}$ can be approximated by spatial dynamic networks (i.e., for any $\varepsilon > 0$ there exists some $\mathcal{S} \in$ SDN such that $|\mathcal{F}\underline{u}(t) - \mathcal{S}\underline{u}(t)| < \varepsilon$ for all $\underline{u} \in U^n$ and $t \in \mathbb{R}$)

if and only if *$\mathcal{F}$ can be approximated by a sequence of (finite or infinite) Volterra series.*

*The claim remains valid if SDN is replaced by SDN**

Proof. We show that the two alternative conditions on $\mathcal{F}$ in the claim of the theorem are equivalent by proving that both conditions are equivalent to a third condition: the condition that $\mathcal{F}$ is time invariant and has fading memory – for the straightforward extension of this notion to filters $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$, where U is now a class of functions from $R^d \times \mathbb{R}$ into $\mathbb{R}$. We say that such filter $\mathcal{F}$ has fading memory if for every $\langle v_1, \ldots, v_n\rangle \in U^n$ and every $\varepsilon > 0$ there exist $\delta > 0, T > 0$, and $K > 0$ so that $|\mathcal{F}\underline{v}(0) - \mathcal{F}\underline{u}(0)| < \varepsilon$ for all $\underline{u} = \langle u_1, \ldots, u_n\rangle \in U^n$ with the property that $\parallel \underline{v}(\underline{x}, t) - \underline{u}(\underline{x}, t) \parallel < \delta$ for all $t \in [-T, 0]$ and all $\underline{x} \in [-K, K]^d$. Note that this condition implies "fading influence" of $u(\underline{x}, t)$ for arguments $\langle \underline{x}, t\rangle$ where $|t|$ or $\parallel \underline{x} \parallel$ are very large.

It is obvious that any k-th order Volterra term of the form (8) is time invariant and has fading memory. Furthermore this property is preserved by taking sums and limits (analogously as in Lemma 2.5). Hence also for filters with inputs $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ we have that any filter which can be approximated by a sequence of finite or infinite Volterra series is time invariant and has fading memory. On the other hand one can extend the proof of Theorem 1 in (Boyd and Chua, 1985) in a straightforward manner to show that any time invariant filter that has fading memory and receives inputs from U^n (for a class U with the properties as in the claim of the theorem) can be approximated arbitrarily closely by Volterra polynomials. For this extension of the argument of (Boyd and Chua, 1985) one just has to verify that this class U is compact with regard to the topology generated by the neighborhoods $\{\underline{u} \in U^n : \parallel \underline{v}(\underline{x}, t) - \underline{u}(\underline{x}, t) \parallel < \varepsilon$ for all $t \in [-T, 0]$ and all $\underline{x} \in [-K, K]^n\}$ for arbitrary $\underline{v} \in U^n$ and $\varepsilon, T, K > 0$.

It is clear that any spatial dynamic network according to Definition 4.1 is time invariant and has fading memory. Thus it only remains to show that any time invariant filter $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$ with fading memory can be approximated arbitrarily closely by spatial dynamic networks. In order to extend the proof of Theorem 1 in (Boyd and Chua, 1985) to this case one just has to observe that the proof of Theorem 3.1 implies that for any two functions $u, v, \in U$ with $u(\underline{x}, t) \neq v(\underline{x}, t)$ for some $t \leq 0$ and $\underline{x} \in \mathbb{R}^d$ there exists some $\underline{x} \in \mathbb{R}^d$ and some filter $\mathcal{H}$ modeling synaptic dynamics as in Definition 2.1 which satisfies $\mathcal{H}u(\underline{x}, \cdot)(0) \neq \mathcal{H}v(\underline{x}, \cdot)(0)$. Thus we have shown that a filter $\mathcal{F} : U^n \to \mathbb{R}^{\mathbb{R}}$ can be approximated by spatial dynamic networks if and only if $\mathcal{F}$ is time invariant and has fading memory. This completes the proof of Theorem 4.2. ■

Remark 4.3.

1. Analogous versions of Corollary 3.2, Remark 3.3, Theorem 3.4, Theorem 3.6, and Theorem 3.7 also hold for the framework of computations on spatio-temporal patterns considered in Theorem 4.2.

2. If one considers a system consisting of many spatial dynamic networks that provide separate outputs for different spatial output locations one also can produce spatio-temporal patterns in the *output* of such systems. Theorem 4.2 implies that exactly those maps $\mathcal{A}$ from spatio-temporal patterns to spatio-temporal patterns can be realized by such systems where the restriction of the output of $\mathcal{A}$ to any fixed output location is a time invariant filter $\mathcal{F}$ with fading memory.

5 Conclusion

We have analyzed in this article the power of feedforward models for biological neural systems for computations on temporal patterns and spatio-temporal patterns. We have identified the class of filters that can be approximated by such models, and shown that this result is quite stable with regard to changes in the model. In particular we have shown that all filters which can be approximated by Volterra series can be approximated by models consisting of a single layer of dynamic synapses and neurons. Furthermore the class of filters that can be approximated does not change if one considers feedforward networks with arbitrarily many layers. In addition the filters in this class are characterized by a very simple property (time invariance and fading memory) that is in general easy to check for a concrete filter. This class of filters is very rich. In fact, one might argue that any filter that is possibly useful for a function of a biological organism belongs to this class.

Since we have included in our analysis the case where *several* temporal patterns are presented simultaneously as input to a neural system, our approach also provides a new foundation for analyzing computations that respond in a particular manner to *temporal correlations* in the firing activity of different pools of neurons. We show that any such computation that can be described by time invariant filters with fading memory (which is for example the case for most conjectured computations involving binding of features belonging to the same object via temporal correlations) can in principle be carried out by a feedforward neural system.

So far the analysis of the possible functional role of short term synaptic dynamics has found the most convincing computational role for synaptic depression: Our results in this article point to a possible computational role for the other important dynamical mechanism in biological synapses: facilitation. We show that via facilitation models for neural systems gain the power to approximate any filter in the very large class of linear and nonlinear filters described above. Furthermore we have shown that this possible function of facilitation does not interfere with any other computational role of synaptic depression, since we have shown that for any fixed depression mechanisms one can find parameters for the synaptic facilitation mechanisms that allow the approximation of arbitrary filters from this class. Apart from this we have also shown in section 3.3 that the same very rich class of linear and nonlinear filters can be approximated by models for neural systems whose synapses just exhibit depression, not facilitation.

The view of neural systems as computational models for computations on temporal patterns or spatio-temporal patterns – rather than on static vectors of numbers – is likely to have significant consequences for the analysis of learning in neural systems. It suggests that learning should be analyzed in the context of adaptive filters. Whereas we do not contribute directly to any learning result in this article, our results identify exactly the class of filters within which such filter adaptation would take place, and thereby prepare the ground for a closer analysis of learning in neural systems from the point of view of adaptive filtering.

Finally we would like to point out that our "universal approximation results" for computations on temporal and spatio-temporal patterns suggest a new complexity measure and a new parameterization for nonlinear filters in this domain, which may be more appropriate in the context of biological neural systems. We show that instead of measuring the complexity of a nonlinear filter $\mathcal{F}$ by the degree of the Volterra polynomial or Wiener polynomial that is needed to approximate $\mathcal{F}$ within a given ϵ, one can also measure the complexity of $\mathcal{F}$ by the number of sigmoidal gates and dynamic synapses that are needed to approximate $\mathcal{F}$ within ϵ. Our results show that both complexity hierarchies characterize the same class of linear and nonlinear filters. However the latter measure is more adequate in the context of neural computation, because already the approximation of a single sigmoidal gate requires a high order Volterra polynomial for a good approximation. Hence the order of the Volterra polynomial required to approximate a given nonlinear filter $\mathcal{F}$ is in general a poor measure for the cost of implementing $\mathcal{F}$ in neural hardware. On the other hand the alternative complexity measure for filters $\mathcal{F}$ that is suggested by our results is closely related to the cost of implementing $\mathcal{F}$ in neural hardware.

In addition our approach via formal models for dynamic networks provides a new parameterization for all filters that are approximable by Volterra series – in terms of parameters that control the architecture as well as the temporal dynamics and scale of synaptic dynamics. Such parameterization is in particular of interest for the analysis of learning (if the goal is to learn a map from spatio-temporal to spatio-temporal patterns), especially since the parameters that occur in our new parameterization appear to be related to those parameters that are "plastic" in biological neural systems.

This article also prepares the ground for an investigation of the required complexity of models for neural systems for approximating specific filters that are of particular interest in this context. Preliminary computer simulation results (Natschläger et al., 1999) suggest that in fact quite small instantiations of the dynamic network models considered in this article suffice to approximate specific quadratic filters. Other topics of current research are the role of noise in this context, and the possible role of lateral and recurrent connections in the network (see (Maass, 1999b)).

Acknowledgement:

We would like to thank Anthony M. Zador for bringing the problems addressed in this article to our attention. In addition we would like to thank Larry Abbott and two anonymous referees for helpful comments.

References

Abbott, L. F. (1994). Decoding neuronal firing and modelling neural networks. *Quarterly Review of Biophysics*, 27:291–331.

Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science*, 275:220–224.

Back, A. D., & Tsoi, A. C. (1991). FIR and IIR synapses, a new neural network architecture for time series modeling. *Neural Computation*, 3:375–385.

Boyd, S., & Chua, L. O. (1985). Fading memory and the problem of approximating nonlinear oparators with Volterra series. *IEEE Trans. on Circuits and Systems*, 32:1150–11.

Dasgupta, B., & Sontag, E. D. (1996). Sample complexity for learning recurrent perceptron mappings. *IEEE Trans. Inform. Theory*, 42:1479–1487.

Dieudonne, J. (1969). *Foundations of Modern Analysis*. Academic Press, New York.

Dobrunz, L., & Stevens, C. (1997). Heterogenous release probabilities in hippocampal neurons. *Neuron*, 18:995–1008.

Fliess, M. (1975). Un outil algebrique: les series formelles non commutatives. In Marche, G., editor, *Mathematical Systems Theory*, pages 122–148. Springer New York, Udine.

Folland, G. B. (1984). *Real Analysis*. Wiley, New York.

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, 34, 1089–1104.

Gallman, P. G., & Narendra, K. S. (1976). Representations of nonlinear systems via the Stone-Weierstrass theorem. *Automatica*, 12:618–622.

Georgopoulos, A. P. (1995). Reaching: coding in motor cortex. In Arbib, M. A., editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge.

Georgopoulos, A. P., Schwartz, A. P., & Ketner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233:1416–1419.

Gerstner, W. (1999). Spiking neurons. In Maass, W. & Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press, Cambridge, 32–53. Preprint electronically available at `http://diwww.epfl.ch/lami/team/gerstner/wg\_pub.html`.

Grossberg, S. (1969). On the production and release of chemical transmitters and related topics in cellular control. *J. Theor. Biol.*, 22, 325–364.

Grossberg, S. (1972). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences*, 15, 253–285.

Grossberg, S. (1984). Some psychophysiological and pharmacological correlates of a developmental, cognitive, and motivational theory. In: *Brain and information: event related potentials*, Karrer, R., Cohen, J., Tueting, P., eds., New York Academy of Science, New York, 58–142.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley.

Hewitt, E. & Stromberg, K. (1965). *Real and Abstract Analysis*. Springer-Verlag, Berlin.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366.

Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons.* Oxford University Press, New York.

Koiran, P., & Sontag, E. D. (1998). Vapnik-Chervonenkis dimension of recurrent neural networks. *Discrete Applied Math.*, 86:63–79.

Leshno, M., Lin, V., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6:861–867.

Maass, W. (1999). On the computational power of winner-take-all, submitted for publication.

Maass, W. (1999). On the role of noise and recurrent connections in neural networks with dynamic synapses, in preparation.

Maass, W., & Natschläger, T. (1999). A model for fast analog computation based on unreliable synapses. *Neural Computation*, in press.

Maass, W., & Zador, A. (1998). Dynamic Stochastic Synapses as Computational Units. *Advances in Neural Information Processing Systems*, vol. 10, MIT-Press (Cambridge), 1998, 194–200; journal version in: *Neural Computation*, vol. 11, 1999, 903–917.

Maass, W. & Zador, A. (1999). Computing and learning with dynamic synapses. In Maass, W. and Bishop, C., editors, *Pulsed Neural Networks*. MIT-Press, Cambridge.

Marmarelis, P. Z. & Marmarelis, V. Z. (1978). *Analysis of Physiological Systems: The White-Noise Approach*. Plenum Press, New York.

Mead, C. (1989). *Analog VLSI and Neural Systems*. Addison-Wesley (Reading).

Murthy, V., Sejnowski, T., & Stevens, C. (1997). Heterogeneous release properties of visualized individual hippocampal synapses. *Neuron*, 18:599–612.

Natschläger, T., Maass, W., & Zador, A. (1999). Temporal processing with dynamic synapses, in preparation.

Palm, G. (1978). On representation and approximation of nonlinear systems. *Biol. Cybernetics*, 31:119–124.

Palm, G. & Poggio, T. (1977). The Volterra respresentation and the Wiener expansion: validity and pitfalls. *SIAM J. Appl. Math.*, 33(2):195–216.

Poggio, T. & Reichardt, W. (1980). On the representation of multi-input systems: computational properties of polynomial algorithms. *Biol. Cybernetics*, 37:167–186.

Rieke, F., Warland, D., Bialek, W., & de Ruyter van Steveninck, R. (1997). *SPIKES: Exploring the Neural Code*. MIT-Press, Cambridge.

Rugh, W. J. (1981). *Nonlinear System Theory*. John Hopkins University Press, Baltimore.

Sandberg, I. W. (1991). Structure theorems for nonlinear systems. *Multidimensional Systems and Signal Processing*, 2:267–286.

Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York.

Sontag, E. D. (1997). Recurrent neural networks: Some systems-theoretic aspects. In Karny, M., Warwick, K., & Kurkova, V., editors, *Dealing with Complexity: a Neural Network Approach*, pages 1–12. Springer-Verlag, London.

Sontag, E. D. (1998a). A learning result for continuous-time recurrent neural networks. *Systems and Control Letters*, 34:151–158.

Sussmann, H. J. (1975). Semigroup representations, bilinear approximations of input-output maps, and generalized inputs. In Marchesini, G., editor, *Mathematical Systems Theory*, pages 172–192. Springer, New York, Udine.

Tsodyks, M., & Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc.Natl.Acad.Sci.*, 1994:719–723.

Tsodyks, M., Pawelzik, K., & Markram, H. (1998). Neural networks with dynamic synapses. *Neural Computation*, 10, 821–835.

Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L. F., & Nelson, S. B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *Journal of Neuroscience*, 17:7926–7940.