The Neurobiological Basis of Cognition: Identification by Multi-Input, Multioutput Nonlinear Dynamic Modeling:

A method is proposed for measuring and modeling human long-term memory formation by mathematical analysis and computer simulation of nerve-cell dynamics

Theodore W. Berger [Fellow IEEE], **Dong Song [Member IEEE]**, **Rosa H. M. Chan [Student Member IEEE]**, and **Vasilis Z. Marmarelis [Fellow IEEE]**
Center for Neural Engineering, USC Biomedical Simulations Resource, Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089-1111 USA

Theodore W. Berger: berger@bmsr.usc.edu; Dong Song: dsong@usc.edu; Rosa H. M. Chan: homchan@usc.edu; Vasilis Z. Marmarelis: marmarelis@hotmail.com

Abstract

The successful development of neural prostheses requires an understanding of the neurobiological bases of cognitive processes, i.e., how the collective activity of populations of neurons results in a higher level process not predictable based on knowledge of the individual neurons and/or synapses alone. We have been studying and applying novel methods for representing nonlinear transformations of multiple spike train inputs (multiple time series of pulse train inputs) produced by synaptic and field interactions among multiple subclasses of neurons arrayed in multiple layers of incompletely connected units. We have been applying our methods to study of the hippocampus, a cortical brain structure that has been demonstrated, in humans and in animals, to perform the cognitive function of encoding new long-term (declarative) memories. Without their hippocampi, animals and humans retain a short-term memory (memory lasting approximately 1 min), and long-term memory for information learned prior to loss of hippocampal function. Results of more than 20 years of studies have demonstrated that both individual hippocampal neurons, and populations of hippocampal cells, e.g., the neurons comprising one of the three principal subsystems of the hippocampus, induce strong, higher order, nonlinear transformations of hippocampal inputs into hippocampal outputs. For one synaptic input or for a population of synchronously active synaptic inputs, such a transformation is represented by a sequence of action potential inputs being changed into a different sequence of action potential outputs. In other words, an incoming temporal pattern is transformed into a different, outgoing temporal pattern. For multiple, asynchronous synaptic inputs, such a transformation is represented by a spatiotemporal pattern of action potential inputs being changed into a different spatiotemporal pattern of action potential outputs. Our primary thesis is that the encoding of short-term memories into new, long-term memories represents the collective set of nonlinearities induced by the three or four principal subsystems of the hippocampus, i.e., entorhinal cortex-to-dentate gyrus, dentate gyrus-to-CA3 pyramidal cell region, CA3-to-CA1 pyramidal cell region, and CA1-to-subicular cortex. This hypothesis will be supported by studies using *in vivo* hippocampal multineuron recordings from animals performing memory tasks that require hippocampal function. The implications for this hypothesis will be discussed in the context of "cognitive prostheses"—neural prostheses for cortical brain regions believed to support cognitive functions, and that often are subject to damage due to stroke, epilepsy, dementia, and closed head trauma.

Keywords

Cognition; hippocampus; memory; modeling; nonlinear; systems analysis

I. INTRODUCTION

Cognitive functions such as language, abstract reasoning, and learning and memory have long been held to represent the most complex operations of the brain. Thus, it is not surprising that cognitive functions also have been the most difficult of brain operations to define in terms of underlying neural function and neural mechanisms. Cognition most often is defined in terms of theoretical constructs, for example, "information" or "recognition," and operations on those constructs, such as "information processing." Theoretical approaches to cognition, although often successful at the level of inferred cognitive operations and behavior, have difficulty in bridging the gap to neuronal functions (e.g., postsynaptic potentials, or PSPs; action potentials, or APs or "spikes") and especially in bridging the gap to mechanisms underlying neuronal function (e.g., presynaptic calcium channel kinetics and neurotransmitter release, receptor-channel kinetics, membrane biophysics, synaptic plasticity, etc.). Without common points of registry for the conceptual hierarchies of a neurobiological framework and any theoretical framework for cognition, it becomes difficult if not impossible, to understand a cognitive process in terms of a corresponding neurobiological process, and *vice versa*. Although fMRI and other imaging methods hold promise for contributing to the solution of this problem, neither the spatial-temporal resolution, nor the generalizability of these technologies are yet at a level to provide the bridge required.

We propose an operational definition of the neurobiological basis of cognition using a combined experimental/theoretical approach designed to measure cognitive processes directly, and to describe them mathematically. Our approach is based on principles of nonlinear systems identification, first developed in the field of engineering [1]–[3]. We and our colleagues have spent much of the last 30 years adapting these principles to neurobiological systems, and specifically to the hippocampus, a brain region responsible for long-term memory formation [4]. In our approach, each neuron is considered the fundamental operating unit of a given neural system, consistent with the "neuron doctrine" of Ramon y Cajal in the early part of the 20th century [5]. Neurons generate output signals in the form of all-or-none APs that propagate to other neurons (typically tens to hundreds of other neurons) along "axons" that end in specialized contacts known as "synapses" (Fig. 1). Each AP input (a neuron may receive hundreds to thousands of such inputs) generates a synaptic response that can be depolarizing (excitatory postsynaptic current, EPSC, or potential, EPSP) or hyperpolarizing (inhibitory postsynaptic current, IPSC, or potential, IPSP). If inputs to a neuron cause the resting membrane potential (typically −70 mV relative to the extracellular fluid) to depolarize to −55 mV or more, a "threshold" is crossed which results in the generation of an output AP (this number for threshold varies considerably from neuron to neuron, and should be considered very "approximate").

All of these concepts deserve much more detailed consideration for an understanding of the biophysical properties of neurons and/or fundamental principles of synaptic transmission [6], [7]. In this paper, however, we will focus on a few elemental concepts that derive from essential properties of neurons and neural networks, and that are key in determining the theoretical and experimental approach used in our research and described here. We wish to first identify these concepts, and then explain how they have shaped our approach to studying neural function at synaptic, neuron, and network levels of organization. We propose that experimental measurements and mathematical modeling of network function, using the formalisms identified, provide the best available direct observation of high-level neural system function, and thus, a real, definable, and available neural counterpart to "cognitive processes."

A. Neurons and Neural Networks as Hierarchically Organized, Dynamical Systems

Among these elemental concepts is that of "dynamics," in other words, the fact that the EPSC and EPSP shown in Fig. 1 do not have a single value amplitude, but instead, have an amplitude that evolves over time. EPSPs reflect EPSCs flowing across the resistance and capacitance of the cell membrane. The amplitude–time course of the EPSCs reflect the probabilistic opening and closing of large numbers of channels in the postsynaptic region of the cell membrane; the channels are activated by neurotransmitter released by vesicles of the presynaptic element, and the binding of that neurotransmitter to the postsynaptic receptor. The dynamics of the relation between receptor binding and channel state are typically described with kinetic models. Because the probability of channel opening is large initially, and then gradually decreases, the EPSC and the subsequent EPSP have the shape that they do: a sharp rise followed by an exponential decay. What also is crucial to an understanding of brain function, however, is that any biological neural network consists of a hierarchical organization of dynamical systems [8]–[12]. The dynamics of molecular interactions of receptor and channel subunits determine the dynamics of EPSCs; dynamics of EPSCs determine the dynamics of EPSPs. Within each layer of the hierarchy, elements can be "triggered" or "activated," but the activity of a given element then evolves largely according to its own internal dynamics.

There also can be interactions between levels in the hierarchy, however. "Internal dynamics" strongly influence the response of a mechanism like an AMPA receptor-channel to an external input; AMPA receptor-channel kinetics are unlikely to change substantially unless there is a genetic mutation of one of its subunits [7]. However, the kinetics of other receptor-channel complexes like the NMDA type, include elements that are voltage-dependent (the voltage-dependent blockade of the NMDA channel by Mg^{2+} must be relieved by depolarization of the local membrane), and thus, are influenced by a property of the next higher level in the hierarchy, i.e., the neuron. The transmembrane voltages induced by other inputs surrounding any one NMDA receptor-channel are integrated by the postsynaptic neuron to determine the local membrane voltage. This local voltage at the level of the neuron is the source of a feedback to the lower level of synapses, to shape the amplitude–time course of the NMDA-mediated EPSC (see [8], [9], [13], for a formalism to describe the neural hierarchy).

B. Neurons and Neural Networks as Hierarchically Organized, Nonlinear Dynamical Systems

Another key concept to understanding brain function underlying cognition is the "nonlinearity" of virtually all synaptic and neural mechanisms. What is meant by nonlinearity is straightforward to define, though not so straightforward to measure, and to measure accurately. The definition of nonlinearity, in the context of neural synaptic transmission, is that the response of a postsynaptic neuron to the second of two successive presynaptic stimuli is not predictable by the principle of superposition. Consider the hypothetical examples shown in Fig. 2. The input pulse, when delivered alone (x_a), generates an output response, (y_a) that exhibits a relatively rapid rise and an exponential decay typical of EPSP-like waveforms. In the second case, (x_b, y_b), two pulses are delivered with an inter-impulse interval such that the second pulse is delivered before the response to the first pulse is completed. This results in postsynaptic responses that are partially overlapping, and notably, the resulting compound EPSP is not equivalent to a simple summation of the two individual EPSPs. Any such deviation from "superposition" is identified as a "nonlinearity." In this hypothetical example, the resulting response is more than the summation predicted by superposition, i.e., a "facilitative" second-order nonlinearity; a response less than that predicted by superposition is identified as a "suppressive" second-order nonlinearity. Importantly, observable overlap between responses

to successive inputs is not required for the generation of nonlinearities. The observable response to a given input may be completed (i.e., the response returns to baseline), but that input event may also have initiated, for example, unobservable activation of biochemical second messenger systems intrinsic to the postsynaptic neuron, and/or excitation of local interneurons that provide feedback to the target cell from which recordings are obtained. The effects of these secondary inputs may not be observable until expressed in the context of another direct synaptic input (see Fig. 2, and [14]).

The first two examples (second and third pairs of panels) of nonlinearities considered in Fig. 2 are second-order nonlinearities: deviations from linear summation of responses to single impulses. We also can consider summation of second-order nonlinearities, i.e., summation of two responses where: 1) each response is elicited by the second of a pair of stimulations and 2) the response to at least one pair includes a nonlinearity. This possibility is shown in Fig. 2, the bottom three pairs of panels. Panels x_b, y_b and x_c, y_c each demonstrate a significant nonlinearity in response to two different interimpulse intervals. When these two facilitative nonlinearities are combined into a triplet, however, any expected summation of the two facilitations instead is revealed as a strong suppression, expressed (x_d, y_d). We have observed this in our own studies of synaptic transmission in intrinsic hippocampal pathways, though we have yet to conclusively identify the explanation. We hypothesize that the first two pulses of the triplet activate a second messenger system which, in turn, hyperpolarizes the cell membrane, e.g., through activation of a Ca^{2+}-activated K^+ conductance.

1) Formal Definition of Nonlinearities—Before this discussion proceeds much further, we should pause to provide a mathematical framework useful for defining and quantitatively measuring nonlinearities. To be brief (see [1], [2], and [15], for more complete discussions of the fundamentals related to nonlinearities of biological systems), the work of Volterra [16], Wiener [17], Marmarelis [1], [2], and others (see Marmarelis reviews [1], [2]) has established that for any nonlinear, time-invariant (stationary) system with finite memory, the system output y can be represented as a functional power series of the input x, as in the single-input, single-output, discrete-time case

$$y(t)=k_0+\sum_{\tau=0}^{M}k_1(\tau)x(t-\tau)+\sum_{\tau_1=0}^{M}\sum_{\tau_2=0}^{M}k_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)+\sum_{\tau_1=0}^{M}\sum_{\tau_2=0}^{M}\sum_{\tau_3=0}^{M}k_3(\tau_1,\tau_2,\tau_3)x\times(t-\tau_1)x(t-\tau_2)x(t-\tau_3)+\ldots$$

(1)

In this formulation, the system *dynamics* are expressed by the temporal convolutions of the input and the Volterra kernel functions k; the system *nonlinearity* is expressed in the form of multiple convolutions of the input and the higher order (above first order) kernel functions. Kernel functions k thus represent the input–output *nonlinear dynamics* of the system. The zeroth-order kernel, k_0, is the value of the output when the input is absent, i.e., spontaneous activity. The first-order kernel, k_1, describes the linear dynamic relation between the input and the output, as a function of the time interval (τ) between the present time and past time. The second-order kernel, k_2, describes the second-order pairwise nonlinear dynamic relation between x and y. The third-order kernel, k_3, describes the third-order triplet-wise nonlinear dynamic relation between x and y, and so on. Higher order kernels, e.g., the fourth-order kernel, are not shown in this equation. The formal relation between the Volterra kernels and the single-pulse, paired-pulse, triple-pulse responses shown in Fig. 2 will be described more fully in Section II [(30), (31)].

2) Relation Between Nonlinearities and Cellular Mechanisms—Cellular
mechanisms exhibiting second- and third-order nonlinearities are common throughout the
nervous system. It is fair to state that the great majority of mechanisms underlying nervous
system functionality exhibit strong second-order nonlinearities, with third and higher order
nonlinearities being common rather than rare. Examples of second- and third-order
nonlinearities for hippocampal EPSP recordings already have been shown in Fig. 2. Note
that the strong facilitation of EPSP amplitude to the second pulse of the triplet almost
certainly reflects residual calcium accumulation presynaptically [18]. The first pulse
activates voltage-dependent calcium channels located presynaptically; the resulting calcium
entry binds with a family of presynaptic molecules to initiate integration of
neurotransmitter-containing vesicles to the presynaptic membrane, and the subsequent
release of neurotransmitter into the presynaptic cleft. The time course for removal of free
calcium from the presynaptic, intracellular space is approximately 50 ms [19]. If a second
pulse activates the same presynaptic fibers within that time period, the calcium entry caused
by the second pulse will sum with the residual calcium from the first pulse, resulting in a
larger amount of neurotransmitter released and thus a larger postsynaptic response.

Note that in Fig. 2, the suppression of the response to the third pulse of the triplet does not
represent a "ceiling effect" (or saturation): the response amplitude to the third pulse in the
triplet is substantially less than the large amplitude response to the second pulse. Instead, the
first two pulses of the triplet initiate intracellular mechanisms and/or feedback circuitry that
actively suppress the glutamate-induced depolarization [20], [21]. In intracellular studies
conducted previously [22]–[24], it was demonstrated experimentally that the majority of the
second-order suppression is induced by GABA-mediated inhibition acting through type A
and type B receptor subtypes.

Many of the electrical stimulation protocols that are commonly used to elicit characteristic
response profiles from target cells, or to reveal particular currents, provide additional
insights into the mechanisms underlying non-linearities. For example, T-type calcium
currents are sometimes studied by slightly hyperpolarizing the neuron cell membrane (to
bring the majority of channels out of inactivation) and, while in that hyperpolarized state,
delivering a depolarization (approximately 10 mV) [25]. These requirements for activating
T-type channels would suggest third-order nonlinearities emerging from the requirements of
excitatory input delivered following previous excitation; the first excitation must be
sufficient to induce GABAergic inhibitory feedback, and the second excitatory barrage must
occur within a specific time window to avoid desensitization of the T-type channels. Other
calcium channels, e.g., the N-type and the L-type, require different conditions for their
activation. Near-selective activation of L-type channels requires a period (e.g., 50 ms) of
depolarization (from rest, −75 mV to approximately 0 mV), followed by an additional
depolarization (e.g., to +20 mV). L-type calcium current then will continue to flow provided
the depolarization remains, given that L-type calcium channels exhibit little-to-no
desensitization. It is difficult to estimate *a priori* the degree of nonlinearity associated with
L-type calcium channel dynamics, but it certainly would be at least of third order, and may
extend across two or more orders of nonlinearity. N-type calcium channel dynamics will lie
somewhere between those of the T-type and the L-type.

From these examples, we hope that at least some principles have become clear. Namely,
kernels and input–output models in general, provide a different arsenal of measures for
looking at the same neurobiological mechanisms examined with other analytical tools used
in the neurosciences [14], [26], [27]. Most of the other methods and approaches, which we
will term here "mechanistic," emphasize products of the reductionist approach: analysis and
properties of a single mechanism, studied while isolated from the myriad of other
mechanisms with which that target mechanism usually interacts. Kernel functions and the

class of input–output models discussed here emphasize interactions between mechanisms. Given that cognitive processes must derive from systems-level dynamics, we would argue that input–output modeling is an essential component of any attempt to link cognition to neurobiological mechanisms.

Finally, input–output modeling has sometimes been called a "black box" approach, based on an assumption that practitioners of the approach do not know the box contents, i.e., the neurobiological mechanisms underlying the dynamics being modeled. This assumption is ludicrous on its face, of course. Our input–output modeling of hippocampus, and input–output modeling of other systems like the retina [28], have been accomplished with the same knowledge of the underlying circuitry, synaptic organization, and pharmacology as studies done on the same systems with mechanistic approaches. In fact, our studies of the role of GABAergic interneurons in second- and third-order nonlinearities of hippocampal dentate granule cells were guided by pharmacologically induced changes in intracellularly recorded membrane potentials [22]–[24]. Changes in the kernel functions occurred in response to interstimulus intervals and pairs of interstimulus intervals matching the time constants of $GABA_A$ and $GABA_B$ receptor kinetics, with drug-induced changes being specific for GABAergic receptor agonists and antagonists—in other words, the input–output studies used techniques, procedures and criteria nearly identical with mechanistic analyses. In the end, however, kernel analyses reveal more about the total system functionality, both because of the effects of broad-band input stimulation (activates many more mechanisms than traditional single-pulse, paired-pulse, or constant frequency stimulation), and because the formalism itself forces data interpretation and problem specification in terms of a neural network or neural systems level of analysis. For example, Fig. 3(a) shows a box diagram of the dentate gyrus in the intact rat, making explicit the relation between dentate granule cells (the principal neurons of the dentate), and many of the known pathways providing feedforward and feedback regulation of granule cells in response to excitation of input from the entorhinal cortex (similar feedback pathways for CA3 and CA1 are not shown). In the context of a continuous (average interimpulse interval: 500 ms), random interval (interval range: 1–5000 ms) impulse train stimulation of excitatory entorhinal input, it can be seen that granule cells will be activated monosynaptically, but also will be stimulated multisynaptically through the commissural, GABAergic, and other feedforward and feedback pathways intrinsic to the dentate. An equivalent representation for the pathways included in the hippocampal slice is shown in Fig. 3(b); the system can be reduced further with pharmacological blockade of GABAergic receptors. In this manner, the underlying anatomical pathways that contribute to dentate granule cell nonlinear dynamics can be readily identified, and used for interpretation of associated input-output models. Thus, input–output modeling is only as "black box," or uninterpretable, as the user. Recently, the relation between input–output models and mechanistic models has been formalized, and we have shown how both approaches can be used in a complementary manner [20], [21], [29].

C. Neurons and Neural Networks as Hierarchically Organized, Nonstationary, Nonlinear Dynamical Systems

The range of different dynamics found in the nervous system, and the magnitude and higher orders of nonlinearity found for those mechanisms studied to date, provide for considerable complexity of temporal pattern encoding. The degree of complexity increases even further when we consider that the dynamics being discussed to this point are not always constant, but instead, can change over time, or are "nonstationary." The learning and adaptive capabilities of the vertebrate and invertebrate nervous systems are well established. In addition, the last four decades of neuroscience research have seen experimental identification of a wealth of long-term, permanent changes in cellular and synaptic mechanisms that are induced by the learning process. All of this evidence has shown that

learning and memory do not involve "out of the ordinary" mechanisms that are reserved only for learning and memory, and that remain hidden and unexploited until environmental circumstances demand their amalgamation and use. In general, the mechanisms involved in learning and memory are the same mechanisms that underlie the biophysics and synaptic transmission of neurons in day-today circumstances: learning and memory simply require more of mechanism *x* or less of mechanism *y*. Given that the effects of mechanisms *x* and *y* are captured by, and contribute to, the kernels under nonlearning conditions, we should expect to see a relatively smooth change in system dynamics during the course of learning, i.e., there should not be a sudden and abrupt incorporation into the system of a radically different set of mechanisms, which would be reflected by a sudden and abrupt change in system nonlinearities. Although not an optimal test of this hypothesis, the above is precisely what we observed with the induction of long-term potentiation (LTP). The induction of LTP was accompanied by a smooth and gradual change in pre-LTP second- and third-order nonlinearities [30]. With regard to the Volterra kernel expressions introduced earlier, it thus is reasonable to incorporate cellular plasticity and learning and memory, or nonstationarities, simply by having the kernel expressions become a function of time, *t*, in addition to remaining a function of τ, the time since a prior input pulse

$$y(t) = k_0(t) + \sum_{\tau=0}^{M} k_1(t, \tau) x(t - \tau) + \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{M} k_2(t, \tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) + \sum_{\tau_1=0}^{M} \sum_{\tau_2=0}^{M} \sum_{\tau_3=0}^{M} k_3(t, \tau_1, \tau_2, \tau_3) x(t - \tau_1) x \times (t - \tau_2) x(t - \tau_3) + \ldots$$

(2)

Of course, there are more neurobiological processes than those underlying learning and memory that change as a function of time, and thus, would be reflected by nonstationarity of neural system kernels. Both the noradrenergic and the serotonergic neurotransmitter systems provide a widely dispersed input to much of the forebrain, thalamus, brainstem, and spinal cord. Both of these systems also change their levels of activity markedly during the sleep–wake cycle, with experiments demonstrating that the actions of norepinephrine and serotonin can significantly alter the responsiveness of recipient neurons to other, nonnoradrenergic and nonserotonergic afferents. For example, we have shown previously that large magnitude changes in noradrenergic levels in hippocampus are associated acutely with substantial changes in second- and third-order nonlinear responsiveness of dentate granule cells to excitatory, glutamatergic input from the perforant path, and inhibitory, GABAergic, input from inhibitory interneurons internal to the dentate gyrus [31].

Other processes having the longest time constants are likely to be those involved in development of the nervous system. Like changes in kernels representing learning, those representing development will not follow a pattern of deviating from a baseline of system characteristics, and then returning to that standard some predictable period of time later, as should be observed in the case of the dynamics of diurnal cycles. Instead, in the case of development, we would expect nonlinear system characteristics that slowly evolve into progressively richer, more stable, and more different (than the original) sets of system properties. This also allows for the exciting possibility that abnormal developmental and aging states that are difficult to diagnose (e.g., autism, schizophrenia, Alzheimer's disease) might be identified and differentiated with the new and varied set of quantifiable descriptors represented by the kernels, and which we propose to be capable of reflecting "system properties" of the neural circuitry underlying cognition.

D. Information Representation in the Ensemble Firing of Populations of Neurons

It has been demonstrated, particularly in cortical systems, that key information guiding trained, intentional behavior is represented in the "ensemble" firing of populations of neurons [32]–[39], i.e., spatiotemporal patterns of electrophysiological activity. The advent of multi-channel single-cell recording has provided the capability for simultaneously observing the firing of tens to hundreds of neurons, so that higher level analyses of the collective relations among subpopulations of neurons can be conducted [32], [40]–[42]. This has allowed confirmation of earlier suggestions of collective, ensemble activity in results from single cell recording studies [43].

How should this collective activity of subpopulations of neurons be interpreted in terms of cognitive processing? Clearly, when a subpopulation of neurons achieves and maintains a given spatiotemporal pattern, or a given "relatedness in activity," and which as a consequence allows for the identification of a relation between that pattern and an external event, it is reasonable to define that spatiotemporal pattern as a "representation" [44]. Representations are transient because neuron firing typically is maintained in one spatiotemporal pattern for only hundreds to thousands of milliseconds (restated, the duration of an identifiable spatiotemporal pattern is typically hundreds to thousands of milliseconds), unless we consider pathological conditions, e.g., rhythmic, cyclical firing characteristic of epilepsy. The latter and physiological rhythmicities, e.g., alpha rhythm, are indicative of "states" rather than the identities of specific external events.

With regard to hippocampus, such representations, or temporarily stable spatiotemporal patterns, could readily map onto individual memories, possibly even individual components of a memory. As the contents of a memory process, temporarily stable spatiotemporal patterns of activity within areas that provide input to hippocampus could constitute "short-term memories." With representations as "content," input–output transformations could be considered "process." Neural systems and brain regions process information by transforming incoming spatiotemporal patterns into different, outgoing spatiotemporal patterns. This statement is not a claim—there simply is no other reasonable interpretation of the basic phenomenology. Thus, information processing underlying cognition involves transformations of neural representations that are dynamic, nonlinear, and often nonstationary (time-varying). While recent advances in multielectrode technology have made it possible to record the simultaneous activities of populations of neurons in behaving animals, modeling such complex system behavior still remains one of the most challenging tasks in computational neuroscience [45]. It is in response to this need that we have invested over 20 years in the development and refinement of a combined experimental-theoretical strategy for quantitatively characterizing, and then modeling, neural systems typical of those routinely found in the mammalian brain.

II. EXPERIMENTAL-THEORETICAL STRATEGY FOR MODELING BRAIN COGNITIVE FUNCTIONS

We formulate here a three-step strategy to model the cognitive function of brain regions in general, and the hippocampus, in particular. In this strategy, we define the cognitive operation of a brain region as the transformation from its input activities to its output activities. Therefore, understanding the cognitive function of a brain region is equivalent to identifying its input–output transfer function S. Since in brain regions, input–output signals are manifested in the form of spatiotemporal patterns of neural spikes, i.e., all-or-none electrical events recorded from individual neurons, all parameters of the transfer function should be derived from the timings of the input/output spikes. The first two steps deal with the stationary and nonstationary aspects of the transfer function, respectively (Fig. 4 left,

middle). For the nonstationary case, the third step seeks to identify the "learning rule" underlying the nonstationarity of the transfer function (Fig. 4 right).

A. Stationary Modeling of Brain Regions

During performance of asymptotically learned behavior, a brain region is modeled as a time-invariant system. Its transformational property is modeled as a stationary process. A time-invariant (stationary) system is one whose transfer function does not depend on time. The modeling goal in this step then is to identify the time-invariant transformation S from multiple input spike trains X to the multiple output spike trains Y (3). Since the mechanisms underlying synaptic transmission and generation of spikes in neurons are inherently nonlinear and dynamical, the stationary model has to be a multiple-input, multiple-output (MIMO) nonlinear dynamical model

$$S:X \rightarrow Y. \tag{3}$$

In our approach, the MIMO model is decomposed into a series multiple-input, single-output (MISO) models (Fig. 5). Each MISO model is then formulated to have both parametric (i.e., mechanistic) and nonparametric (i.e., descriptive) components [46], [47]. First, the overall model structure is parameterized to be "neuron-like." It captures the stereotypical features of spiking neurons and explicitly includes variables that can be interpreted as the principal cellular processes such as the postsynaptic potential, the spike-triggered after-potential, the pre-threshold noise, and the spike-generating threshold. This configuration partitions the system nonlinear dynamics in a physiologically realistic manner, and thus facilitates comparison with intracellular recording results. The more versatile features of spiking neurons, i.e., the transformation from the input spikes to postsynaptic potentials and the transformation from the output spike to the after-potential, on the other hand, are modeled nonparametrically with the Volterra series, taking advantages of its flexibility in capturing nonlinear dynamics.

1) Model Configuration—The MISO model structure consists of five components (Fig. 5): 1) a feedforward block K transforming the input spike trains x to a continuous hidden variable u that can be interpreted as the postsynaptic potential; 2) a feedback block H transforming the *preceding* output spikes to a continuous hidden variable a that can be interpreted as the after-potential; 3) a noise term ε that captures the system uncertainty caused by both the intrinsic neuronal noise and the unobserved inputs; 4) an adder generating a continuous hidden variable w that can be interpreted as a prethreshold potential; and 5) a threshold function generating output spikes when the value of w crosses θ. The model can be expressed by the following equations:

$$w=u(k, x)+a(h, y)+\varepsilon(\sigma) \tag{4}$$

$$y=\begin{cases} 0 & \text{when } w<\theta \\ 1 & \text{when } w \geq \theta. \end{cases} \tag{5}$$

K takes the form of a Volterra model, in which u is expressed in terms of the inputs x by means of the Volterra series expansion as

$$u(t) = k_0 + \sum_{n=1}^{N} \sum_{\tau=0}^{M_k} k_1^{(n)}(\tau) x_n(t - \tau)$$

$$+ \sum_{n=1}^{N} \sum_{\tau_1=0}^{M_k} \sum_{\tau_2=0}^{M_k} k_{2s}^{(n)}(\tau_1, \tau_2) x_n(t - \tau_1) x_n(t - \tau_2)$$

$$+ \sum_{n_1=1}^{N} \sum_{\tau_2=1}^{n_1-1} \sum_{\tau_1=0}^{M_k} \sum_{\tau_2=0}^{M_k} k_{2x}^{(n_1,n_2)}(\tau_1, \tau_2) x_{n_1}$$

$$\times (t - \tau_1) x_{n_2}(t - \tau_2) + \ldots \qquad (6)$$

The zeroth-order kernel, k_0, is the value of u when the input is absent, for example, when

there is spontaneous variations in membrane potential first-order kernels, $k_1^{(n)}$, describe the linear relation between the nth input x_n and u, as functions of the time intervals (τ) between the present time and the past time. In other words, for each of the multiple inputs to the system, first-order kernels account for the effects of a single input event (a spike, or action potential) on the system membrane potential output, u, regardless of when those single input events may have occurred in the past, and thus, regardless of any other inputs that may have occurred between the past time designated by a particular (τ) and the present time. Second-

order self-kernels $k_{2s}^{(n)}$ describe the second-order nonlinear relation between the nth input, x_n, and u, as functions of the two time intervals (τ_1, τ_2) between the present time and the two respective past times. Thus, second-order kernels account for the modulatory effects of an input event occurring in the past on the system membrane potential output, u, evoked by a second input event occurring in the present, when both events occur on the same input. The previous input pulse may increase the response evoked by the present input, i.e., cause facilitation, or may reduce the response evoked by the present input, i.e., cause suppression.

Second-order cross-kernels $k_{2x}^{(n_1,n_2)}$ describe the second-order nonlinear interactions between each unique pair of input events (x_{n1} and x_{n2}) as they affect u, when each of those pulse events occurs on different inputs. N is the number of inputs. M_k denotes the memory length of the feedforward process. Higher order kernels, e.g., third-and fourth-order kernels, are not shown in this equation, but should be obvious by extension from the explanations above.

Similarly, H takes the form of a first-order Volterra model as in

$$a(t) = \sum_{\tau=1}^{M_h} h(\tau) y(t - \tau) \qquad (7)$$

where h is the linear feedback kernel. M_h is the memory length of the feedback process (note τ starts from 1 instead of 0 to avoid predicting the current output with itself). The noise term ε is modeled as a Gaussian white noise with standard deviation σ.

In summary, what the Volterra representation states is that subthreshold variation in membrane potential for any one neuron can be accounted for by variation in the temporal pattern of past action potentials for any one input, or, variation in the spatiotemporal pattern of past action potentials for a population of inputs to that neuron. In total, with all of its components, the model states that, for a population of neurons (input) that provide synaptic input to a second population of neurons (output), variation in the spatiotemporal pattern of past action potentials for the input neurons predicts the spatiotemporal pattern of action

potentials for the output population of neurons. We know from what are now tenants of fundamental neuroscience that, in general, such an input–output relation must be true. Outstanding issues relate more to whether or not such a relationship can be quantified or modeled, and whether or not experimental evidence supports such a model to the extent that it can be used to predict the effects of arbitrary input patterns. We report here that both of the latter questions can be answered in the affirmative.

2) Model Estimation—With the model structure defined as above, the next step is to estimate all model parameters, i.e., feedforward kernels k, feedback kernels h, prethreshold noise standard deviation σ, and threshold θ, from the timings of the input/output spikes. The biggest challenges in Volterra modeling is the large number of open parameters (coefficients) to be estimated, especially in the cases of high dimensional input and high order model. To solve this problem, Laguerre expansion of the Volterra kernels (LEV) and statistical model selection techniques are employed [46], [47].

With LEV, Volterra kernels (k and h) are expanded with orthonormal Laguerre basis functions b [20], [48], [49]. Equations (6) and (7) are rewritten into

$$u(t)=c_0+\sum_{n=1}^{N}\sum_{j=1}^{L}c_1^{(n)}(j)v_j^{(n)}(t)+\sum_{n=1}^{N}\sum_{j_1=1}^{L}\sum_{j_2=1}^{j_1}c_{2s}^{(n)}(j_1,j_2)v_{j_1}^{(n)}(t)v_{j_2}^{(n)}(t)+\sum_{n_1=1}^{N}\sum_{n_2=1}^{n_1-1}\sum_{j_1=0}^{L}\sum_{j_2=0}^{L}c_{2x}^{(n_1,n_2)}\times(j_1,j_2)v_{j_1}^{(n)}(t)v_{j_2}^{(n)}(t)+\ldots$$

(8)

$$a(t)=\sum_{j=1}^{L}c_h(j)v_j^{(h)}(t)$$

(9)

where v are the convolution of input–output spike trains (x and y) and Laguerre basis functions b

$$v_j^{(n)}t=\sum_{\tau=0}^{M_k}b_j(\tau)x_n(t-\tau),\ \ v_j^{(h)}(t)$$
$$=\sum_{\tau=1}^{M_h}b_j(\tau)y(t-\tau).$$

(10-11)

$c_1^{(n)},c_{2s}^{(n)},c_{2x}^{(n_1,n_2)}$, and c_h are the sought Laguerre expansion coefficients of $k_1^{(n)},k_{2s}^{(n)},k_{2x}^{(n_1,n_2)}$, and h, respectively (c_0 is equal to k_0). The number of basis functions (L) is typically much smaller than the memory length (M_k and M_h), so the total number of coefficients is greatly reduced [46], [47].

All model parameters can be estimated using a maximum-likelihood method. The negative log-likelihood function L is

$$L(y|x,k,h,\sigma,\theta)=-\sum_{t=0}^{T}\ln P(y|x,k,h,\sigma,\theta)$$

(12)

where T is the data length, and P is the probability of generating the recorded output y

$$P(y|x,k,h,\sigma,\theta)=\begin{cases} \text{Prob}(w \geq \theta|x,k,h,\sigma,\theta) & \text{when } y=1 \\ \text{Prob}(w<\theta|x,k,h,\sigma,\theta) & \text{when } y=0. \end{cases}$$

(13)

Since ε is assumed to be Gaussian, the conditional firing probability intensity function P_f (the conditional probability of generating a spike, i.e., Prob($w \geq \theta|x, k, h, \sigma, \theta$) in (13)) at time t can be calculated with the Gaussian error function (integral of Gaussian function) *erf*

$$P_f(t)=0.5 - 0.5erf\left(\frac{\theta - u(t) - a(t)}{\sqrt{2}\sigma}\right)$$

(14)

where

$$erf(s)=\frac{2}{\sqrt{\pi}}\int_0^s e^{-t^2}\, dt.$$

(15)

P at time t then can be calculated as

$$P(t)=\begin{cases} P_f(t) & \text{when } y=1 \\ 1 - P_f(t) & \text{when } y=0 \end{cases}$$

(16)

or,

$$P(t)=0.5 - [y(t) - 0.5]erf\left(\frac{\theta - u(t) - a(t)}{\sqrt{2}\sigma}\right).$$

(17)

Model coefficients c then can be estimated by minimizing the negative log-likelihood function L

$$\tilde{c} = \arg\min(L(c)).$$

(18)

It is shown that this model is equivalent to a generalized linear model (GLM) [50], [51] with inputs and preceding output structured with Volterra models [46], [47]. For this reason, this model can be termed as generalized Volterra model (GVM) [47], [52]. Note that u, a, and n are dimensionless variables, so without loss of generality, σ and θ can be set to 1 in estimation, and later restored from the estimated coefficients.

The second step of model estimation involves the selection of optimal subsets of model coefficients. Mathematically, this step is necessary for further reducing the number of model coefficients to avoid overfitting. More importantly, this step identifies the significant inputs (represented by the first- and second-order self-kernels) and nonlinear interactions between inputs (represented by the second-order cross-kernels) of each output neuron and results in more interpretable models [47]. For a given output neuron, the selected input neurons are the

ones that have functional connections to the output neuron; the selected (second) cross-kernels indicate the pairs of inputs that exhibit nonlinear summations in the synaptic potential (u) of the output neuron. The statistical model selection procedure involves a forward step-wise model selection method [53] and a cross-validation method that have been described previously [47].

3) Kernel Reconstruction and Interpretation—The model coefficients $\hat{c}$ and $\hat{\sigma}$ can be obtained from the estimated Laguerre expansion coefficients, $\tilde{c}$, as in

$$\widehat{c}_0=0, \ \widehat{c}_1^{(n)}=\frac{\tilde{c}_1^{(n)}}{1-\tilde{c}_0}, \ \widehat{c}_{2s}^{(n)}=\frac{\tilde{c}_{2s}^{(n)}}{1-\tilde{c}_0},$$
$$\widehat{c}_{2x}^{(n_1,n_2)}=\frac{\tilde{c}_{2x}^{(n_1,n_2)}}{1-\tilde{c}_0}, \ \tilde{c}_h=\frac{\tilde{c}_h}{1-\tilde{c}_0}, \ \tilde{\sigma}=\frac{1}{1-\tilde{c}_0}.$$

(19-24)

Feedforward and feedback kernels are then reconstructed as

$$\widehat{k}_0=0,$$
$$\widehat{k}_1^{(n)}(\tau)=\sum_{j=1}^{L}\widehat{c}_1^{(n)}(j)b_j(\tau),$$
$$\widehat{k}_{2s}^{(n)}(\tau_1,\tau_2)=\sum_{j_1=1}^{L}\sum_{j_2=1}^{j_1}\frac{\widehat{c}_{2s}^{(n)}(j_1,j_2)}{2}\times[b_{j_1}(\tau_1)b_{j_2}(\tau_2)+b_{j_2}(\tau_1)b_{j_1}(\tau_2)],$$
$$\widehat{k}_{2x}^{(n_1,n_2)}(\tau_1,\tau_2)=\sum_{j_1=1}^{L}\sum_{j_2=1}^{L}\widehat{c}_{2x}^{(n_1,n_2)}(j_1,j_2)b_{j_1}(\tau_1)b_{j_2}(\tau_2),$$
$$\widehat{h}(\tau)=\sum_{j=1}^{L}\widehat{c}_h(j)b_j(\tau).$$

(25-29)

Threshold θ is equal to one.

The normalized kernels provide an intuitive representation of the system input–output nonlinear dynamics. Single-pulse and paired-pulse response functions (r_1 and r_2) of each input can be derived as [20], [47]

$$r_1^{(n)}(\tau)=\widehat{k}_1^{(n)}(\tau)+\widehat{k}_{2s}^{(n)}(\tau,\tau), \text{ and}$$
$$r_2^{(n)}(\tau_1,\tau_2)=2\widehat{k}_{2s}^{(n)}(\tau_1,\tau_2)$$

(30-31)

$r_1^{(n)}$ is simply the PSP elicited by a single spike from the nth input neuron; $r_2^{(n)}$ describes the nonlinear effect of pairs of spikes from the nth input neuron that is different from the simple summation of their single PSPs, i.e., $r_1^{(n)}(\tau_1)+r_1^{(n)}(\tau_2)$. $\widehat{k}_{2x}^{(n_1,n_2)}(\tau_1,\tau_2)$ represents the nonlinear effect of pairs of spikes with one spike from neuron n_1 and one spike neuron n_2. h represents the output spike-triggered after-potential (Fig. 6).

4) Model Validation and Prediction—The cross-validation procedure in model selection guarantees the resulting model to have predictive power over novel datasets since the out-of-sample likelihood function has to be decreased during model selection [47]. Selected inputs/cross-terms and estimated parameters/coefficients can be readily used to make further inferences about the functional connectivity and neuronal dynamics as shown

in the previous section. However, one also needs to evaluate quantitatively the goodness-of-fit of the model. One way of doing this is to evaluate the continuous firing probability intensity predicted by the model with the recorded output spike train. According to the time-rescaling theorem, an accurate model should generate a conditional firing intensity function P_f that can rescale the recorded output spike train into a Poisson process with unit rate [54]. By further variable conversion, interspike intervals should be rescaled into independent uniform random variables on the interval (0, 1). The model goodness-of-fit then can be assessed with a Kolmogorov-Smirnov (KS) test. If the model is correct, all points should lie closely to (e.g., within the 95% confidence bounds) the 45-degree line of the KS plot. Another way is to quantify the similarity between the recorded output spike train y and the predicted output spike train $\hat{y}$ after a smoothing process. First, $\hat{y}$ is realized through simulation. Secondly, $\hat{y}$ and y are convolved with a Gaussian kernel and then compared by calculating their correlation coefficient [46].

5) Application to Hippocampal CA3-CA1 Dynamics—This method has been successfully implemented in the modeling of hippocampal CA3-CA1 dynamics [46], [47]. In the hippocampus, CA1 pyramidal neurons are primarily driven by CA3 pyramidal cells. Output of the CA1 region thus can be considered a nonlinear transformation of the CA3 spike trains. In Drs. Sam Deadwyler and Hampson's laboratories at Wake Forest University, rats are trained to perform a memory-dependent behavioral task—delayed nonmatch-to-sample task. CA3 and CA1 spike trains are simultaneously recorded when the rats are performing the task, and then used to build the MIMO model (Fig. 7). Results show that the MIMO model can be reliably estimated from the CA3 and CA1 spike trains. The model: a) accurately (but stochastically) predicts the CA1 spatiotemporal pattern based on the CA3 spatiotemporal pattern (Fig. 8); b) provides intuitive representations of the CA3-CA1 transfer function in means of feedforward kernels, feedback kernels, noise standard deviation; and c) reveals the functional CA3-CA1 connectivity with its significant model terms (see [46], [47] for more details).

B. Nonstationary Modeling of Brain Regions

Our modeling approach also must deal with the non-stationarities of hippocampal regions. In a nonstationary (time-varying) system, the input–output transfer function depends also on time (32). The modeling goal is to track the emergence and evolution of the MIMO nonlinear dynamics during learning and memory formation.

$$S(t){:}X \rightarrow Y. \tag{32}$$

1) Estimating a Time-Varying MIMO Model—We have formulated a nonstationary modeling methodology for the above-described model structure using a point-process adaptive filtering framework [55]. In this approach, model coefficients (c) are taken as state variables while the input–output spikes are taken as observable variables. Using adaptive filtering methods, state variables can be recursively updated as the observable variables unfold in time. The underlying change of system input–output properties then is represented by the time-varying Volterra kernels ($k(t)$ and $h(t)$) reconstructed with the time-varying coefficients ($c(t)$).

Firstly, the probability of observing an output spike at time t, i.e., $P_f(t)$, is predicted by the GVM at time $t − 1$ based on the inputs up to t and output before t (14). Secondly, the difference between $P_f(t)$ and the new observation of output $y(t)$ is used to correct the GVM model coefficients. Using the stochastic state point process filtering algorithm [56],

coefficient vector $C(t)$ and its covariance matrix $W(t)$ are both updated iteratively at each time step t

$$W(t)^{-1}=[W(t-1)+Q]^{-1}+\left[\left(\frac{\partial \log P_f(t)}{\partial C}\right)^T P_f(t)\left(\frac{\partial \log P_f(t)}{\partial C}\right)-\left(y(t)-P_f(t)\right)\frac{\partial^2 \log P_f(t)}{\partial C \partial C^T}\right]_{C(t-1)}$$ (33)

$$C(t)=C(t-1)+W(t)\times\left[\left(\frac{\partial \log P_f(t)}{\partial C}\right)^T\left(y(t)-P_f(t)\right)\right]_{C(t-1)}$$ (34)

where Q is the coefficient noise covariance matrix. Including W as the "learning rate" allows reliable and rapid tracking of the model coefficients C representing the system nonlinear dynamics.

2) Simulation Studies—We have intensively tested this nonstationary algorithm with synthetic input–output spike train data obtained through simulations [55]. The tested systems have various model structures involving different model orders, e.g., first and second order (including self-and cross-kernels). The number of system inputs ranges from moderate to larger scale (e.g., 32-input), which matches the maximal number of available units in a typical experimental dataset. The system nonstationarity to be tracked takes a variety of forms such as: a) step (jump) change; b) linear change; and c) LTP/LTD-like changes. Results show that the nonstationary algorithm can reliably and accurately track the underlying system nonstationarities and represents them in the time-varying Volterra kernels (see Fig. 9 for a second-order, two-input, step-change example). In all cases, the estimated kernels converge rapidly (with a 10–100 s timescale) to the target kernels without interfering each other.

C. Identification of the Learning Rule

The nonstationarity in the transfer function of a given brain region is determined by the experiences of the animal. In the brain region, the experiences are internally represented as the flow of the input/output spatiotemporal patterns of spike trains. A fundamental question to ask is whether it is possible to reconstruct the nonstationarity of the transfer function of a brain region, which is characterized in Step 2, using its input–output spike trains and a learning rule defining how to modify the transfer function based on the input–output spike trains (Fig. 4 right). Such a learning rule is critical for understanding the underlying mechanisms of cognition processes, e.g., Hebbian-like synaptic modification during learning and memory formation

$$L:X, Y \rightarrow S.$$ (35)

We postulate to conduct mathematical analyses and computer simulations of neuronal network nonlinear, nonstationary dynamics to identify such potential learning rules. As a first step, a neuronal network model can be built and initialized based on the MIMO nonlinear dynamics identified from naïve animals. The functional connections between input and output neurons will be determined based on the feedforward kernels. The spike-dependent intrinsic properties of the output neurons will be determined by the feedback kernels. In the next step, learning processes in the brain region will be simulated by feeding

the network model with the input sequence recorded from the modeled brain region during learning. The input–output transfer function *S*(*t*) will be updated by the input and output patterns following a learning rule *L*. Finally, the learning rule is substantiated through mathematical analyses, and the associated parameters are optimized so that the emergence and changes of the transfer function characterized in Step 2 can be replicated in the simulation. The candidate learning rules include the following.

1. Input/output frequency-dependent learning. This learning rule mimics the classical Bienenstock-Cooper-Munro (BCM) form of synaptic modification rules, in which the changes of the transfer function depend only on the frequencies of the input/ output spike trains [57].

2. Input/output timing-dependent learning. This learning rule mimics the spike-timing-dependent plasticity. Changes of the transfer function depend on the timings of both input and output spikes [58].

3. Input/output pattern-dependent learning. This will be a general form of learning rule, in which changes in the transfer function are determined by the spatiotemporal patterns of the input/output spikes [59]. Interactions between multiple input/output spikes will be explicitly included and analyzed.

We expect the final outcome of this step to be a generative model of the identified nonstationarities of the hippocampal population nonlinear dynamics.

III. CONCLUSIONS AND DISCUSSION

In this paper, we have dealt with the issue of the neurobiological bases of cognition. More specifically, we have argued that nonlinear input–output properties of populations of neurons are potential neurobiological indices of cognitive processing. We have demonstrated both here and previously [14], [20], [21], [23], [24], [26], [27], [46], [47], [60] that nonlinear input–output properties of single neurons and populations of neurons can be measured experimentally (electrophysiologically), and for mathematical modeling, can be readily incorporated within a theoretical framework of nonlinear systems identification. We also have presented here some of the most recent methodological advances in nonlinear systems modeling that provide the critical capabilities for achieving systems-level descriptors of neural function—systems-level descriptors that can be proposed and investigated as potential correlates of cognitive function. These new methodologies allow input–output properties to be defined in the context of high-order nonlinearities, nonstationarities (synaptic plasticity) of nonlinearities, and population, ensemble coding of neural information.

Before discussing these concepts and approaches in the context of the cognitive function of the hippocampus, it is important to state some assumptions. First, we assume that cognitive functions reflect the highest levels of neural function, i.e., neural operations that involve entire systems of neurons. For example, the cognitive function of creating new long-term memories from existing short-term memories is performed by the hippocampal formation, which is a collection of cortical neural structures consisting of the entorhinal cortex, the dentate gyrus, the CA3 pyramidal cell region (the regio inferior of the hippocampus), the CA1 pyramidal cell region (the regio superior of the hippocampus), and the subiculum [5], [61]. The hippocampus proper—dentate, CA3, and CA1—is considered the "intrinsic trisynaptic pathway" of the hippocampus, and is the minimum circuitry involved in the short-term memory to long-term memory transformation. Second, we assume that the collective functional properties of the dentate, CA3, and CA1, when combined together, are equivalent to the cognitive function of "long-term memory formation." Third, we assume that the functional properties of the components of the hippocampus proper identified above,

and for that matter most any brain region, can be assessed as "input–output properties," i.e., the manner in which incoming signals are processed into different, outgoing signals. At a neural level, the composite input–output properties of the major, intrinsic pathways of a brain region *are* its function. When the kernels are estimated accurately for the appropriate order nonlinearity, and for neural data generated under "natural" conditions, the kernels: 1) describe how the neural system, at each one of its major layers or subsystems, responds to the range of input signals associated with the set of behaviors and/or cognitive states of interest; 2) describe how neural correlates of the behavior of interest (#1) are transformed from the system input to the system output, and at each of its major layers or subsystems; and 3) allow prediction of system and subsystem output for a wide range of activity conditions.

Clinical studies conducted over the last 60 years have clarified that the hippocampal formation is responsible for long-term memory formation [4], [62], [63]. The hippocampal system does not store memories itself, but instead, re-encode short-term memory so that information is compatible with existing long-term memory. Precisely what "compatibility" means remains unknown, but as an example, compatibility might mean that appropriate first-order associations for a given episodic memory had been identified. Long-term memory is stored in a distributed manner, probably throughout neocortex. With the hippocampus defined as the set of brain systems above, "long-term memory formation" must be equivalent to the total re-encoding process performed as inputs propagate from the dentate gyrus to the CA1 region. How can this "re-encoding process" be assessed and understood? As stated above, and as demonstrated in previous sections of this paper, we assume that the functional properties of any network of neurons (or for that matter, any neuron, or any neuron component, e.g., channel, etc.) can be represented in terms of its input–output properties, or in this case, its nonlinear multiple-input, multiple-output properties. Given the arguments made earlier, and from data described above, it is our position that neurons should be conceived of as nonlinear dynamical processing elements. Because of the inherent nonlinear properties of hippocampal neurons and the nonlinearities inherent in the processes of synaptic transmission, input spatiotemporal patterns of spike train activity are transformed into different, output spatiotemporal patterns of spike train activity. The nature and degree of nonlinear transformation will almost certainly vary as a function of hippocampal region because of differences in principal cell morphology, and/or intrinsic conductances (e.g., distribution, type of active channels), and/or local circuitry. Nonetheless, as activity propagates from the entorhinal cortex to the subiculum, each layer of the hippocampus (dentate gyrus, CA3, and CA1) progressively reencodes short-term memory representations into long-term memory representations.

The total re-encoding process whereby short-term memories become long-term memories can be assessed experimentally, and modeled mathematically, in the manner demonstrated previously with regard to multi-input, multioutput properties of the CA3-CA1 hippocampal system. If the same analyses were performed for the entorhinal-dentate and the dentate-CA3 subsystems of the hippocampus, then computer simulations of the functioning of all three subsystems of the hippocampus would be attainable. We have investigated previously the possibility of analytical solutions to the combination of subsystem nonlinear characterizations to achieve larger nonlinear system input–output models, and *vice versa* for system decomposition, but these studies were of single-input, single-output cases only [64]–[66]. Experimental verification of such a simulated model of the functioning of the intrinsic, hippocampal "trisynaptic pathway," though difficult, is possible (the requirement would be simultaneous recordings of neural activity from two sites in the hippocampal formation separated by two or more synapses, e.g., layer II of the entorhinal cortex and the CA1 pyramidal cell region).

Considering all of the above, we believe it is experimentally and theoretically feasible to characterize each of the subregions of the hippocampus: dentate, CA3, CA1, and then to integrate the dynamics of each layer into a model of the intrinsic, trisynaptic pathway of the hippocampal system, though we are a long way from demonstrating this. The nonlinear transformations of the entire circuit should be equal to the total nonlinear transformations required to convert short-term memory into long-term memory—though this also is an example of an hypothesis that should be tested by such a combined theoretical-experimental approach. The meaning of the transformations of any one layer is unknown, and again, this identifies an important area of future study. Clinical and experimental animal studies have provided compelling clues as to the function of the entire hippocampus, but we have only a few hypotheses as the functional role of each hippocampal subsystem. Input–output studies of each individual component of the hippocampus will provide quantifications of the properties of each of the dentate, CA3, and CA1 fields, and in the process, also provide hints as to subsystem function to which we previously have not had access. The major point, however, is that a combined theoretical-experimental path can be defined for achieving a biologically based, animal model of a highly important cognitive function—long-term memory formation.

The relevance of this approach to neural prostheses is that it follows from the positions argued here that it may be possible for the complexities of higher brain processing related to concept formation, representations, hierarchically organized associations, etc., and potentially even consciousness, i.e., the brain functions least understood in neural terms at present, and most difficult to repair following brain damage, to be represented mathematically as a set of kernels. We have presented an example of such a characterization with modeling of the CA3-CA1 transformation contribution to long-term memory. Such a set of kernels could even be parameterized for context, for example, for the sleep-wake cycle, and as we have shown in previous work, can be reduced to hardware circuitry. What is remarkable about a kernel-based model, in addition to the attributes identified above, is the degree of "compactness" of the input–output relation: all of the mechanisms underlying the highly nonlinear behavior of hippocampal (or other class) neurons, including the contribution of interneurons, and notably, the contribution of unknown mechanisms yet to be discovered, are included in the model, and as shown here, the model in turn can accurately predict system output to arbitrary input patterns. This is a major advantage of our approach compared to, for example, linear or low-order nonlinear models that are the bases of neural prostheses to replace lost upper extremity functionality. We are in the process of testing the hypothesis that kernel functions for the hippocampus can interact with the endogenous tissue to reinstate normal long-term memory capability after hippocampal dysfunction has been induced experimentally. If successful, this experimental-modeling work will lay the foundation for a general strategy to develop neural pros-theses for any one of multiple cognitive functions. Given the availability of such models, additional research investigating the nonlinear transformations of a given brain region with the purported cognitive functions of the same neural system could provide substantial insights into the relations between neural and cognitive dynamics.

Acknowledgments

References

1. Marmarelis, VZ. Nonlinear Dynamic Modeling of Physiological Systems. Hoboken: Wiley-IEEE Press; 2004.

2. Marmarelis, VZ.; Marmarelis, PZ. Analysis of Physiological Systems: The White-Noise Approach. New York: Plenum; 1978.

3. Casti, JL. Nonlinear System Theory. New York: Academic; 1985.

4. Milner, B. Memory and the medial temporal regions of the brain. In: Pribram, KH.; Broadbent, DE., editors. Biology of Memory. New York: Academic; 1970. p. 29-50.

5. Ramon y Cajal, S. The Structure of Ammon's Horn. Springfield, IL: Charles C. Thomas; 1968.

6. Johnston, D.; Wu, SM. Foundations of Cellular Neurophysiology. Cambridge, MA: MIT Press; 1995.

7. Hille, B. Ionic Channels of Excitable Membranes. Sunderland, MA: Sinauer Assoc; 1992.

8. Chauvet GA. Hierarchical functional organization of formal biological systems: A dynamical approach. III. The concept of non-locality leads to a field theory describing the dynamics at each level of organization of the (D-FBS) sub-system. Philos Trans Roy Soc B, Biol Sci 1993;339:463–481.

9. Chauvet, GA. Theoretical Systems in Biology: Hierarchical and Functional Integration. Vol. 3. Oxford, U.K: Pergamon; 1996.

10. Chauvet GA. An n-level field theory of biological neural networks. J Math Biol 1993;31:771–795. [PubMed: 8263424]

11. Chauvet GA. S-propagators: A formalism for the hierarchical organization of physiological systems. Application to the nervous and the respiratory systems. Int J Gen Syst 1999;28:53–96.

12. Chauvet GA. On the mathematical integration of the nervous tissue based on the S-propagator formalism: I. Theory. J Integr Neurosci 2002;1:31–68. [PubMed: 15011264]

13. Chauvet GA, Berger TW. Hierarchical model of the population dynamics of hippocampal dentate granule cells. Hippocampus 2002;12:698–712. [PubMed: 12440584]

14. Berger TW, Eriksson JL, Ciarolla DA, Sclabassi RJ. Nonlinear systems analysis of the hippocampal perforant path-dentate projection. II. Effects of random impulse train stimulation. J Neurophysiol 1988;60:1076–1094. [PubMed: 3171657]

15. Marmarelis VZ, Berger TW. General methodology for nonlinear modeling of neural systems with Poisson point-process inputs. Math Biosci 2005;196:1–13. [PubMed: 15963534]

16. Volterra, V. Theory of Functionals and of Integral and Integro-Differential Equations. New York: Dover; 1959.

17. Wiener, N. Nonlinear Problems in Random Theory. New York: Technol. Press MIT/Wiley; 1958.

18. Zucker RS, Regehr WG. Short-term synaptic plasticity. Annu Rev Physiol 2002;64:355–405. [PubMed: 11826273]

19. Erler F, Meyer-Hermann M, Soff G. A quantitative model for presynaptic free Ca2+ dynamics during different stimulation protocols. Neurocomputing 2004;61:169–191.

20. Song D, Marmarelis VZ, Berger TW. Parametric and non-parametric modeling of short-term synaptic plasticity. Part I: Computational study. J Comput Neurosci Feb;2009 26:1–19. [PubMed: 18506609]

21. Song D, Wang Z, Marmarelis VZ, Berger TW. Parametric and non-parametric modeling of short-term synaptic plasticity. Part II: Experimental study. J Comput Neurosci Feb;2009 26:21–37. [PubMed: 18504530]

22. Berger, TW.; Harty, TP.; Barrionuevo, G.; Sclabassi, RJ. Modeling of neuronal networks through experimental decomposition. In: Marmarelis, VZ., editor. Advanced Methods of Physiological System Modeling. New York: Plenum; 1989.

23. Harty TP, Berger TW, Sclabassi RJ, Barrionuevo G. Nonlinear systems analysis of the in vitro hippocampal dentate gyrus. I. Characterization of granulecell response to perforant path input. submitted for publication.

24. Harty TP, Berger TW, Sclabassi RJ, Barrionuevo G. Nonlinear Systems Analysis of the In Vitro Hippocampal Dentate Gyrus. II. Contribution of $GABA_A$ and $GABA_B$ Receptor Function. Submitted for publication.

25. Grace AA. In vivo and in vitro intracellular recordings from rat midbrain dopamine neurons. Ann NY Acad Sci 1988;537:51–76. [PubMed: 2849358]

26. Berger TW, Eriksson JL, Ciarolla DA, Sclabassi RJ. Nonlinear systems analysis of the hippocampal perforant path-dentate projection. III. Comparison of random train and paired impulse stimulation. J Neurophysiol 1988;60:1095–1109. [PubMed: 3171658]

27. Sclabassi RJ, Eriksson JL, Port RL, Robinson GB, Berger TW. Nonlinear systems analysis of the hippocampal perforant path-dentate projection. I. Theoretical and interpretational considerations. J Neurophysiol 1988;60:1066–1076. [PubMed: 3171656]

28. Puchalla JL, Schneidman E, Harris RA, Berry MJ. Redundancy in the population code of the retina. Neuron May 5;2005 46:493–504. [PubMed: 15882648]

29. Song, D.; Wang, Z.; Marmarelis, VZ.; Berger, TW. A modeling paradigm incorporating parametric and non-parametric methods. Proc. 26th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC 2004); 2004. p. 647-650.

30. Berger, TW.; Sclabassi, RJ. Long-term potentiation and its relation to changes in hippocampal pyramidal cell activity and behavioral learning during classical conditioning. In: Landfield, PW.; Deadwyler, SA., editors. Long-Term Potentiation: From Biophysics to Behavior. New York: Alan R. Liss; 1988. p. 467-497.

31. Robinson GB, Fluharty SJ, Zigmond MJ, Sclabassi RJ, Berger TW. Recovery of hippocampal dentate granule cell responsiveness to entorhinal cortical input following norepinephrine depletion. Brain Res 1993;614:21–28. [PubMed: 7688646]

32. Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA. Cognitive control signals for neural prosthetics. Science Jul 9;2004 305:258–262. [PubMed: 15247483]

33. Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, Kim J, Biggs SJ, Srinivasan MA, Nicolelis MA. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature Nov 16;2000 408:361–365. [PubMed: 11099043]

34. Deadwyler SA, Hampson RE. Ensemble activity and behavior: What's the code?'. Science 1995;270:1316–1318. [PubMed: 7481817]

35. Eichenbaum H, Kuperstein M, Fagan A, Nagode J. Cue-sampling and goal-approach correlates of hippocampal unit-activity in rats performing an odor-discrimination task. J Neurosci Mar;1987 7:716–732. [PubMed: 3559709]

36. Georgopoulos AP, Schwartz AB, Kettner RE. Neuronal population coding of movement direction. Science Sep;1986 233:1416–1419. [PubMed: 3749885]

37. Nicolelis MA. Brain-machine interfaces to restore motor function and probe neural circuits. Nat Rev Neurosci 2003;4:417–422. [PubMed: 12728268]

38. Pouget A, Dayan P, Zemel RS. Inference and computation with population codes. Annu Rev Neurosci 2003;26:381–410. [PubMed: 12704222]

39. Salinas E, Abbott LF. Vector reconstruction from firing rates. J Comput Neurosci Jun;1994 1:89–107. [PubMed: 8792227]

40. Eichenbaum H, Pettijohn D, Deluca AM, Chorover SL. Compact miniature microelectrode-telemetry system. Physiol Behav Jun;1977 18:1175–1178. [PubMed: 928540]

41. Buzsaki G. Large-scale recording of neuronal ensembles. Nat Neurosci May;2004 7:446–451. [PubMed: 15114356]

42. Chapin JK. Using multi-neuron population recordings for neural prosthetics. Nat Neurosci May; 2004 7:452–455. [PubMed: 15114357]

43. Berger TW, Rinaldi PC, Weisz DJ, Thompson RF. Single-unit analysis of different hippocampal cell types during classical conditioning of rabbit nictitating membrane response. J Neurophysiol Nov;1983 50:1197–1219. [PubMed: 6644367]

44. Deadwyler SA, Bunn T, Hampson RE. Hippocampal ensemble activity during spatial delayed-nonmatch-to-sample performance in rats. J Neurosci 1996;16:354–372. [PubMed: 8613802]

45. Brown EN, Kass RE, Mitra PP. Multiple neural spike train data analysis: State-of-the-art and future challenges. Nat Neurosci May;2004 7:456–461. [PubMed: 15114358]

46. Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Nonlinear dynamic modeling of spike train transformations for hippocampal-cortical prostheses. IEEE Trans Biomed Eng Jun;2007 54(6):1053–1066. [PubMed: 17554824]

47. Song D, Chan RHM, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Nonlinear modeling of neural population dynamics for hippocampal prostheses. Neural Netw Nov;2009 22:1340–1351. [PubMed: 19501484]

48. Marmarelis VZ. Identification of nonlinear biological systems using Laguerre expansions of kernels. Ann Biomed Eng Nov./Dec;1993 21:573–589. [PubMed: 8116911]

49. Zanos TP, Courellis SH, Berger TW, Hampson RE, Deadwyler SA, Marmarelis VZ. Nonlinear modeling of causal interrelationships in neuronal ensembles. IEEE Trans Neural Syst Rehabil Eng Aug;2008 16:336–352. [PubMed: 18701382]

50. McCullagh, P.; Nelder, JA. Generalized Linear Models. 2. Boca Raton, FL: Chapman & Hall/CRC; 1989.

51. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol Feb;2005 93:1074–1089. [PubMed: 15356183]

52. Song D, Hendrickson P, Marmarelis VZ, Aguayo J, He J, Loeb GE, Berger TW. Predicting EMG with generalized Volterra kernel model. Proc Conf IEEE Eng Med Biol Soc 2008;2008:201–204.

53. Kutner, MH.; Nachtsheim, CJ.; Neter, J.; Li, W. Applied Linear Statistical Models. 5. Boston, MA: McGraw-Hill/Irwin; 2004.

54. Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. Neural Comput 2002;14:325–346. [PubMed: 11802915]

55. Chan RHM, Song D, Berger TW. Tracking temporal evolution of nonlinear dynamics in hippocampus using time-varying Volterra kernels. Proc 30th Annu Int Conf IEEE EMBS 2008:4996–4999.

56. Eden UT, Frank LM, Barbieri R, Solo V, Brown EN. Dynamic analysis of neural encoding by point process adaptive filtering. Neural Comput May;2004 16:971–998. [PubMed: 15070506]

57. Bienenstock EL, Cooper LN, Munro PW. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. J Neurosci Jan;1982 2:32–48. [PubMed: 7054394]

58. Froemke RC, Dan Y. Spike-timing-dependent synaptic modification induced by natural spike trains. Nature Mar 28;2002 416:433–438. [PubMed: 11919633]

59. Bing, L.; Dibazar, A.; Berger, TW. Nonlinear Hebbian learning for noise-independent vehicle sound recognition. Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN 2008) (IEEE World Congr. Comput. Intelligence); 2008. p. 1336-1343.

60. Berger TW, Chauvet G, Sclabassi RJ. A biological based model of functional properties of the hippocampus. Neural Netw 1994;7:1031–1064.

61. Lorente de No R. Studies on the structure of the cerebral cortex. II: Continuation of the study of the Ammonic system. J Psychol Neurol 1934;46:113–177.

62. Eichenbaum H, Dudchenko P, Wood E, Shapiro M, Tanila H. The hippocampus, memory, and place cells: Is it spatial memory or a memory space?'. Neuron 1999;23:209–226. [PubMed: 10399928]

63. Squire LR, Zola-Morgan S. The medial temporal lobe memory system. Science Sep 20;1991 253:1380–1386. [PubMed: 1896849]

64. Chian MT, Marmarelis VZ, Berger TW. Characterization of unobservable neural circuitry in the hippocampus with nonlinear systems analysis. Proc 4th Joint Symp Neural Comput 1997;7:43–50.

65. Chian MT, Marmarelis VZ, Berger TW. Decomposition of neural systems with nonlinear feedback using stimulus-response data. Neurocomputing 1999;26:641–654.

66. Sclabassi, RJ.; Kosanovic, BR.; Barrionuevo, G.; Berger, TW. Computational methods of neuronal network decomposition. In: Marmarelis, VZ., editor. Advanced Methods of Physiological System Modeling. Vol. III. New York: Plenum; 1994. p. 55-86.

Biographies

Theodore W. Berger (Fellow, IEEE) received the Ph.D. degree from Harvard University, Cambridge, MA, in 1976; his thesis work received the James McKeen Cattell Award from the New York Academy of Sciences.

He conducted postdoctoral research at the University of California, Irvine from 1977 to 1978, and was an Alfred P. Sloan Foundation Fellow at the Salk Institute from 1978 to 1979. He joined the Departments of Neuroscience and Psychiatry at the University of Pittsburgh in 1979, being promoted through to Full Professor in 1987. Since 1992, he has been Professor of Biomedical Engineering and Neurobiology at the University of Southern California, and was appointed the David Packard Chair of Engineering in 2003. He became Director of the Center for Neural Engineering in 1997, an organization which helps to unite USC faculty with cross-disciplinary interests in neuroscience, engineering, and medicine. He has published over 170 journal articles and book chapters, and is the coeditor of a book recently published by the MIT Press on *Toward Replacement Parts for the Brain: Implantable Biomimetic Electronics as Neural Prostheses*. His research interests are in the development of biologically realistic, experimentally based, mathematical models of higher brain (hippocampus) function; application of biologically realistic neural network models to real-world signal processing problems; VLSI-based implementations of biologically realistic models of higher brain function; neuron–silicon interfaces for bidirectional communication between brain and VLSI systems; and next-generation brain-implantable, biomimetic signal processing devices for neural prosthetic replacement and/or enhancement of brain function.

Prof. Berger has received a McKnight Foundation Scholar Award, twice received an NIMH Research Scientist Development Award, and was elected a Fellow of the American Association for the Advancement of Science. While at USC, he has received an NIMH Senior Scientist Award, was given the Lockheed Senior Research Award in 1997, was elected a Fellow of the American Institute for Medical and Biological Engineering in 1998, received a Person of the Year "Impact Award" by the AARP in 2004 for his work on neural prostheses, was a National Academy of Sciences International Scientist Lecturer in 2003, and an IEEE Distinguished Lecturer in 2004–2005. He received a "Great Minds, Great Ideas" award from the *EE Times* in 2005, and in 2006 was awarded USC's Associates Award for Creativity in Research and Scholarship.

Dong Song (Member, IEEE) received the B.S. degree in biophysics from the University of Science and Technology of China, Hefei, in 1994 and the Ph.D. degree in biomedical engineering from the University of Southern California (USC), Los Angeles, in 2003.

From 2004 to 2006, he worked as a Postdoctoral Research Associate at the Center for Neural Engineering at USC. He is currently a Research Assistant Professor at the Department of Biomedical Engineering at USC. His main research interests include nonlinear systems analysis of nervous system, cortical neural prosthesis, electro-physiology of hippocampus, long-term and short-term synaptic plasticity, and the development of modeling methods incorporating both parametric and nonparametric modeling techniques.

Prof. Song is a member of Biomedical Engineering Society, American Statistical Association, and Society for Neuroscience.

Rosa H. M. Chan (Student Member, IEEE) received the B.Eng degree in automation and computer-aided engineering from the Chinese University of Hong Kong (CUHK), Hong Kong, in 2003. She is currently working toward the Ph.D. degree in the Department of Biomedical Engineering of University of Southern California.

Her research interest is in the development of cortical neural prosthesis.

Ms. Chan was awarded both the Croucher Scholarship and the Sir Edward Youde Memorial Fellowship for overseas study.

Vasilis Z. Marmarelis (Fellow, IEEE) was born in Mytiline, Greece, on November 16, 1949. He received the Diploma degree in electrical engineering and mechanical engineering from the National Technical University of Athens in 1972 and the M.S. and Ph.D. degrees in engineering science (information science and bioinformation systems) from the California Institute of Technology, Pasadena, in 1973 and 1976, respectively.

After two years of postdoctoral work at the California Institute of Technology, he joined the faculty of Biomedical and Electrical Engineering at the University of Southern California, Los Angeles, where he is currently Professor and Director of the Biomedical Simulations Resource, a research center funded by the National Institutes of Health since 1985 and dedicated to modeling/simulation studies of biomedical systems. He served as Chairman of

the Biomedical Engineering Department from 1990 to 1996. He is coauthor of the book *Analysis of Physiological System: The White Noise Approach* (New York: Plenum, 1978; Russian translation: Moscow, Mir Press, 1981; Chinese translation: Academy of Sciences Press, Beijing, 1990), editor of three research volumes on *Advanced Methods of Physiological System Modeling* (Plenum, 1987, 1989, 1994) and author of a monograph on *Nonlinear Dynamic Modeling of Physiological Systems* (IEEE Press & Wiley Interscience, 2004). He has published more than 100 papers and book chapters in the areas of system modeling and signal analysis. His main research interests are in the areas of nonlinear and nonstationary system identification and modeling, with applications to biology and medicine. Other interests include spatiotemporal and multi-input/multioutput modeling of nonlinear systems, with applications to neural information processing, closed-loop system modeling, and high-resolution 3–D ultrasonic imaging and tissue classification.

Prof. Marmarelis is a fellow of the American Institute for Medical and Biological Engineering.
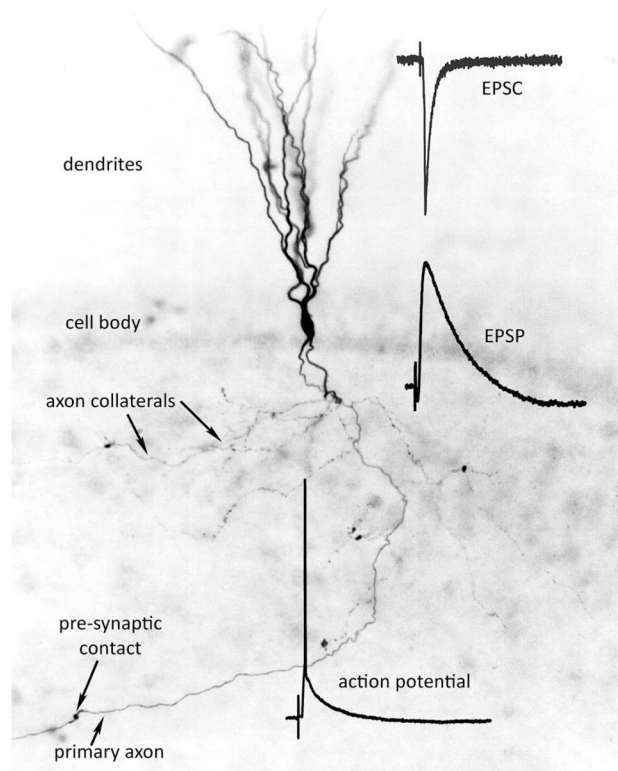
Fig. 1.
Intracellularly filled hippocampal dentate granule cell, providing visualization of most of the anatomical components of a granule cell neuron. Labels identify major aspects of granule cell anatomy. Right: example intracellularly recorded responses showing an excitatory postsynaptic current (EPSC), an excitatory postsynaptic potential (EPSP), and an action potential (AP) generated when magnitude of the EPSP exceeded threshold. See text for more explanation.
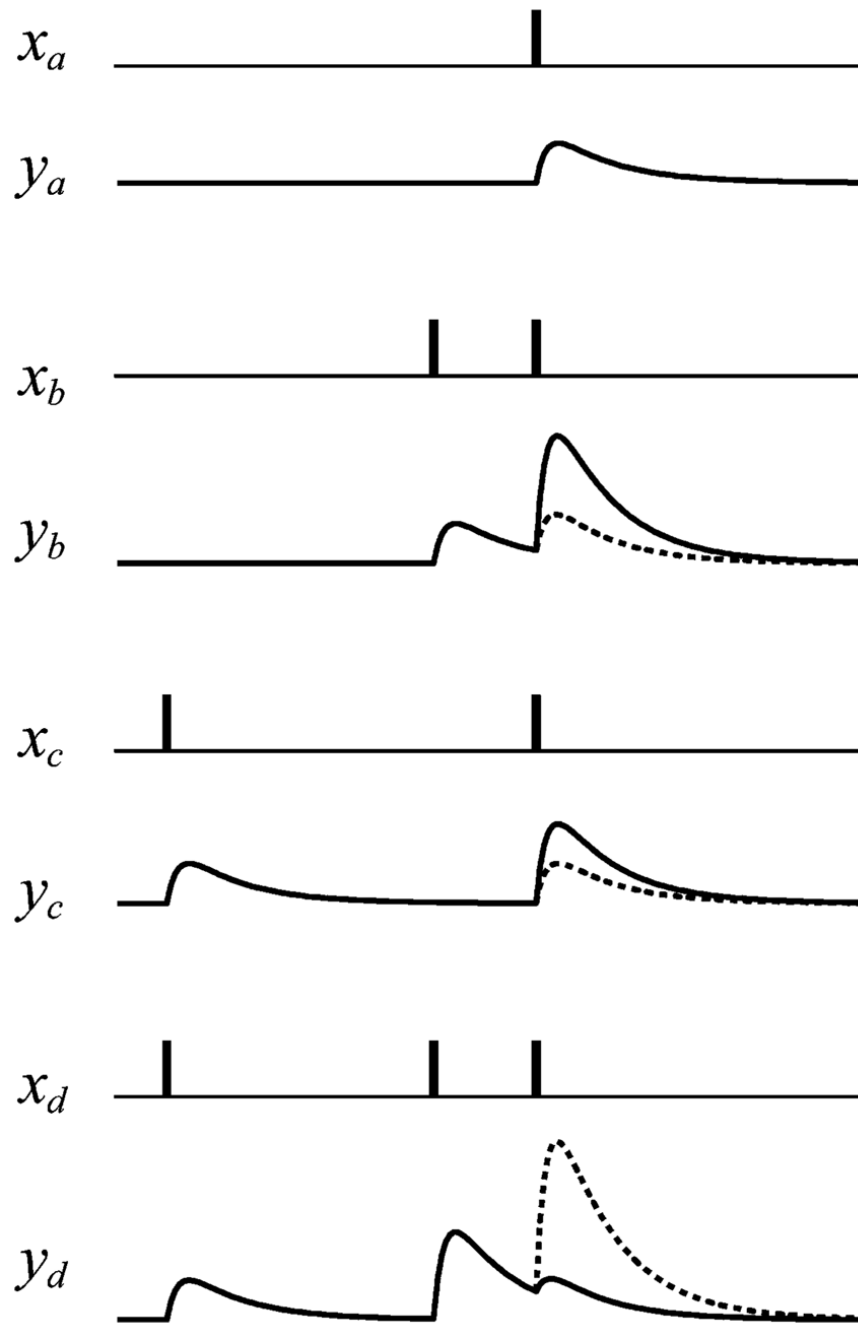
x_a

y_a

x_b

y_b

x_c

y_c

x_d

y_d

Fig. 2.
Hypothetical (based on many, real biological experiments) illustrating typical nonlinear interactions between two and three pulses delivered to excitatory synaptic inputs to hippocampal granule and pyramidal neurons. First (top) trace: single pulse stimulation (x_a) and evoked EPSP (y_a). Second trace: two pulses delivered such that the second pulse is delivered before the response to the first pulse is complete. The dashed EPSP demonstrates the amplitude–time course of the EPSP that would have occurred if the input–output system x_b, y_b had been linear. Because the system is nonlinear, however, a strong facilitation occurs, much larger than predicted based on superposition, with the response to the second pulse being 3–4 times larger than that elicited by the first pulse. The third panel, x_c, y_c,

shows another pair of pulses, with a longer interstimulus interval, that nonetheless also produces facilitation. The fourth (bottom) panel, x_d, x_d illustrates the consequences of combining both pairs of intervals. Again, the dashed line shows the expected response if the two second-order nonlinearities combine in a linear manner. Instead, a strong suppression occurs, revealing a negative third-order nonlinearity. Corresponding applications of quadruplets and quintuplets would uncover an even wider scope of nonlinearities, translating into a rich differential sensitivity to temporal patterns.
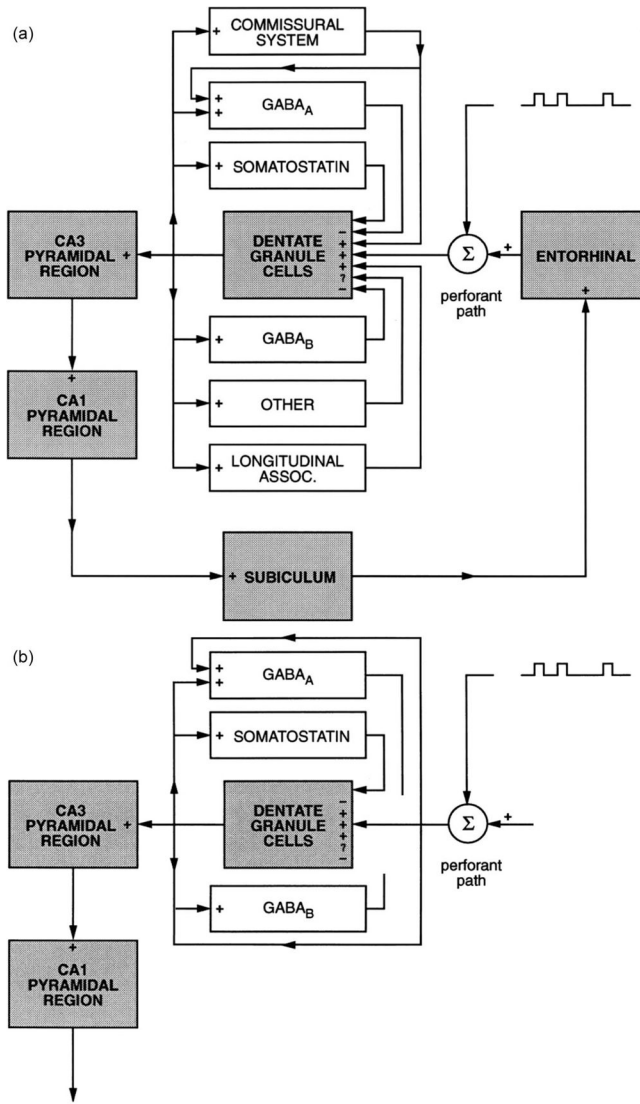
Fig. 3.
Block diagram of most, known anatomical pathways that provide feedforward and feedback regulation of granule cells in response to excitation of input from the entorhinal cortex, and thus are the source of some of the nonlinearities of granule cells, the output neurons of the dentate. (a) Feedforward and feedback pathways in the intact animal. (b) Equivalent representation for the pathways included in the hippocampal slice.
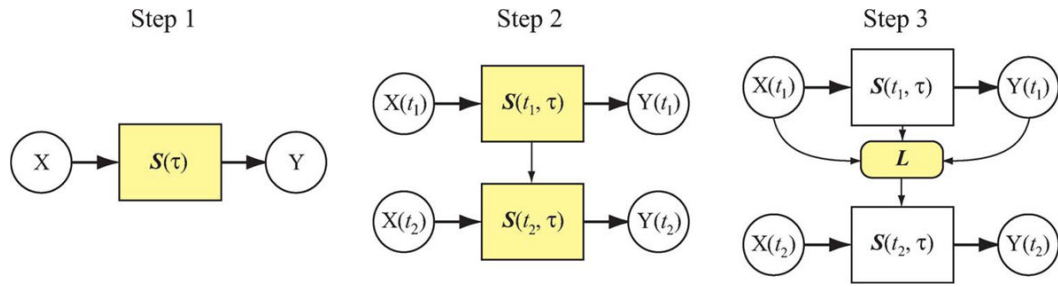
Fig. 4.
Schematic diagram of the three-step modeling strategy. X: input sequences; Y: output sequences; S: transfer functions; L: learning rule for S. In Step 1, S is not a function of time. In Step 2 and 3, S varies with time. During learning, S evolved as a result of input and output activities following the learning rule. Colored boxes indicate the functions need to be identified in each step.
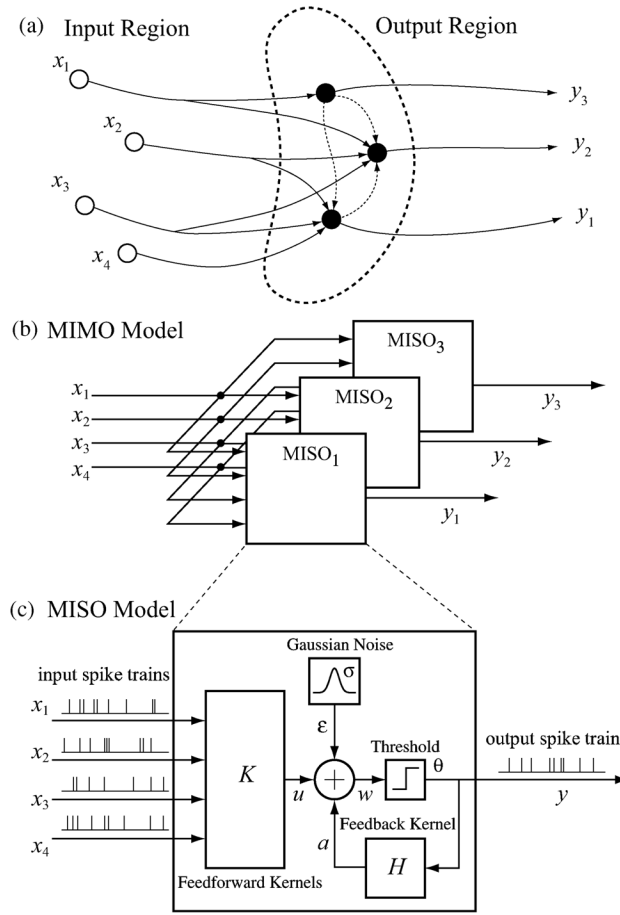
Fig. 5.
MIMO model for population neural dynamics. (a) Schematic diagram of spike train propagation between two brain regions. (b) MIMO model as a series of multiple-input single-output (MISO) models. (c) Structure of a MISO model.
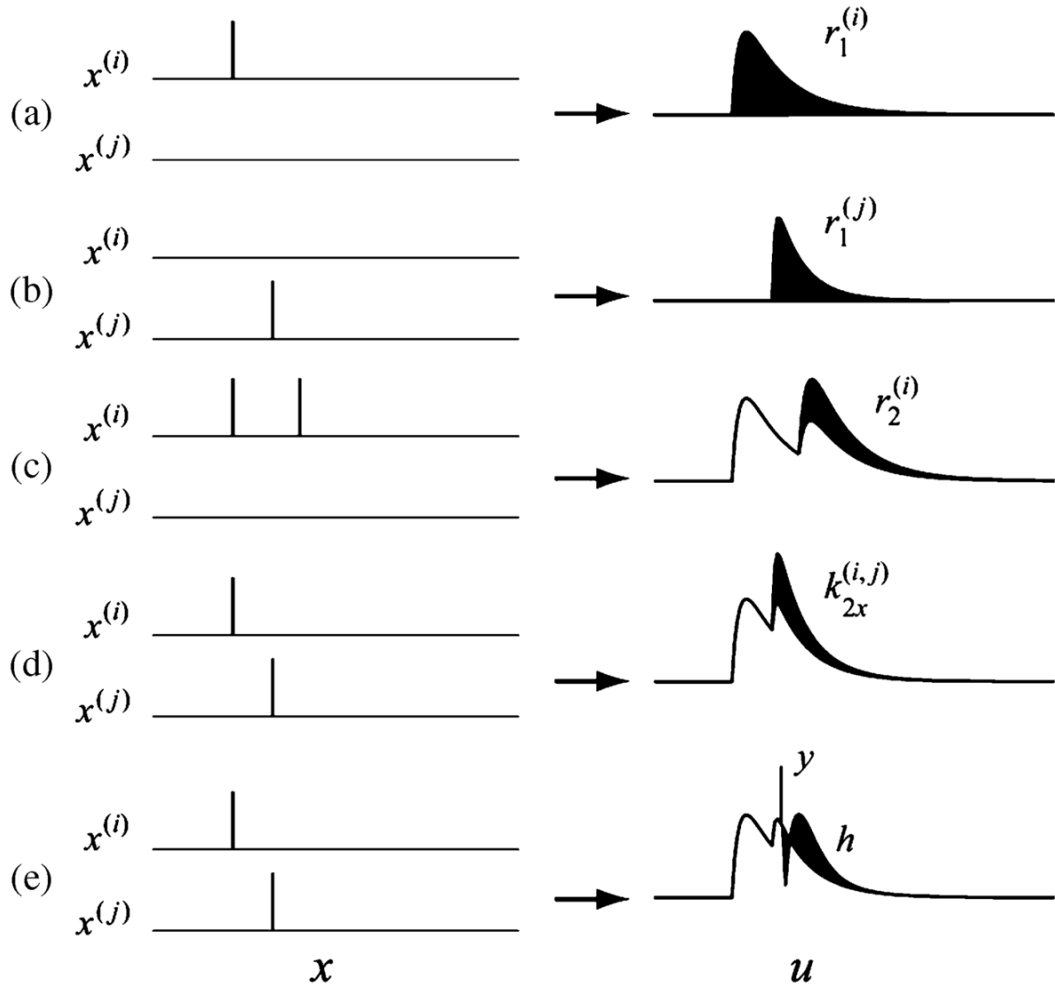
Fig. 6.

Interpretations of the feedforward and feedback kernels. $r_1^{(i)}$ is the response in u elicited by a single spike from the ith input neuron; $r_2^{(i)}$ describes the joint nonlinear effect of pairs of spikes from the ith input neuron in addition to the linear summation of their first-order responses. $k_{2x}^{(i,j)}$ represents the joint nonlinear effect of pairs of spikes from neuron i and j. h represents the output spike-triggered after-potential on u. Black areas: effect of each kernel on u.

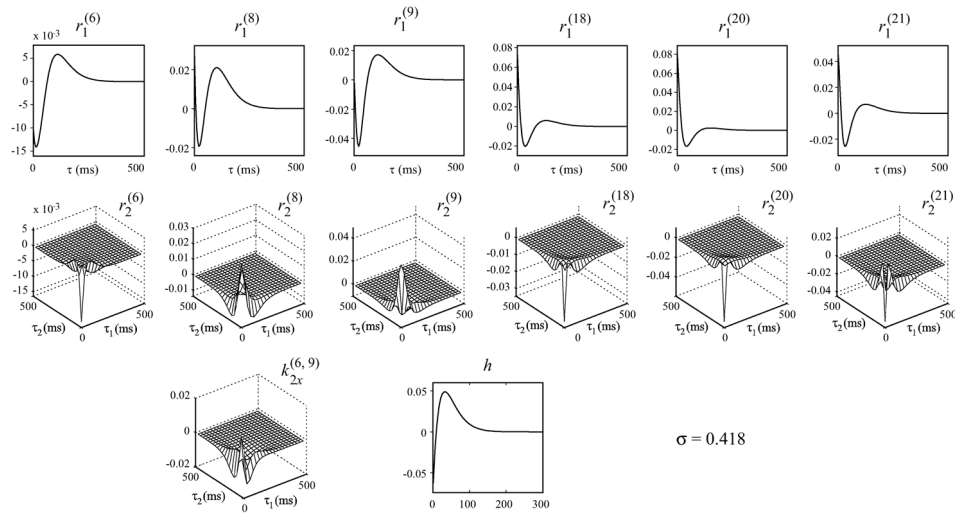**Fig. 7.**
A stationary multi-input single-output (MISO) model of hippocampal CA3-CA1. r_1 are the single-pulse response functions; $k_{2\,s}$ are the paired-pulse response functions for the same input neurons. k_{2x} are the cross-kernels for pairs of neurons. This particular MISO model has six r_1, six r_{2x}, and one k_{2x}.
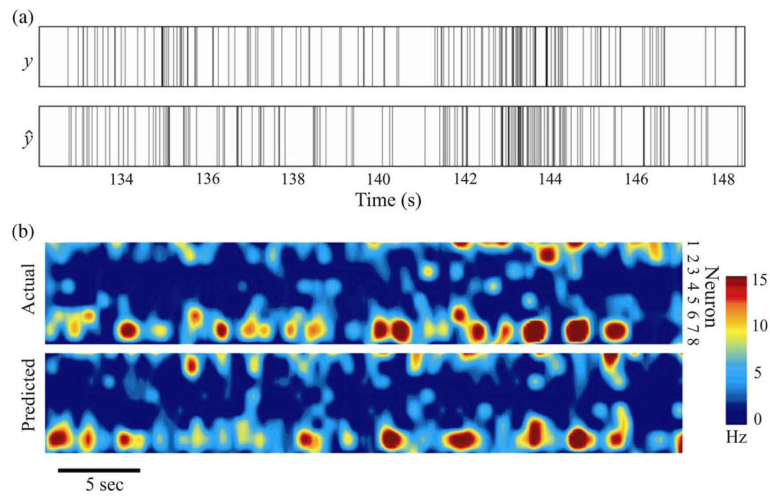
Fig. 8.
Model prediction with MISO and MIMO models. (a) Actual output spike train (top panel) and output spike train predicted by a MISO model (bottom panel). (b) Output spatiotemporal pattern predicted by a MIMO model. (a) and (b) are both out-of-sample results.
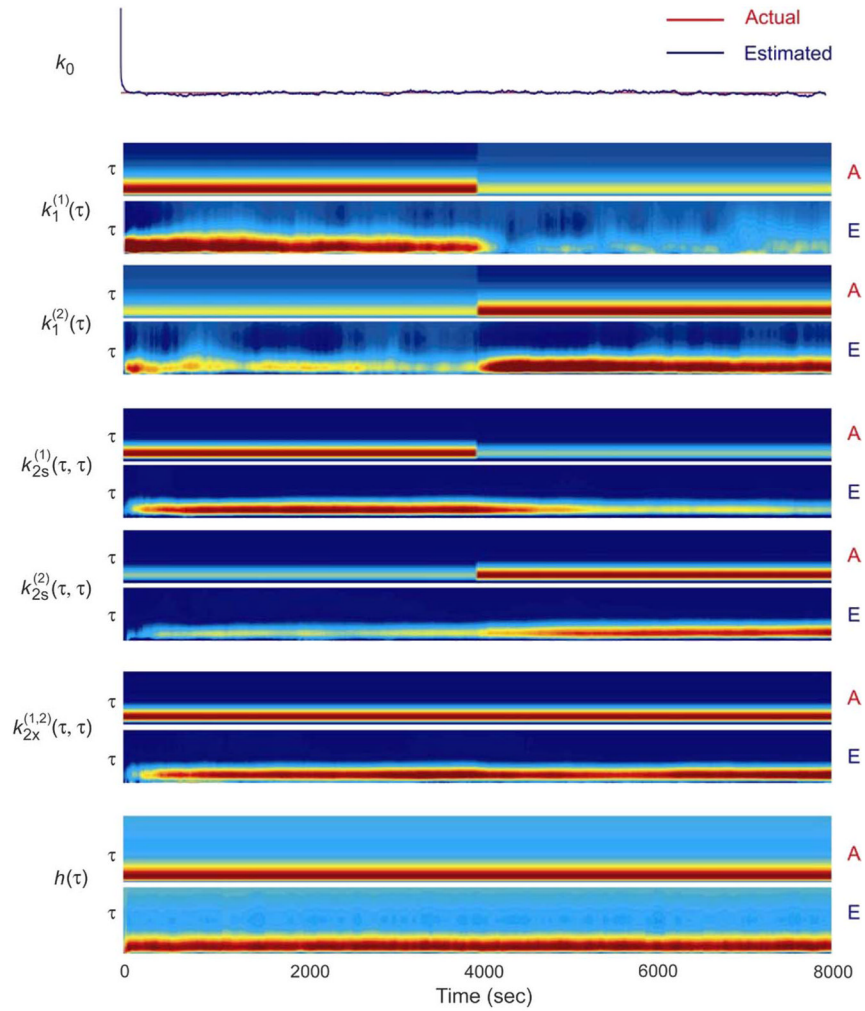
Fig. 9.
A second-order, two-input-single-output system tracked with the nonstationary algorithm. First-order kernels (k_1) and second-order self-kernels (k_{2s}) have step changes at 4000 s. Zeroth-order kernel (k_0), second-order cross-kernels (k_{2x}) and feedback kernel (h) remain constant. Delay time expand (τ) is 500 ms for k_1, k_{2s}, and k_{2x}, 1000 ms for h. The amplitude of kernel is indicated by the color. Only diagonal values of second-order kernels are plotted for simplicity. A: actual kernels; E: estimated kernels.