

Progressively Introducing Quantified Biological Complexity Into a Hippocampal CA3 Model That Learns Trace Conditioning

William B Levy⁽¹⁾, Kai S. Chang⁽¹⁾ and Andrew G. Howe⁽²⁾

(1) *University of Virginia, Department of Neurosurgery, P.O. Box 800420, Charlottesville, VA 22908, USA.*

(2) *Informed Simplifications, LLC, 520 Panorama Road, Earlysville, VA 22936, USA.*

Abstract : Quantifying the performance of a cognitive-behavioral model on a temporal paradigm requires mapping time onto the computational cycles of the simulation. We present a family of four minimal models of the hippocampus CA-3 simulated at different time resolutions. Behavioral results from the hippocampally-dependent trace classical conditioning paradigm show that rabbits can learn to properly anticipate US presentation for a specific range of trace interval time periods. Therefore, our hippocampal model should successfully anticipate US presentation for the same specific range of trace interval durations. Each model attempts to learn two different trace interval lengths. The results reinforce prior findings where we map time into the computational cycles of a minimal model. Further, our results support the the following idea : as the time resolution of a simulation increases, an increasing number of biological processes must be explicitly modeled to maintain behavioral performance for a temporal paradigm.

Progressively Introducing Quantified Biological Complexity Into a Hippocampal CA3 Model That Learns Trace Conditioning

William B Levy⁽¹⁾, Kai S. Chang⁽¹⁾ and Andrew G. Howe⁽²⁾

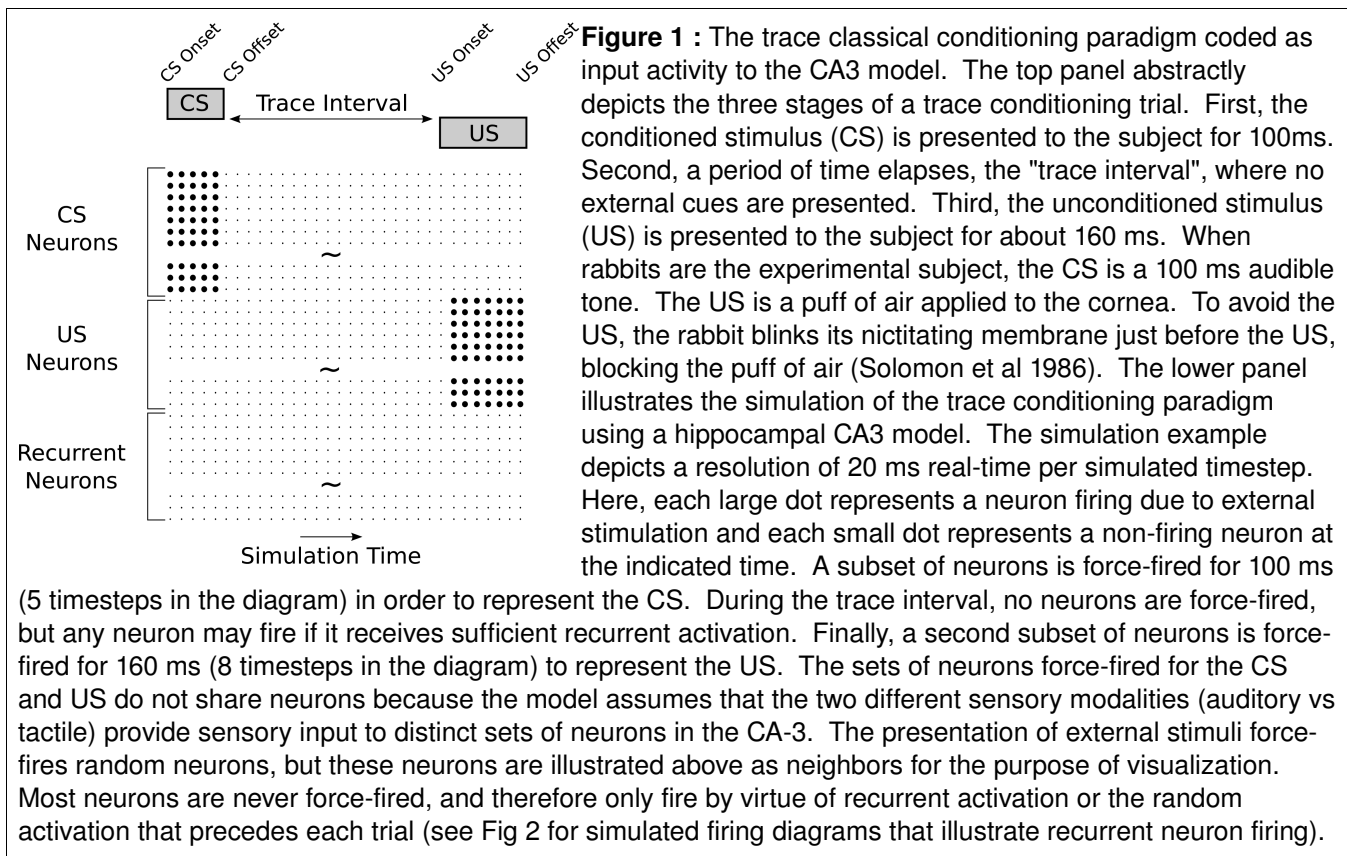
(3) *University of Virginia, Department of Neurosurgery, P.O. Box 800420, Charlottesville, VA 22908, USA.*

(4) *Informed Simplifications, LLC, 520 Panorama Road, Earlysville, VA 22936, USA.*

INTRODUCTION AND METHODS

Minimal modeling can create strong, quantitative conjectures that identify a small number of critical biological features at the microscopic level. "Critical" means that these features are necessary, and are all that is necessary, to reproduce the appropriate macroscopic observations. In the macroscopic case of hippocampally-mediated cognitive function, this has been particularly worthwhile because the hippocampus is required for what appear to be unrelated classes of tasks (see Levy 1996) : temporal (e.g. trace conditioning), spatial (poly-sensory, asynchronous sequences such as spatially-based associations) and abstract association (transitive inference, transitive patterning).

Our minimal modeling studies (summarized in Levy et al 1995, 2005a & Levy 1996) identify a small number of critical functional features - random connectivity, moderate external activation, time-spanning synaptic modification, and moderately well-behaved activity control - that are necessary to reproduce the cognitive-behavioral functions. Identifying these critical functions then leads to meaningful mesoscopic descriptions of CA-3 computations, e.g., random recoding and sequence prediction (Levy 1989, Levy et al 2005a). However, in this same overall context, there are many more questions to ask about microscopic biological details of hippocampal function. Indeed, every observable instance of biochemistry and biophysics that contributes to normal function in region CA-3 is subject to functional conjectures and is a candidate for more detailed scrutiny.



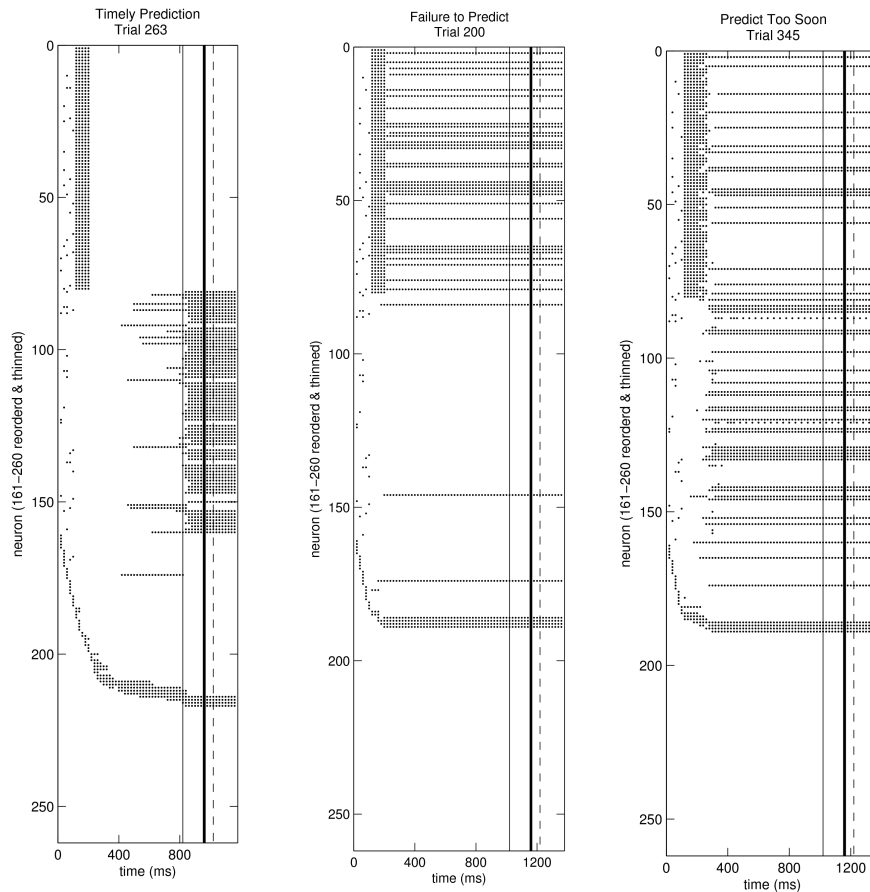


Figure 2 : Raster diagrams of neuron firing during testing (no US presentation, no synaptic modification) for successful learning and two classes of failure to learn. The left panel depicts successful learning and timely prediction of the US. Most US neurons are silent throughout the trace interval period. Successful prediction and avoidance occurs when many US neurons fire a few timesteps before US onset. The middle panel depicts a failure to bridge – the network learns a sequence of firing neurons that do not activate the US neurons. The right panel depicts another failure mode; in this mode, a plurality of US neurons fire much too early for timely prediction. Three vertical lines within each raster diagram partition simulation time relative to US onset. The dotted (right) line indicates US onset. The thick solid line (middle) indicates late prediction; any prediction on or after this timestep is too late for avoidance of the US. The thin solid line (left) indicates the earliest timestep for successful prediction; prediction before this timestep is considered too early. Only a fraction of recurrent neurons are illustrated and are reordered for visualization of sequential firing (see Levy et al 2005b or Howe & Levy 2007 for further explanation).

Here, we describe a logical progression of quantified models for ascertaining the role of individual biological components in producing a hippocampally-dependent, cognitive/ behavioral function. Although not the only strategy, it is our preference to move as methodically as possible, inching from the simple to the complex. Therefore, we start with a successful minimal model and MINIMALLY extend it to determine the biological relevance of microscopic components.

In this preliminary communication, we illustrate a promising sequence for developing model complexity; a sequence that may be applicable beyond region CA-3 of the hippocampus. In the future, we will report the detailed interactions between the components that make up the successively more complex models. Here we present a

sequence of minimally expanded versions of our minimal model. The key idea is that the components of the original model are expanded via gradual enhancement of temporal resolution. However, each enhancement is only acceptable if the model maintains quantitative agreement with a temporally characterized hippocampal paradigm.

A Well-quantified, Hippocampally-dependent Paradigm

The air-puff-to-cornea version of the trace conditioning paradigm is particularly valuable with regard to its hippocampal dependence (no striatum) and temporal sensitivities (Solomon et al 1986, Moyer et al 1990). Both the duration of the conditioned stimulus (CS), and the trace length itself, interactively determine the learnable trace interval. In rabbits (Gormezano et al 1983), and our CA-3 model (Rodriguez & Levy 2000, Howe & Levy 2007), the maximum learnable interval (one in which the subject can produce a successful, defensive eye-blink) is about 700 ms when using a 100 ms CS.

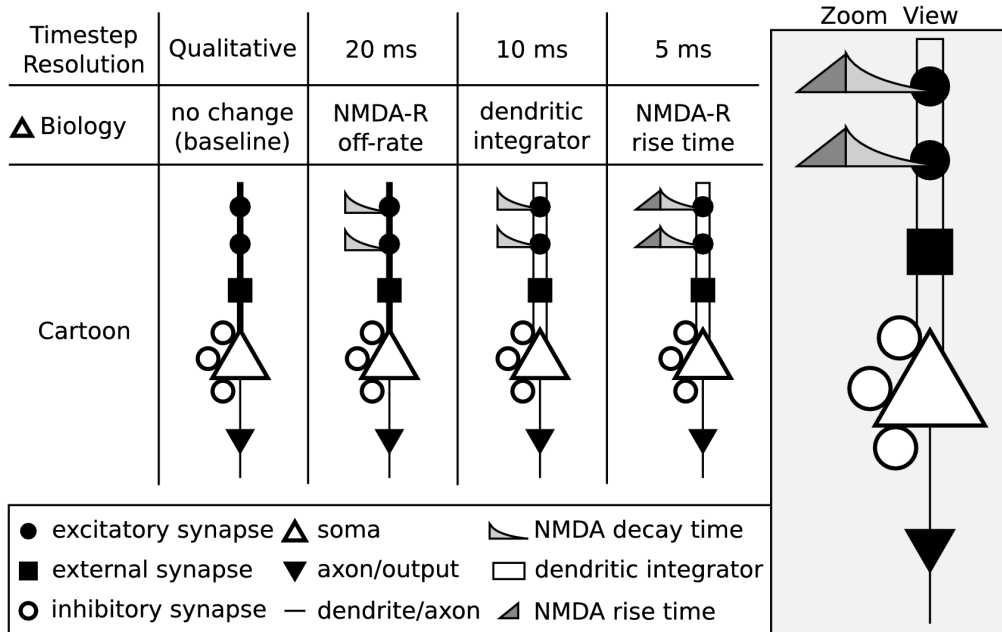
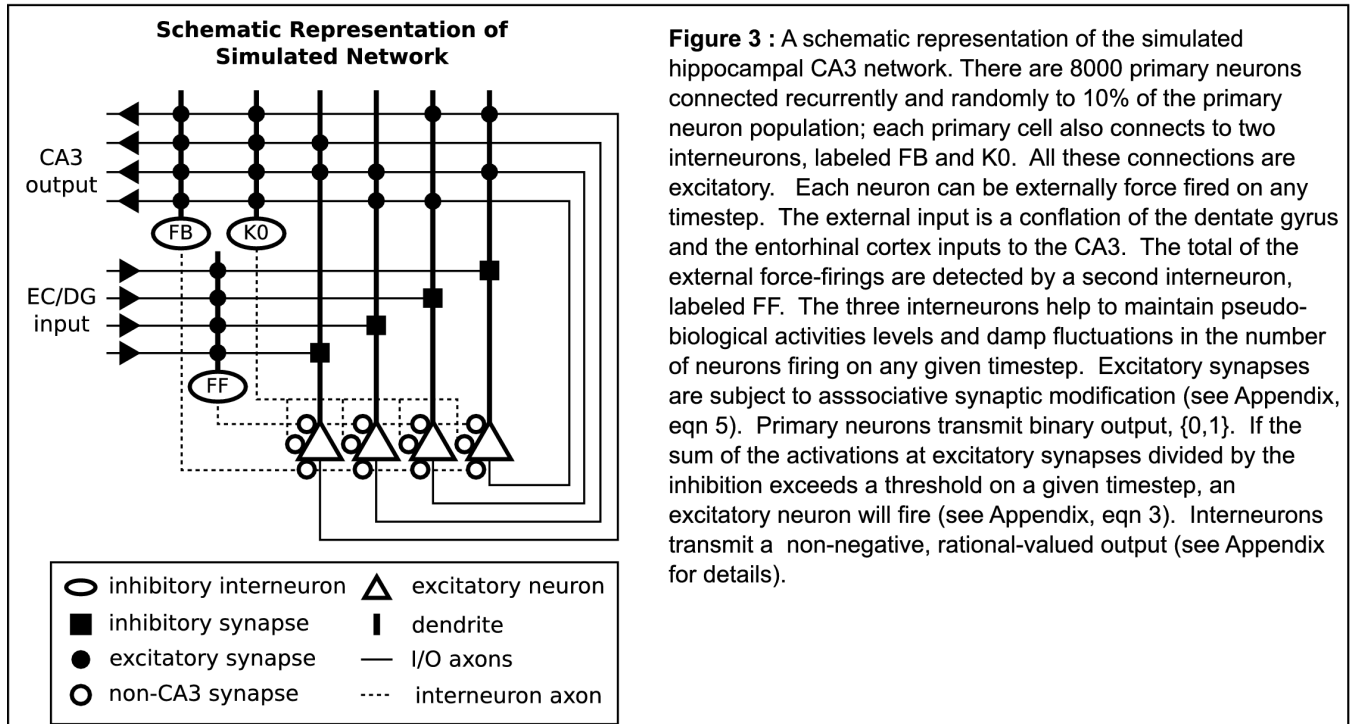


Figure 4 : The initial simplicity of minimal modeling gives way to a family of models of increasing complexity. As time resolution increases (qualitative, 20 ms, 10 ms, 5 ms) additional biology becomes explicit. At the first level of complexity, the qualitative timescale, synapses and a soma are sufficient – all biological events are abstracted to 1 timestep. At the second level of complexity (20 ms), external and internal timescales are matched via the first (slowest) time constant. In particular, glutamate binding to the NMDA-R decays with a 100 ms time constant and this is matched to the longevity of stimuli. (The decay of this binding over simulated timesteps is represented with the light gray “ramp” symbol.) At the third level of complexity (10 ms), the transfer of current from the dendrite to the soma takes more than one timestep. In particular, simulations suggest that a dendritic integration of 20 ms must be represented by a two-timestep boxcar (for 10 ms resolution). (The wide, white rectangle represents a dendritic element which integrates across timesteps.) At the fourth level of complexity (5 ms), the time to peak NMDA-R efficacy becomes quantitatively relevant. (The dark gray right triangle symbol represents this rise in current over multiple timesteps.) On the right of the diagram is a large version of the 5 ms cartoon.

A longer CS can increase the learnable trace (Wu & Levy 2004). Figure 1 illustrates the basic trace paradigm and shows the distinct, externally activated neurons that are, by assumption, coding for the conditioned stimulus (CS) and unconditioned stimulus (US).

Figure 2 shows three raster diagrams of different US neuron firing behaviors during a testing trial after at least 250 training trials. The left diagram depicts timely prediction of the US for a 700 ms trace interval. Successful prediction is judged by virtue of the firing times of the US neurons; the hippocampal prediction of the US can be used by some motor region to defend against the air-puff with a nictitating membrane blink. This raster diagram also displays a glimpse of the sequence (bridge) of firing neurons that connect the CS neurons to the US neurons (see Levy et al 2005b for more details about bridging over time and over neural space). The remaining two raster diagrams illustrate simulations of the two experimentally observed failure modes due to long trace intervals. The middle raster diagram shows complete failure to predict or recall the US (US neurons do not fire) for a 1000 ms trace interval. This simulation result is equivalent to the learning failure displayed where the rabbit does not blink at all. The right diagram shows an inappropriately early prediction of the US. This simulation result is equivalent to the learning failure where the rabbit blinks its nictitating membrane too early.

The Model And Its Sequential Development

The details of the current model, particularly how it

differs from our previous models, can be found in the appendix.

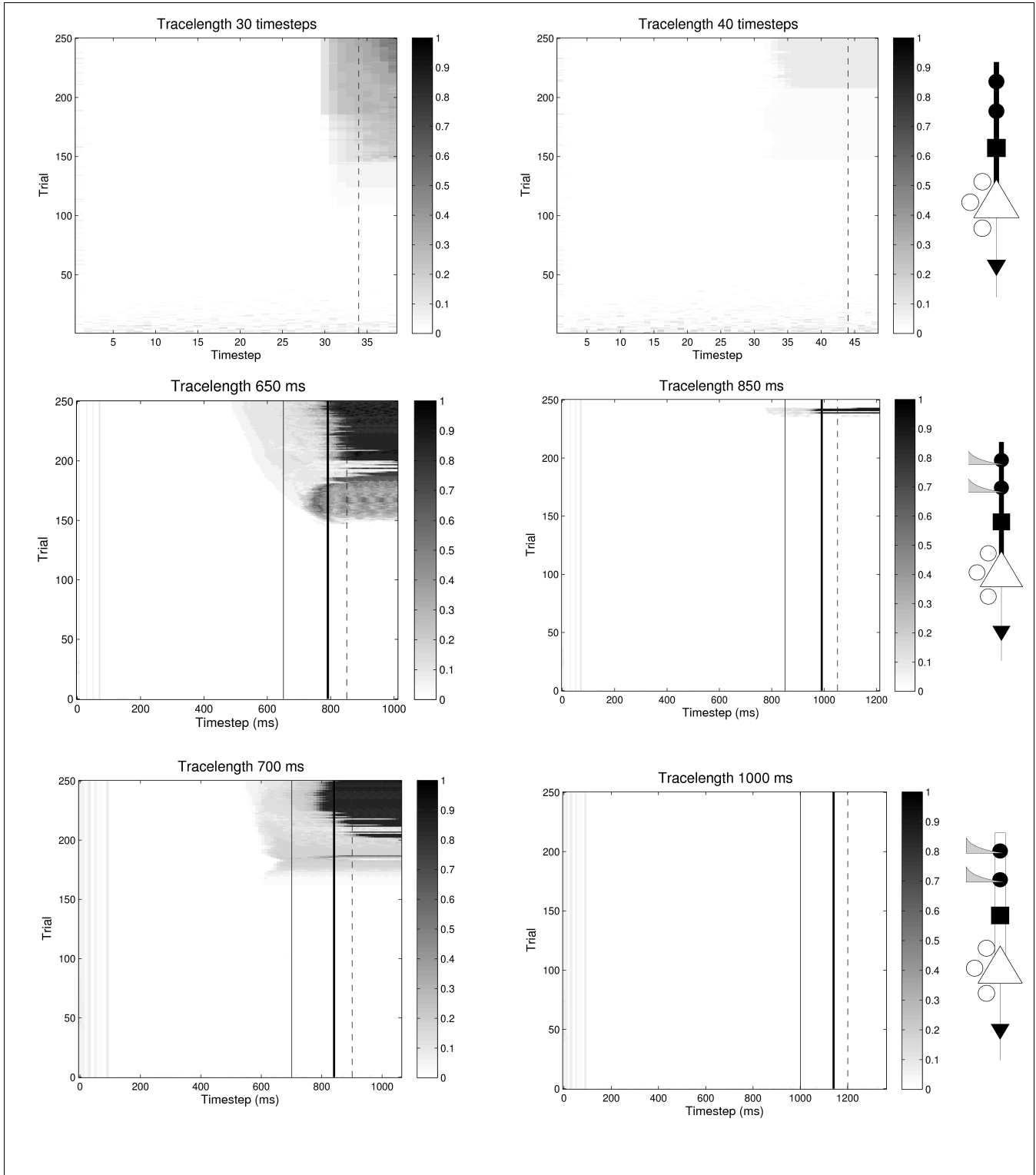
Figure 3 schematically presents the overall CA3 model consisting of randomly and sparsely interconnected, spiking neurons; further description is found in the legend.

The program of gradually enhanced complexity is illustrated in Fig. 4. The minimum neuron is qualitative in terms of time scale (far left, Fig 4). Traditional since McCulloch and Pitts (1943), the implemented biology does not provide a time scale relative to the network's updating cycle. One could pretend that a time scale is induced by the external stimuli durations (e.g. the 100 ms CS duration), but this seems overly crude if not arbitrary; we could use any number of update cycles to represent the 100 ms CS or other external intervals (which must be consistent with each other according to particulars of the training paradigm).

Turning our attention to the second level of model complexity, the qualitative becomes quantitative. We conjecture that the slowest and most important process in our data-fitting simulations is the off-rate-constant of the NMDA-receptor (NMDA-R). The experimental value of the off-rate is approximately 100 ms. The activation of this receptor controls synaptic modification in our model. Thus, if we set our updating equation for $\bar{z}_i(t)$ such that an e-fold decay occurs in five updating cycles (i.e., five timesteps), then one timestep represents 20 ms. Indeed, our best results in the past have used this exact approach (e.g. Sederberg & Levy 1997, Rodriguez & Levy 2001).

Figure 5 (next page): Introducing a boxcar integrator keeps the maximum learnable interval relatively stable across time resolutions. The first row displays results from a timescale-free competitive network. Timescale is present for rows 2 and 3 where the time resolution is increased from 20 ms to 10 ms, respectively. To compensate for these changes, a 20 ms boxcar is introduced at all timescales although it is implicit without specification at 20 ms. Using a gray scale, each diagram displays fractional US activity across all trials and timesteps during trace conditioning test trials. The vertical dashed line indicates when US neurons turn on in training trials. The two solid black lines mark the space where a successful prediction is possible: if US activity rises sharply in this region, the trial is successful. If firing of the US neurons occurs prior to the thin black line, the failure mode is blink-too-soon. If it occurs after the thick black line, the failure mode is blink-too-late (or not at all). For the competitive network, the CS is active from 0-100 ms. For the free-running networks, a 100 ms random firing sequence has been added to stabilize activity, so the CS is active from 100-200 ms. Each row pairs a successful learning, illustrating the maximum learnable trace interval, with a simulation that failed to learn a longer trace interval.

Figure 5



Select Simulation Parameters for Figure 5

Row	Resolution	Boxcar length	Activity	interneuron parameters				
				K_{FF}	K_{FB}	$K_0(0)$	λ_{FF}	λ_{FB}
First	qualitative	1 timesteps	10%*	n/a	n/a	n/a	n/a	n/a
Second	20 ms	1 timesteps	2.5 Hz	0.01100	0.03802	1.17000	1	1
Third	10 ms	2 timesteps	2.5 Hz	0.01200	0.04290	0.78400	4	4

* 10% of total population active per timestep (winners take all)

Moving to the next, more complex, model, we match this e-fold decay to ten timesteps. At this temporal resolution, it is necessary to control the time constant of dendritic integration to produce appropriate US prediction. In particular, a 10 ms dendritic integration time (no explicit integrator at this resolution) yields simulations that fail to appropriately predict the US. But, by adding a simple dendrosomatic integrator (a two-timestep, sliding-boxcar averaging device, see eqn (2) in the appendix), the simulated network appropriately predicts US onset. Thus, we return to a 20 ms neuron in terms of dendritic function via a minimalist form of dendritic integration.

The 5 ms time resolution requires enhancement of our simulation software. Testing of the new algorithm is not yet complete. However, we posit that the onset-rate constant of the NMDA-R is the most important variable to consider when enhancing temporal resolution from 10 ms to 5 ms.

RESULTS AND DISCUSSION

Examples generated by the four different models are presented in Fig 5. Each model is used for two different trace intervals; thus, the left-right pairs arise from the same biology but a shorter (650 or 700 ms) versus longer (850 or 1000 ms) trace interval. These figures plot the fraction of US neuron firings on each timestep of every test trial via a gray scale. (For the quantitative models, during testing, the CS is presented from 100 ms to 200 ms and then no input is given; the first two vertical lines indicate the boundary of an acceptable US prediction while the last vertical line indicates US onset during a training trial). For each figure in the left column, after sufficient training trials, the simulated network produces anticipatory US neuron firing.

In all these models, it may appear as if we have ignored the rate constant of synaptic modification, μ . In fact, the rate of synaptic modification depends on activity levels, rather than timestep resolution, because postsynaptic firing is required to trigger either potentiation or depression of a synaptic weight. In these models, we keep average activity constant, in terms of total firings-per-simulated-second over the whole network, therefore μ is not changed. The original setting of μ , for any one activity level, is mediated by matching the published learning curves (see Rodriguez & Levy 2001).

A notable aspect of this research program is the manner in which it reveals the implicit complexity of the original minimal model. That is, the model enhancements were always implicitly present in the simpler models. However, they were safely ignored due to the relationship between the temporal resolution of the simulation and the timescale of the neural

component. For example, if the timescale of dendritic integration is approximately 20 ms, then a simulation with 20 ms resolution appears to ignore dendritic integration but, in fact, the simulation is implicitly using this time constant.

Future Studies

It is apparent that even this simplified approach is not without its difficulties. For example, un-named implicit processes exist in our 20 ms model and must be considered when re-scaling temporal resolution. The particularly vexing issue is the on-rate time constant of the NMDA-R. The value we would like to use is 10 ms. However, because some delay in the synaptic modification rule is an absolute requirement, the minimum possible delay is 20 ms when the resolution is 20 ms. Thus, this non-physiological delay needs to be explored at higher temporal resolutions to confirm the meaning of the 20 ms model. Likewise, the self-inhibitory conductances that occur after an action potential are implicitly present at the lower temporal resolutions because the neurons can fire no faster than the resolution allows: 50 Hz for 20ms resolution, 100Hz for 10 ms resolution, etc. Thus, each of these variables must be carefully examined one-by-one and also examined interactively with the other variables.

In the future we can capitalize on the multiplicity of components in at least one way: when there is more than one biological time constant, any one of them can serve as the basis of mapping real time into the network model. This allows us to examine the role of any one component via parameter sweeps while maintaining the timescale of the model. For example, we can investigate the NMDA-R off-rate interaction with CS duration and the learnable trace interval using the dendritic integrator time-constant for fixing temporal resolution.

Further Extensions of the Model

The conflating of dendritic-somatic integration with conduction lags must eventually be suppressed in favor of explicit and separate representations. Presumably this might become necessary when simulating at or below the 2-5 ms range of temporal resolution (and network updating cycles). Another biological feature of suspected importance is the relative refractory period of the pyramidal cells. This is certain to become important at the higher temporal resolutions. Indeed, if there is a desire to keep firing rates biological (i.e., slow enough). There must be some version of a relative refractory process, such as is mediated by voltage- or calcium-dependent potassium conductances. Finally, the asymptotic approximation ($t \rightarrow \infty$) is used for $y_i(t)$ must ultimately give way to the use of the underlying differential equation(s).

Appendix : Mathematical Details of the Model

(1)	$y_j(t) = \frac{exc(t)}{exc(t) + K_0(t) + K_{FB} inhib_{FB}(t) + K_{FF} inhib_{FF}(t)}$	<i>net somatic excitation</i>
(2)	$exc(t) = \frac{1}{t_{max}} \sum_{t'=0}^{t_{max}-1} \sum_{i=1}^n w_{ij}(t-t') z_{ij}(t-t'-1)$	<i>dendritic excitation</i>
(3)	$z_j(t) = \begin{cases} 0 & \text{if } \left[y_i(t) < \frac{1}{2} \right] \wedge [x_j(t) = 0] \\ 1 & \text{otherwise} \end{cases}$	<i>spike generator</i>
(4)	$\bar{z}_i(t) = \begin{cases} 1 & z_i(t-1) = 1 \\ \alpha \cdot \bar{z}_i(t-1) & \text{otherwise} \end{cases}$	}
(5)	$w_{ij}(t+1) = w_{ij}(t) + \mu \cdot z_j(t) \cdot (\bar{z}_i(t) - w_{ij}(t))$	
(6)	$w_{il_{FB}}(t+1) = w_{il_{FB}}(t) + \lambda \cdot z_i(t) \cdot \left[\frac{\sum_{k=1}^n z_k(t)}{n} - a \right]$	<i>modification of excitatory synapses to inhibitory neurons</i> K_{FF} and K_{FB} (K_{FB} shown)

The net somatic excitation, y_j , of a simulated neuron j is governed by equation 1. This equation is the steady-state solution to a differential equation of the voltage as a function of excitation with a Nernst battery valued one and inhibition (also rest conductance) with a Nernst battery valued zero. The synaptically produced, dendritic excitation (exc) is allowed to have a boxcar integrator; this occurs when t_{max} is greater than one. When $t_{max}=1$, the excitation only depends on the active neurons from the last timestep. Inhibition arises from the three interneurons and is weighted via the corresponding K term.

At each timestep, a primary neuron transmits $z_i(t)$ via a binary {0,1} signal. Neuron firing is determined via equation (3). A neuron does not fire if its excitation, $y_i(t)$, is less than $\frac{1}{2}$ and the binary-valued external input, $x_i(t)$ is not forcing (i.e., $x_j(t)=1$ forces $z_j(t)=1$).

Associative synaptic modification of a recurrent excitatory synapse, w_{ij} , is determined by the recent

<i>Definitions</i>	
desired fractional activity per timestep	a
feedforward inhibitory constant	K_{FF}
feedback inhibitory constant	K_{FB}
NMDA-R off-rate decay constant	α
inhibitory synaptic modification constant	λ

<i>Timescale Invariant Parameters</i>		
number of neurons	n	8000
fractional connectivity	c	0.10
excitatory synaptic modification constant	μ	0.00375

history of presynaptic excitation $\bar{z}_j(t)$ and modification only occurs when the postsynaptic cell fires. The time constant α reflects the off-rate constant of the NMDA-R. Both potentiation and depression occur (i.e., $z_j(t)=1$) so that this synapse learns the conditional statistic $E[\bar{z}_i | z_i = 1]$ (Levy et al 1990).

The K_0 interneuron transmits across synapses valued $K_0(t)$; this interneuron has a constant spontaneous activity of one. The value of $K_0(t)$ is slowly adjusted by an error-corrector mechanism (thus the purpose of its inputs). The error-corrector compares the total activity over one trial to the desired total activity and adjusts $K_0(t)$ proportionately.

Feedback inhibition is a weighted sum of excitatory firing over time and is scaled with K_{FB} . This neuron is excited by an equation that is similar to equation 2. Feedforward inhibition is similar, but the $z_{ij}(t-t'-1)$ from equation 2 becomes $x_i(t-t')$. All the boxcars are matched for time constant of integration at each temporal resolution.

Synaptic modification of the excitatory inputs to the inhibitory neurons is controlled by an error corrector mechanism. An additional activity-detecting neuron that can generate the error signal is implicit in the model but would be explicit in the detailed biology. Equation 6 is the modification equation for the FB neuron. A similar equation (not shown) applies to the FF neuron but uses its inputs, $x_k(t)$.

Literature Cited

- Gormezano I, Kehoe EJ & Marshall BS (1983) Twenty years of classical conditioning research with the rabbit. *Progress in Psychobiology and Physiological Psychology*, 10 : 197-267.
- Howe AG & Levy WB (2007) A hippocampal model predicts a fluctuating phase transition when learning certain trace conditioning paradigms. *Cognitive Neurodynamics*, 1(2): 143-155.
- Levy WB (1989). A computational approach to hippocampal function. In RD Hawkins & GH Bower (Eds.), *Computational models of learning in simple neural systems* : 243–305. New York: Academic Press.
- Levy WB, Colbert CM, & Desmond NL (1990) Elemental adaptive processes of neurons and synapses: A statistical/computational perspective. In MA Gluck & DE Rumelhart (Eds.), *Neuroscience and connectionist models* :187–235. Hillsdale, NJ: Lawrence Erlbaum Assoc., Inc.
- Levy WB, Wu X & Baxter RA (1995) Unification of hippocampal function via computational/encoding considerations. In Lautrup B & Brunak S (Eds.), *International Journal Neural Systems*, 6(supp) : 71-80. Singapore; World Scientific Publishing Company.
- Levy WB (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, 6 : 579–590.
- Levy & Sederberg P (1997) A neural network model of a hippocampally mediated trace conditioning. *IEEE International Conference on Neural Networks*, 1 : 372-376
- Levy WB, Hocking AB & Wu X (2005a) Interpreting hippocampal function as recoding and forecasting. *Neural Networks*, 18 : 1242-1264.
- Levy WB, Sanyal A, Rodriguez P, Sullivan DW & Wu X (2005b) The formation of neural codes in the hippocampus: Trace conditioning as a prototypical paradigm for studying the random recoding hypothesis. *Biological Cybernetics*, 92 : 409–426.
- McCulloch WS & Pitts WA (1943) A logical calculus of the ideas immanent in neural nets. *Bulletin of Mathematical Biophysics*, 5(115).
- Moyer JR Jr., Deyo RA & Disterhoft JF (1990) Hippocampectomy disrupts trace eye-blink conditioning in rabbits. *Behavioral Neuroscience*, 104: 243-252.
- Rodriguez P & Levy WB (2001) A model of hippocampal activity in trace conditioning: Where's the trace? *Behavioral Neuroscience*, 115: 1224–1238.
- Solomon PR, Vander Schaaf ER, Thompson RF & Weisz DJ (1986) Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behavioral Neuroscience*, 100: 729-744.
- Wu X & Levy WB (2004) Increasing the CS and US longevity increases the learnable trace interval. *Neurocomputing* 65–66 : 283–289.