

Recurrently Connected Silicon Neurons with Active Dendrites for One-Shot Learning

John V. Arthur and Kwabena Boahen

Department of Bioengineering, University of Pennsylvania
Philadelphia, PA 19104, U.S.A.

jarthur@seas.upenn.edu and boahen@seas.upenn.edu

Abstract—We describe a neuromorphic chip designed to model active dendrites, recurrent connectivity, and plastic synapses to support one-shot learning. Specifically, it is designed to capture neural firing patterns (short-term memory), memorize individual patterns (long-term memory), and retrieve them when primed (associative recall). It consists of a recurrently connected population of excitatory pyramidal cells and a recurrently connected population of inhibitory basket cells. In addition to their recurrent connections, the excitatory and inhibitory populations are reciprocally connected. The model is novel in that it utilizes recurrent connections and active dendrites to maintain short-term memories as well as to store long-term memories.

I. NEUROMORPHIC SYSTEMS

Neuromorphic engineers aspire to match the computational abilities of neurobiological systems by morphing the microanatomy and physiology of these systems into custom hardware [1]. To date numerous sensory systems have been constructed. The principal example of these systems is the silicon retina [2]. The silicon retina includes thirteen cell types, wired according to the anatomical structure of the retina. However, such sensory systems are pointless without the higher-order regions to which they project. Therefore, neuromorphic engineers have developed a silicon model of axon guidance, which learns cortical topographic maps based solely on input correlations [3]. The logical progression is next to model deeper, more complex, and more plastic brain regions such as the hippocampus.

We have designed and fabricated a chip based on the neuroanatomy and neurophysiology of the hippocampus. Modeled after the hippocampus' CA3 region, this neuromorphic chip is designed to perform aspects of episodic memory, the memory of sequences of events in space and time. Specifically, it is designed to perform the aspects of episodic memory necessary for one shot-learning, which are to capture neural firing patterns (short-term memory), memorize individual patterns (long-term memory), and retrieve them when primed (associative recall).

II. ONE-SHOT LEARNING IN HARDWARE

Hardware associative memory networks, such as the one described in [4], require distinct learning (training) and recall phases to perform memory tasks. These memory systems require numerous presentations of each pattern to be stored during the off-line training phase. Once trained, the system can recall memories when presented with subsets of the stored patterns. In contrast, neurobiological associative memory networks, such as the hippocampus, perform one-shot learning. With just one presentation of a pattern, they can store a pattern. These networks appear to require no separate learning and recall phases. Instead, all learning occurs on line in real-time.

Our model is designed to capture episodes in short-term memory in a single presentation and to transfer each pattern in the episode into long-term memory over time. We realize short-term memory by modeling active basal dendrites of pyramidal cells [5]. We realize long-term memory by clustering strengthened synapses on basal dendrites [6]. We realize associative recall by exceeding the dendritic threshold to recruit inactive pyramidal cells in a stored pattern. Figure 1 shows our pyramidal cell population with recurrent connections to the basal dendrites and inputs to the apical dendrites.

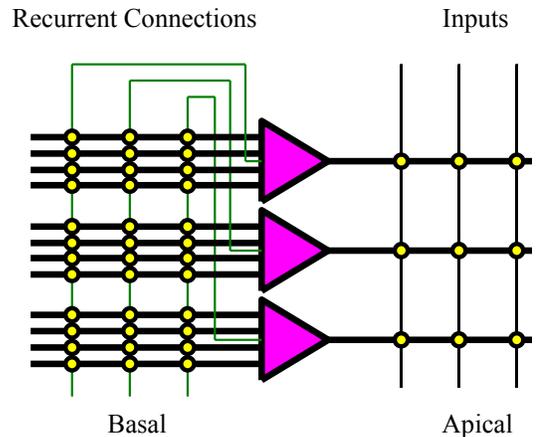


Fig. 1. Pyramidal cells recurrently connect to the basal dendrites of other pyramidal cells in the network.

A. Phase Code in One-Shot Learning

Our model uses a phase code to store sequential patterns of activity in short-term memory. Each activity pattern in the sequence reactivates at a specific phase of the intrinsic theta (5-12Hz) oscillation. The theta oscillation is divided into about seven slots by the intrinsic gamma (30-80Hz) oscillation. Each slot can store one short-term memory for a total of seven patterns in the sequence. The phase of the slot contains information about the sequential position of the short-term memory, a *phase code* [7].

The phase-code repeatedly reactivates multiple short-term memories in sequence at theta frequency, providing multiple opportunities to store each item in long-term memory with only a single presentation. Long-term memories are stored by strengthening recurrent synaptic weights among neurons active in the same pattern, which creates attractor states in the network [8]. These attractor states can be recalled by exciting a subset of neurons in the original pattern, which recruit their peers via the strengthened recurrent synapses.

The basal-dendrite model provides a mechanism for reactivating patterns. Reactivation results from the basal dendrites' active behavior. They contain model ionic channels (simplified voltage- and calcium-dependent calcium channels and calcium-dependent potassium channels from [9]) capable of generating repeated *calcium spikes* at theta frequency when provided with a constant or slowly decaying input. The slowly decaying input is provided by model NMDA synapses from other pyramidal cells in the pattern.

B. Long-term Potentiation in One-Shot Learning

The excitatory synapses that only contain NMDA receptors are called *silent synapses*. Silent synapses pass miniscule current unless the postsynaptic neuron is sufficiently depolarized, because of the NMDA channel's voltage dependence. Therefore, recurrent silent synapses only excite neurons that were depolarized by an external input, maintaining the short-term memory without recruiting other neurons. The pattern of activity is repeated over and over again, reactivating the short-term memory and providing multiple opportunities for transfer to long-term memory.

The basal-dendrite model also provides a mechanism for long-term memory: increased synaptic strength among pyramidal cells active in the same pattern. To increase synaptic strength, pyramidal cells augment their purely NMDA silent synapses from other pyramidal cells in the pattern with AMPA receptors in a process called *long-term potentiation* (LTP) [10]. LTP is known to depend on the magnitude of the NMDA current flowing through a silent synapse [10]. Therefore, in the model, when a NMDA current flowing from a silent synapse to a basal dendrite exceeds a threshold, that silent synapse is potentiated on that basal dendrite.

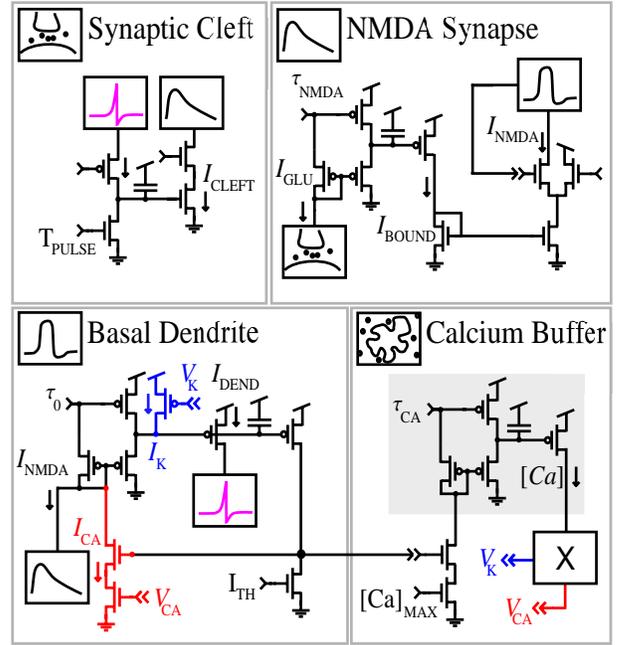


Fig 2. Transistor implementations are shown for a synaptic cleft, NMDA synapse, basal dendrite (all simplified), and calcium buffer, with circuit icons in the top-left of each box. The *synaptic cleft* circuit generates a current pulse (I_{CLEFT}) of about one millisecond duration from the current impulse generated by the presynaptic action potential. The *NMDA synapse* circuit passes the synaptic-cleft current pulse through a low-pass filter (LPF) to compute the glutamate bound to the NMDA receptors (I_{BOUND}). The NMDA current (I_{NMDA}) is computed from the glutamate bound to NMDA receptors and the dendritic potential. The *basal dendrite* circuit, which receives synaptic input and stimulates the soma (magenta) with its output, is also based on a LPF, with model potassium (blue) and calcium (red) channels added. The calcium channels are opened when the dendritic potential (I_{DEND}) exceeds the calcium threshold (I_{TH}). The calcium current is integrated by a LPF in the *calcium buffer* circuit to obtain the dendritic calcium concentration ($[Ca]$), which controls the model potassium and calcium channels (via a scaling circuit).

When synapses from many pyramidal cells active in a pattern are potentiated on a basal dendrite the long-term memory is stored and can be associatively recalled. Associative recall occurs when inputs from stimulated pyramidal cells converge onto the same dendritic branch of an inactive peer and exceed the dendritic threshold. In this fashion, an entire pattern may be recalled given a subset of the pattern as input.

III. SILICON SYSTEM

The primary circuit used in the CA3 Chip is a current-mode first-order low-pass filter (Figure 2 *gray box*). The low-pass filter (LPF) is mathematically identical to a resistor-capacitor circuit. However, in the LPF currents represent both voltages and conductances. The equation that describes the LPF is given by

$$\frac{CU_t}{I_t} \frac{dI_{OUT}}{dt} = -I_{OUT} + I_{IN} \quad (1)$$

where I_{IN} is the input current, I_{OUT} is the output current, and I_{τ} , which represents a conductance, sets the time constant (C is the capacitance and $U_t = 25\text{mV}$ is the thermal voltage). Changing a current that represents a conductance is equivalent to changing that conductance. We can implement neurons with voltage-dependent conductances by making currents that represent conductances depend on currents that represent voltages. Therefore, we can implement a large set of conductance-based neuron models in silicon. Each neuron, synapse, membrane compartment, ionic channel population, and potentiation circuit is modeled by a LPF with different inputs and properties.

A. Synapses

We implement three types of synaptic conductances: AMPA, GABAA, and NMDA. All three types include a synaptic cleft circuit. The synaptic cleft circuit models the time course of neurotransmitter release, which lasts for about one millisecond (Figure 2). The neurotransmitter release is modeled as a current pulse, I_{CLEFT} , with width T_{PULSE} (set by the voltage applied to the gate of the leak transistor) and amplitude I_{MAG} (set by the voltage applied to the gate of the series-connected transistor). We use this current directly to model the AMPA conductance since this conductance is extremely fast.

The GABAA circuit models the amount of neurotransmitter bound to the receptor and the resulting conductance, I_{GABAA} , using a single LPF. Hence,

$$\frac{CU_t}{I_{\tau GABAA}} \frac{dI_{GABAA}}{dt} = -I_{GABAA} + I_{GABA} \quad (2)$$

where I_{GABA} represents the neurotransmitter in the synaptic cleft and $I_{\tau GABAA}$ sets the time constant.

The NMDA circuit models the glutamate bound to NMDA receptors and the conductance of the NMDA receptor-gated channel. The glutamate bound to the NMDA receptors, I_{BOUND} , is modeled as a current pulse filtered by a LPF, as described by

$$\frac{CU_t}{I_{\tau NMDA}} \frac{dI_{BOUND}}{dt} = -I_{BOUND} + I_{GLU} \quad (3)$$

where I_{GLU} represents the neurotransmitter in the cleft and $I_{\tau NMDA}$ sets the time constant. The NMDA conductance, represented by the current I_{NMDA} , is obtained by assuming this channel's voltage dependent block has the form

$$I_{NMDA} = I_{BOUND} \frac{I_{DEND}^N}{I_{DEND}^N + I_{THNMDA}^N} \quad (4)$$

where I_{DEND} represents the basal dendrite's potential, I_{THNMDA} is the NMDA threshold, and $N = 5$ is the Hill coefficient of the voltage dependence.

We use these three conductances to model four types of synapses: ampaergic (with AMPA conductance), silent (NMDA), potentiated (both), and gabaergic (GABAA).

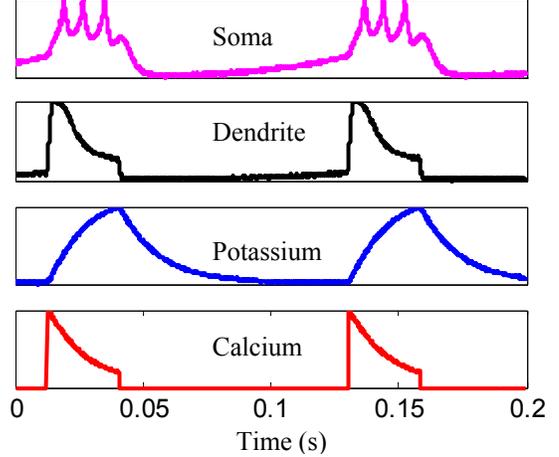


Fig. 3. Calcium spikes generated by the basal dendrite circuit (black) cause the soma to burst (magenta). The calcium spike occurs when the dendrite surpasses a threshold, turning on the calcium current (red). The accumulation of calcium turns on the potassium (blue), which terminates the spike. This data is taken from a silicon neuron with a constant input to basal dendrite.

B. Neurons

We implement two types of neurons: basket cells and pyramidal cells. The basket cell is the simpler of the two neuron types. It consists of a LPF augmented with model voltage-dependent sodium and potassium channels for spike generation (modified from [12]). It receives GABA receptor mediated inhibition from the basket cell population (including itself) and AMPA receptor mediated excitation from the pyramidal cell population.

The pyramidal cell consists of a soma and forty-five basal dendrites, each of which receives input from every one of the pyramidal cell's forty-five silent synapses. The soma is identical to the basket cell, receiving gabaergic inhibition from the basket cell population.

The basal dendrite circuit models the generation of calcium spikes by voltage- and calcium-dependent calcium channels and calcium-dependent potassium (Figure 2). It represents the dendritic potential as a current, I_{DEND} , which is described by

$$\frac{CU_t}{I_K + I_{\tau 0}} \frac{dI_{DEND}}{dt} = -I_{DEND} + (I_{IN} + I_{CA}) \frac{I_{\tau 0}}{I_K + I_{\tau 0}} \quad (5)$$

where I_{IN} is the synaptic input current, I_{CA} (red) is the calcium current, I_K (blue) is the potassium current, and $I_{\tau 0}$ sets the time constant when I_K is zero.

I_{CA} , which acts to increase I_{DEND} , is modeled by a threshold and a scaling circuit, as described by

$$I_{CA} = I_{CA_MAX} \frac{[Ca]_{MAX} - [Ca]}{[Ca]_{MAX}} u(I_{DEND} - I_{TH}) \quad (6)$$

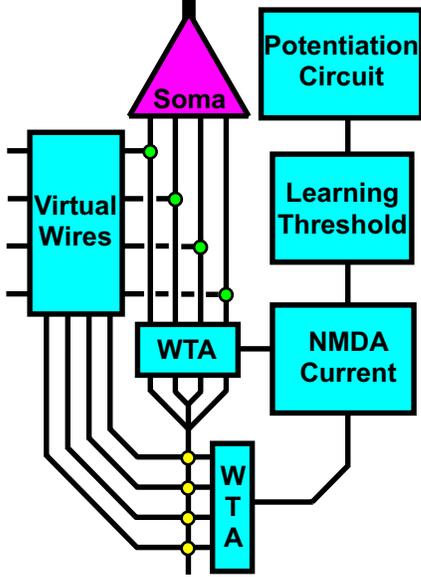


Fig. 4. Long-Term Potentiation occurs when the NMDA current, which depends on both dendritic potential and synaptic input, surpasses a threshold. Only one silent synapse (yellow) can be potentiated (green) in a given infinitesimal time window. This synapse should be the one with the largest NMDA current. The source of the largest current is the most active silent synapse, and the sink of the largest current is the most active basal dendrite. Two winner-take-all (WTA) circuits select the most active silent synapse and basal dendrite. If the NMDA receptor dependent current between them exceeds a threshold, a signal is sent off chip.

where $[Ca]$ is the dendritic ‘calcium concentration’; it has a maximum value of $[Ca]_{MAX}$. I_{CA_MAX} is the maximum calcium current and $u(x)$ is the Heaviside function. Hence, I_{CA} turns on when I_{DEND} exceeds I_{TH} , the calcium threshold, initiating a calcium spike.

I_K , which shunts I_{DEND} , is modeled by a scaling circuit described by

$$I_K = I_{K_MAX} \frac{[Ca]}{[Ca]_{MAX}} \quad (7)$$

where I_{K_MAX} is the maximum potassium current. Hence I_K turns on when $[Ca]$ rises.

$[Ca]$ is modeled by a LPF (Figure 2), representing various endogenous calcium buffers, and is described by

$$\frac{CU_t}{I_{tCA}} \frac{d[Ca]}{dt} = -[Ca] + [Ca]_{MAX} \cdot u(I_{DEND} - I_{TH}) \quad (8)$$

where I_{tCA} sets the calcium time constant. As $[Ca]$ rises, I_K increases and I_{CA} decreases, terminating the calcium spike (Figure 3).

C. Potentiation circuit

A pyramidal cell has a single potentiation circuit that evaluates pair wise correlations between all forty-five of its basal dendrites and all forty-five of its silent synapses. Therefore, each potentiation circuit is constantly evaluating

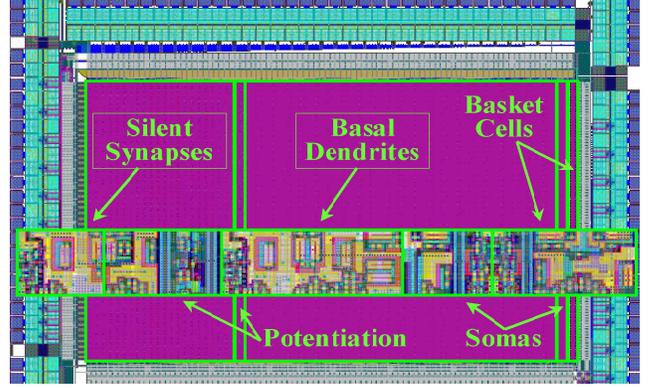


Fig 5. Layout of CA3 Chip with 100 pyramidal cells, each with 45 basal dendrites and 45 silent synapses, is shown. These two 45 by 100 arrays take up most of the chip area. Between them is a column of 100 potentiation circuits and they are bordered on the right by two columns, one with 100 somas and the other with 100 basket cells. The insert shows a close up view of the layouts of these individual circuits, which are tiled horizontally, to build each pyramidal cell. This horizontal slice is 17.9 microns tall; the entire chip is 3.9mm by 2.5mm.

2025 correlations. The potentiation circuit is able to evaluate all correlations by looking only at the basal dendrite and silent synapse that have the highest correlation at any point in time, which must be the most active of each (Figure 4). The most active is selected through a winner-take-all competition with its peers [13]. The winning currents are sent to the potentiation circuit. The potentiation circuit evaluates the NMDA receptor gated current between this most active pair. If the current exceeds threshold, the potentiation circuit sends the basal dendrite and silent synapse’s addresses off chip in an address event. Circuitry off chip implements binary valued LTP of the silent synapse, moving it a potentiated site on the basal dendrite with which it is well correlated (by swapping memory locations in a lookup table).

We have designed and submitted a chip that implements the basal-dendrite model, the CA3 Chip (Figure 5). The CA3 Chip was fabricated through MOSIS in TSMC’s (Taiwan Semiconductor Manufacturing Company) 0.25-micron CMOS process. The CA3 Chip contains 430,000 transistors in 10mm^2 . These transistors model 100 inhibitory basket cells, 100 pyramidal cells, each with 45 basal dendrites and 45 silent synapses, and 100 potentiation circuits. The neurons fire action potentials in real-time, sending spikes off chip and receiving spikes on chip using the address-event representation [11].

IV. RESULTS

We investigated performance of the CA3 Chip in short-term memory, long-term memory, and associative recall tasks. Our setup uses a Tektronix Pattern Generator to briefly stimulate silent synapses and basal dendrites of pyramidal cells on the CA3 Chip. Recurrent connections are implemented with RAM (random access memory) and a

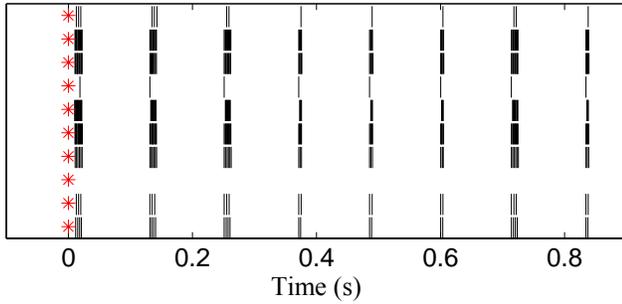


Fig. 6. In the short-term memory task, ten randomly chosen silicon pyramidal cells are stimulated by a brief external input (red stars) and most (9/10) respond by repeatedly firing action potentials (black raster lines) at theta frequency, maintained by NMDA-like recurrent connections and active basal dendrites. No other pyramidal cells fired.

Lattice CPLD (complex programmable logic device). We acquire address-event (spike) data in real-time using a National Instruments DIO (digital input-output) card in the PC and custom software, which displays the spike rasters.

For short-term memory experiments, ten of one hundred pyramidal cells are randomly selected. One basal dendrite from each of these ten pyramidal cells is selected. These selected basal dendrites are briefly excited from the pattern generator with three spikes one microsecond apart (yields same results for inputs several milliseconds apart). The rasters are recorded.

The CA3 Chip failed to store short-term memory because the pyramidal cells could not maintain synchrony by exciting each other. However, we were able to entrain them to the basket cells, which were slowed down to oscillate at theta frequency for this purpose (Figure 6). Thus, we were able to realize short-term memory behavior through synchronization by inhibition, but not through synchronization by excitation as we originally intended. Co-opting the inhibitory basket cells in this fashion precluded storing multiple simultaneous short-term memories. Thus, we were unable to test this functionality.

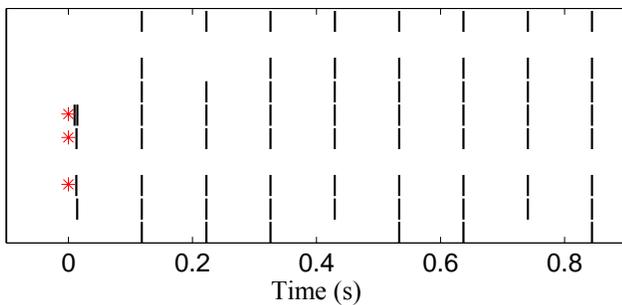


Fig. 8. In the associative recall task, three of ten silicon pyramidal cells in a previously stored memory are stimulated by a brief external input (red stars) and most (8/10) respond by repeatedly firing action potentials (black raster lines) at theta frequency, maintained by potentiated recurrent connections and active basal dendrites. No other pyramidal cells fired.

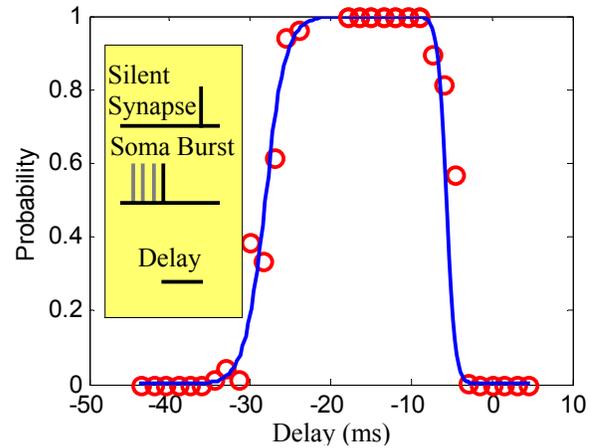


Fig. 7. In the long-term memory task, suprathreshold input to a randomly chosen pyramidal cell's basal dendrite and silent synapse are paired such that the last spike in the somatic burst arrives with some delay after the silent synapse input. If the delay is between 5 and 30 milliseconds, the potentiation circuit will likely send a command to potentiate the silent synapse on the stimulated basal dendrite.

For long-term memory experiments, one basal dendrite and one silent synapse are each excited. The basal dendrite receives enough excitation to cause a calcium spike, eliciting a burst of action potentials at the soma. The spikes from the pyramidal cell as well as any learning signals from the potentiation circuit are recorded. The delay between the time the silent synapse receives input and the time of the last spike in the burst is computed. By varying the delay between the times at which the basal dendrite and silent synapse are stimulated, and performing multiple trials, we can measure the probability of potentiation at each delay relative to the end of the burst.

The CA3 Chip successfully sends a potentiation signal when a silent synapse and basal dendrite are well correlated. The potentiation circuit exhibits a form of spike timing dependent plasticity (Figure 7). If the silent-synapse activation precedes the last spike of the burst by up to 30 milliseconds, it is likely to be potentiated. However, during high activity (where many pyramidal cells and basket cells are spiking) digital noise causes the analog potentiation circuits to send frequent erroneous signals that contain only a silent-synapse or basal-dendrite address, instead of both together. These errors have prevented us from implementing real-time learning of long-term memories by the system.

For associative recall experiments, an activity pattern (or several patterns) consisting of ten of one-hundred pyramidal cells is stored in RAM. One basal dendrite on each pyramidal cell in the pattern receives potentiated synaptic input from every other pyramidal cell in the pattern. Three of the ten pyramidal cells in the memory are randomly selected. The basal dendrites involved in the memory of the selected pyramidal cells are excited from the pattern generator with three spikes one microsecond apart (yields same results for inputs several milliseconds apart). The rasters are recorded.

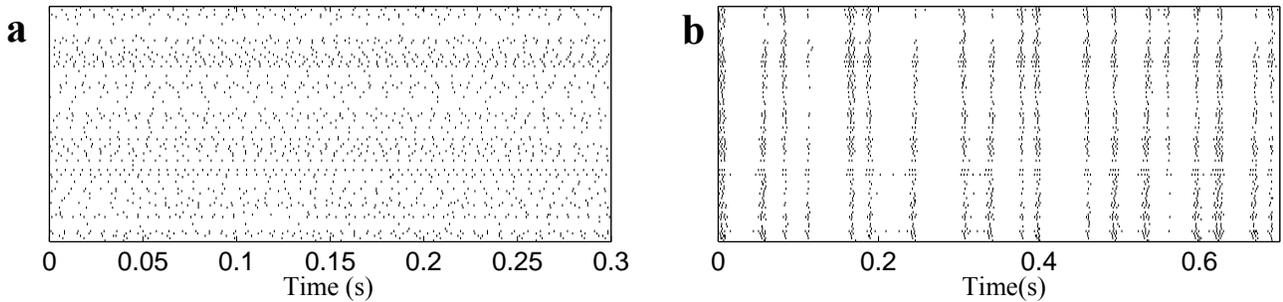


Fig. 9. Silicon basket cells fire action potentials (black raster lines) (a) at variable rates in response to a constant current input when disconnected from their peers, (b) but synchronize when connected with global inhibition.

The CA3 Chip successfully performs associative recall. When three of ten pyramidal cells involved in a previously stored memory pattern are activated, they recruit most or all of their unstimulated peers in the memory (Figure 8). However, the CA3 Chip is unable to store many patterns with overlapping pyramidal cell activities. We have stored only six memories. The dendritic thresholds are sensitive to transistor mismatch. Therefore, some pyramidal cells can be recruited with just one spike, whereas others require numerous spikes. The most excitable dendrites quickly recruit other excitable dendrites that are active in the overlapping patterns. Within seconds every pyramidal cell in the network fires synchronously at a frequency determined by the inhibition.

V. DISCUSSION

The CA3 Chip performs short-term memory, long-term memory, and associative recall in a limited manner. It is limited by its sensitivity to transistor mismatch, which creates variability in the calcium-spike thresholds and in the oscillation frequencies. These mismatched oscillation frequencies disrupted synchrony by excitation but did not disrupt synchrony by inhibition. Hence, the basal dendrites (firing rate coefficient of variation = 1.02) were unable to synchronize on their own, whereas the basket cells (firing rate coefficient of variation = 0.95) were able to synchronize them as well as synchronize on their own (Figure 9). Thus, inhibition proved to be a more robust synchronization mechanism than excitation.

In the next silicon model, we plan to compensate for the variable oscillation frequency of the calcium spikes by employing a second population of interneurons to provide dendritic inhibition to coax the dendrites to synchronize. This dendritic synchronization should allow storage of multiple short-term memories, isolated by somatic inhibition from basket cells. We will compensate for the variable calcium spike thresholds by employing adaptive mechanisms analogous to spike frequency adaptation or a homeostatic mechanism to control the dendritic threshold. We will compensate for the potentiation circuit noise by introducing

an additional CPLD into the system to filter erroneous events, allowing the goal of real-time one-shot storage of patterns as long-term memories to be achieved.

ACKNOWLEDGEMENTS

We thank Kai Hynna for improving the spike generating circuit and Paul Merolla and Brian Taba for help in testing. The Office of Naval Research funded this work.

REFERENCES

- [1] C. Mead, "Neuromorphic Electronic Systems", *IEEE Proc.*, vol. 78, pp. 1629-1636, 1990.
- [2] K. A. Zaghoul, K. Boahen, "Optic Nerve Signals in a Neuromorphic Chip I: Outer and Inner Retina Models", *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 657-666, 2004.
- [3] B. Taba, K. A. Boahen, "Topographic Map Formation by Silicon Growth Cones", in: *Advances in Neural Information Processing Systems 15* (MIT Press, Cambridge), pp. 1163-1170, 2003.
- [4] A. Johannet, L. Personnaz, G. Dreyfus, J. D. Gascuel, M. Weifeld, "Specification and Implementation of a Digital Hopfield-Type Associative Memory with On-Chip Training", *IEEE Trans. Neural Networks*, vol. 3, pp. 529-539, 1992.
- [5] G. Ariav, A. Polsky, J. Schiller, "Submillisecond Precision of the Input-Output Transformation Function Mediated by Fast Sodium Dendritic Spikes in Basal Dendrites of CA1 Pyramidal Neurons", *J. of Neurosci.*, vol. 23, pp. 7750-7758, 2003.
- [6] P. Poirazi, B. W. Mel, "Impact of Active Dendrites and Structural Plasticity on the Memory Capacity of Neural Tissue", *Neuron*, vol. 29, pp. 779-796, 2001.
- [7] O. Jensen, M. A. P. Idiart, and J. E. Lisman, "Physiologically realistic formation of autoassociative memory in networks with theta/gamma oscillations: role of fast NMDA channels", *Learning and Memory*, vol. 3, pp. 243-256, 1996.
- [8] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *PNAS*, vol. 79, pp. 2554-2558, 1982.
- [9] D. S. Wei, Y. A. Mei, A. Bagal, J. P. Kao, S. M. Thompson, C. M. Tang, "Compartmentalized and Binary Behavior of Terminal Dendrites in Hippocampal Pyramidal Neurons", *Science*, vol. 293, pp. 2272-2275, 2001.
- [10] R. C. Malenka, R. A. Nicoll, "Long-Term Potentiation—A Decade of Progress?", *Science*, vol. 285, pp. 1870-1874, 1999.
- [11] K. A. Boahen, "Point-to-Point Connectivity Between Neuromorphic Chips Using Address-Events", *IEEE Trans. on Circuits and Systems II*, vol. 47, pp. 416-434, 2000.
- [12] E. R. Culurciello, R. Etienne-Cummings, K. A. Boahen, "A Biomorphic Digital Image Sensor", *IEEE J. Solid State Circuits*, vol. 38, pp. 281-294, 2003.
- [13] J. Lazzaro, S. Ryckebush, M. A. Mahowald, C. A. Mead, "Winner-Take-All Networks of O(n) Complexity", *Advances in Neural Information Processing Systems*, vol. 1, pp. 703-711, 1989.