Towards a Nanoscale Artificial Cortex

Alice C. Parker, Aaron K. Friesz, and Afshaneh Pakdaman
Department of Electrical Engineering EEB 348 MC 2562
University of Southern California, Los Angeles, California, 90089-2562 USA
parker@eve.usc.edu

Abstract – This paper presents an approach to predicting the feasibility of artificial brains in the future. We focus on biomimetic neural models and electronic circuits that implement those models. Complexities in modeling biological neural tissue are discussed. Estimates are given for the size of artificial neural systems based on CMOS technology in 2021, without considering interconnections. We propose some solutions to the problem of interconnecting neurons. However, the best solution to this issue is an ongoing research topic.

Keywords – biomimetic, neuromorphic, electronic, artificial brain, neuron, prediction

1. Introduction

From the early days of vacuum tube and relay electronics, researchers have developed electronic models of neurons designed to *emulate* neural behavior with electrical signals that mimic in some ways the measured potentials of biological neurons. However, historically, the size and cost of available electronics made construction of complex brain-like structures infeasible. As technologies have become smaller and less expensive, there is a possibility of constructing neural structures on the scale of a human brain in the foreseeable future. Therefore, we have begun to examine the feasibility of building extremely large-scale neural systems.

Autonomous vehicle navigation, identity determination, robotic manufacturing, and medical diagnostics are all engineering challenges that could benefit from technological solutions that involve artificial neural structures. A common capability required for each of the above challenges is image understanding, an activity at which humans excel. A further motivation for creating artificial brain structures is neural prosthetics. While other very difficult engineering problems must be solved to interface biological and artificial neurons, understanding the technological requirements and possibilities for isolated artificial brain structures could accelerate the hope that such prostheses would become a reality.

This paper is aimed at answering one simple question: when are we going to be able to construct an artificial brain of reasonable size and cost that exhibits almost real-time behavior? While there are hundreds of brain structures that have somewhat different anatomical characteristics [1], this paper focuses on cortical structures because of the density and complexity of their interconnections. This paper also focuses on custom biomimetic circuits designed to emulate neurons, sometimes called neuromorphic circuits [2].

Our focus on circuits is in contrast to computer models that *simulate* neural behavior using conventional multiprocessors. In recent years, several artificial brain projects have been launched that rely on computer simulations of neural behavior, hosted on multiprocessors, such as the IBM artificial brain project in cooperation EFPL [3]. Emulation can have several advantages over simulation, generally being faster, sometimes running at nearly real time for small problems. For brain emulation, speed may not be as critical since neurons are slow in comparison to electronics, but performance becomes an issue when the interconnection hardware is extensively time-shared to interconnect many parts of the brain.

The USC artificial brain research is progressing along two lines:

- First, we are researching the feasibility of emulating brain structures that are biomimetic, that reflect the subtleties of the biological neuron behavior, and
- Second, we are engaging in some predictions of the future of each potential technology that enable us to speculate on the size and performance of a brain emulator.

This paper describes the latter of these two research directions – predictions of future technological progress that will support artificial brain structures. The paper focuses on one specific prediction question – when and if nanoscale CMOS technology can eventually be used to construct artificial brain structures. We are making numerous assumptions in this paper in order to simplify this prediction problem, and to arrive at some preliminary conclusions. These predictions will become increasingly accurate as each assumption is confirmed, refined or rejected. As research progresses and new information is made available, the prediction model will evolve.

2. Challenges in Modeling Biomimetic Neurons and the Brain

One of the most difficult of the challenges in modeling the brain is the massive interconnectivity. Cortical neurons possess an average of 10,000 and up to 100,000 synaptic connections. [4] With approximately 100 billion neurons in the human cortex, and approximately 60 trillion synaptic connections, connectivity in the artificial brain cortex will be a major challenge. Even in the cerebellum there are postulated to be 100 billion small granule cells, each with up to 100 synaptic connections. [5]. While some connections originate in proximal neurons, some originate in distal neurons, posing an interconnection problem for the candidate modeling technologies. One of the foci of the predictions is on speculation regarding possible interconnection strategies that support the massive synaptic connections of the cortex.

A second challenge involves the complexities of the synaptic connections, and variations in neural processing not just in different regions of the brain but also in different types of neural cells involved in proximal processing [5, 6, 7]. Modeling a complete list of neural complexities and variations is probably not required for good emulation, but major features that contribute in some significant way to neural behavior must be considered. First, we consider some complexities that are found in most neurons, and then we examine some variations between neurons that are significant to behavior at the cellular level. These complexities can be categorized into chemical complexities, synaptic structural variations, and dendritic structural variations. Most of these variations and complexities are discussed in the reference [5].

One of the complexities of neural tissue is the existence of transmitters, chemical messengers that can decrease or increase the excitability of the postsynaptic receptors to stimuli by the pre-synaptic cells, possibly by altering cell membrane conductance. A further complication of transmitter function is via the retrograde process that directly or indirectly modulates transmitter release in the presynaptic junction, a form of extremely local feedback. Retrograde actions typically occur more slowly than presynaptic to postsynaptic activation. In arriving at the predicted size of electronic neurons, we estimate that several transistors per synapse will be used to model this delay.

Transmitters acting on secondary messengers can have short or long-term effects on synaptic junction activation, referred to as short-term facilitation and depression, and long-term potentiation (LTP) and long-term depression (LTD). Facilitation increases the likelihood of the neuron firing, and depression decreases the likelihood. The activation probability of a given synaptic junction is up- or down-regulated

by the amount and timing of presynaptic and postsynaptic activity. A final transmitter complication involves the occurrence of multiple transmitters at a single synapse, sometimes providing conflicting messages of potentiation and depression.

A very significant second complexity of neural structure is the existence of synaptic divergence (*fan out* in engineering terms) and convergence (*fan in*). A single axon can fan out to several presynaptic connections, or several synapses can form around a single axon. Multiple synapses can converge (fan in) to a single postsynaptic terminal, either from a single oversized presynaptic terminal or from multiple presynaptic terminals. Axons can influence the activation of other axons directly, either by sharing ion flow, or by forming synaptic connections, axon to axon. Likewise, there can be dendrodendritic connections that act as synaptic junctions. Synaptic divergence can enhance the signal to noise ratio and hence is useful in constructing brain emulators that are fault tolerant, like the brain itself. A more subtle type of interconnection involves synchronization, where activity in one neural process influences activity in other processes.

Multiple connections, either fanning into or out of a neuron tend to support the same sign activities, either excitatory or inhibitory. Some synaptic connections have a low probability of firing. Multiple connections increase the probability of activation occurring, and therefore increase the "safety factor" of the sub-circuit firing properly. Multiple synapses at the inputs to a single neuron that produce *excitatory* or *inhibitory post synaptic potentials* (EPSPs and IPSPs) can combine, or sum spatially or temporally to produce a potential that is a non-linear summation of the single potentials. In addition, the action of the inhibitory potentials can depend on the type of inhibition. Modeling this non-linear summation is considered important in capturing essential brain functioning.

A further complication is imposed by *feed-forward inhibition*. Here excitatory synapses make direct connections from a pre-synaptic to post-synaptic connection, while other synapses from the "output" neuron connect through a relay neuron to the "input" neuron through an inhibitory synapse. There is a delay through the relay neuron creating an excitatory-inhibitory sequence in the output neuron. Another neuronal behavior that might be important to capture is *recurrent inhibition*, where excitatory potentials in a neuron back propagate to dendrites on that neuron that activate other neurons through dendrodendritic junctions. Those neurons then inhibit the original neuron through inhibitory synapses that create IPSPs in the original neuron. *Lateral inhibition* also occurs. An EPSP in a neuron can activate IPSPs in neighboring neurons. Some feedback structures produce rhythmic activity or oscillations. Some neurons produce bursts of spikes with a single activation, others produce a single spike. The periods and duty cycles of the spike trains are sometimes significant, and variations in frequency cause different responses.

Dendritic structural variations have a first-order impact on the behavior of individual neurons and neuronal circuits. The location of excitatory and inhibitory synapses on the dendrite branches and spines determines the functions realized by the combinations of synaptic inputs [8]. In Section 5, we will illustrate a candidate electronic design for a dendritic tree that has functions dependent on the location of the synaptic connections.

3. The Prediction Approach

The central goal of this paper is to predict when the construction of a reasonably sized bio-mimetic artificial brain will be possible. The sheer size of the structure using current technologies is projected to be enormous; the interconnections between neurons pose further problems due to the large number of synaptic connections for a typical cortical neuron. The structure and behavior of individual neurons is

complex and highly variable. In addition to conventional CMOS technology, nanotechnology offers new and promising alternatives for molecular-scale integration of artificial brains. Although those alternatives are being actively investigated in our group, this particular paper focuses on CMOS nanotechnology.

First, we examine the future trends in CMOS technologies coupled with examination of artificial neuron circuits built in CMOS, in order to project when and if CMOS circuits might be small enough to make a CMOS artificial brain feasible. An ancillary study on biomimetic CMOS neural structures is yielding statistics on neuron size and interconnectivity. For this part of the prediction, spike train artificial neurons have been investigated, along with learning circuits, modeling the effects of neurotransmitters, and various memory configurations.

Second, the prediction approach also involves postulating architectures for artificial brains that support a high degree of interconnectivity. One possible avenue we have investigated is the use of on and off-chip networks that connect neurons, with addressable synapses of individual neurons considered, using a network-on-chip approach developed at USC [9]. This is in contrast to the bus-oriented interconnection scheme Boahen employs [10].

4. **Related Work**

A number of researchers have designed, fabricated and tested artificial neurons, and the trend is toward increasingly biomimetic neurons. Others have used general-purpose computers to simulate neural networks. Schüffny et al. provide a good survey of research performed up to 1999 [11]. A very small number of projects have as a goal construction of an entire artificial brain or cortical columns consisting of many neurons, either from general-purpose computers [3] or more specialized architectures such as cellular automata [12] or asynchronous ARM processors [13]. Finally, there are researchers that focus on specific brain structures like the retina, or applications, like image recognition. While many neurons in the literature have some biomimetic features [e.g. 14, 15, 16, 17, 18], the complete range of neural variations has not been implemented in a single model or even in the variety of neuron models distributed throughout the research community.

The most notable research on artificial neurons includes Mead's artificial retina [2]. This significant body of work originated with Mahowald and Mead's pioneering research [19]. Boahen, who studied with Mead, also concentrates on retinal processing, including the visual cortex [20]. Many researchers describe analog neural circuits, with only a few describing mixed signal circuits. The closest electronic models to biological neurons are the mixed-signal models, including Liu and Frenzel's spike train neuron [17], Pan's bipolar neuron [21] and the cellular neural network research by Chua et al. [22, 23, 24]. Chiju et al. extends the CNN work and tests their neural model on specific applications [25]. Sato et al. [26] use stochastic logic to obtain analog behavior from digital circuits. Chen and Shi [27] use pulse width modulation. Linares-Barranco et al. [28] describe a CMOS implementation of oscillating neurons. Fu et al. [29] present thin-film analog artificial neural networks.

A basic CMOS neuron with learning capabilities is found in Chao's MS thesis [30]. The basic neural structure, the *Parker* neuron, was designed by the author, and the learning circuitry was researched by Chao. Fig. 1a illustrates the original dendrite circuit, while Fig. 1b illustrates a recent enhancement of the dendrite circuit to accept spike train inputs. Fig. 1c provides a layout of the Parker neuron with two synapses. Other noteworthy neurons capable of learning have been proposed [31, 32, 33, 34, 35, 36, 37, 38,39] Koosh and Goodman [32] put a digital computer in the loop for training, control and weight updates, and the neural network is analog, a style realized by several research groups. Commercial neural networks incorporating learning are available, albeit only weakly bio-mimetic, and are in use by the high-

energy physics community.

A recent paper by Wells [40] proposes a neurocomputer architecture intended to solve the problems of interconnectivity, variable synaptic weights and learning, issues not solved completely by any electronic neuron models published to date. Moravec [41] has performed predictions of when inexpensive general-purpose computers will match the human brain in processing power, although his predictions seem optimistic based on our own estimates. Finally, Jeff Hawkins, inventor of the PalmPilot, presents an eloquent if somewhat informal discussion of the brain [4] that motivates our bio-mimetic assumptions.

Elias [15] has performed modeling of dendritic trees that are similar to our models. The primary difference involves his use of resistors and capacitors, and our use of transistors. In addition, his synapses involve single transistors, and ours account for variability in neurotransmitter concentrations and learning. His model is more analog than ours, while ours is more of a pulse and timing circuit.

5. Size and Speed Predictions of Biomimetic Artificial Brain Structures

5.1 Prediction of Individual Neuron Properties

CMOS technology is a mature technology that allows us to predict the construction of biomimetic neurons with complexities beyond current artificial neuron circuits. We began our CMOS brain prediction by designing and laying out prototype VLSI mixed-signal neural circuits and using the transistor counts of these circuits as a base. While digital, analog and mixed-signal circuits are all represented in the literature, the focus of this prediction research will be on mixed-signal circuits, each of which represents a single neuron.

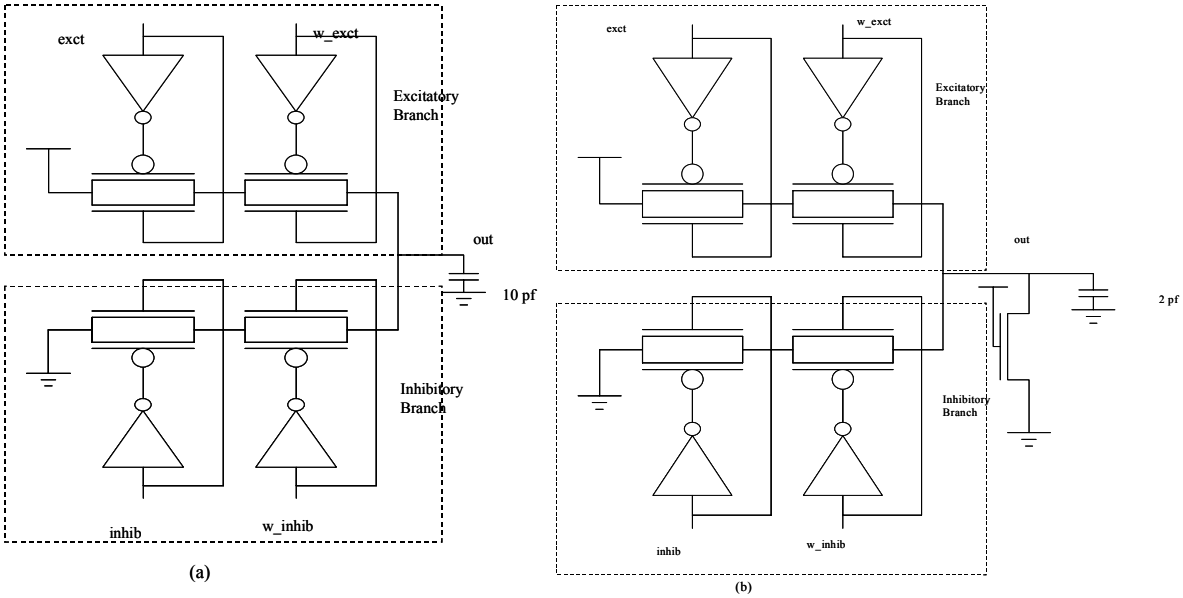
These circuits form cores comprising systems on chip, connected by on-chip buses and networks. First, we assume that the cores use conventional CMOS technology, based on the International Technology Roadmap for Semiconductors [42], and predict area, delay and power to 2019, when the technology roadmap ends.

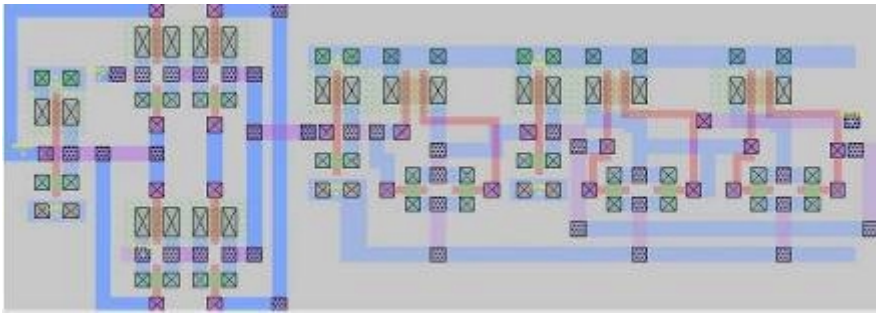
We begin our CMOS prediction using an accurate biomimetic neural electronic circuit, the Liu-Frenzel neuron [17], along with our more area-efficient basic neuron (Figure 1) and then add capabilities like dendritic branches with location-specific behavior. Thus, we will predict structures composed of two neural designs, a design representing a possible lower bound on area and one that represents a more complex design. Our layouts of portions of the Liu and Frenzel neuron, verified by SPICE, are shown in Figure 2. A canonical neuron with dendritic branches with location-specific synaptic behavior [15] is shown in Figure 3a. We consider in our predictions a simplified CMOS implementation of the dendritic branch as shown in Figure 3b. The many possible variations of inhibition [43] are not shown. This particular synaptic structure is taken from the *Parker* neuron [30]. The Liu-Frenzel neuron [17] could have been extended in a similar manner.

We have extended the Parker neuron and the spike train neuron described in Liu and Frenzel [17] to accept a sequence of spike inputs as activation, with the likelihood of activation increasing with increasing spike frequency and duty cycle, as shown in Figure 1b. This is performed by collecting charge on the gate of a synaptic transistor with the occurrence of each spike, and leaking the charge through a high-impedance transistor channel so that infrequent and/or narrow spikes, those below a certain threshold of frequency and/or duty cycle, would not cause a charge accumulation, while frequent, broad spikes would. The leakage transistor could be tuned to discharge at a faster or slower rate depending on its gate voltage, emulating the effects of chemical transmitters on the likelihood of activation. SPICE simulations have verified this straightforward structure.

We predict that CMOS brain structures will be constructed with branching structures like those illustrated in

Fig. 3b to implement a variety of dendritic fan-in structures. Consider a fan-in synaptic connection as described by Shepherd [5], where a single axon forms multiple presynaptic connections to the same postsynaptic connection. The transistor implementation of that particular situation could consist of multiple parallel PMOS transistors with sources tied to Vdd, and drains tied together representing the same dendritic branch or spine, and all with the same input. Fan-out could be supported as in conventional CMOS circuits by resizing of output driving transistors or additional buffers that support increased capacitance caused by the fan-out. Therefore, our future predictions will consider such fan-out support.





(c)

Figure 1: A basic Parker CMOS neuron dendrite (a) The original dendrite circuit (b) The dendrite circuit modified to accept spike train inputs (c) Layout of Parker's original neuron

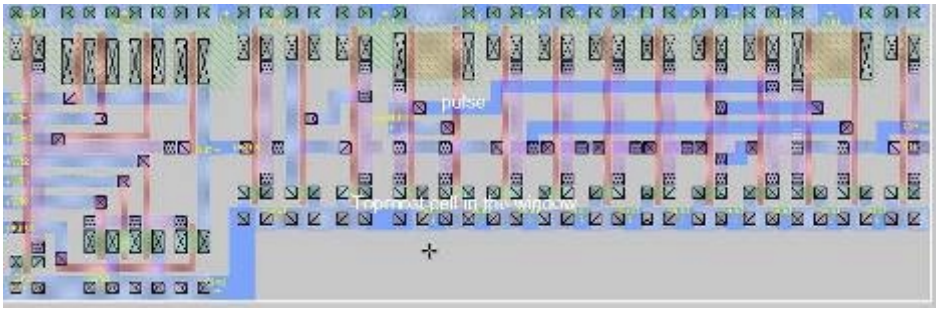


Figure 2: Our layout of the Liu-Frenzel neuron

The effect one axon has on a neighboring one can be emulated via capacitive coupling. Large metal pads overlaid on adjacent metal layers and connected to the axons of neighboring neurons can cause neighboring neurons to exhibit EPSP and possibly firing (activation) when the original neuron fires. Such capacitors could conceivably be implemented on higher levels of metal without affecting the basic neural structure on the substrate, and hence will not be included in neural area predictions. However, these coupling capacitors might affect wiring space.

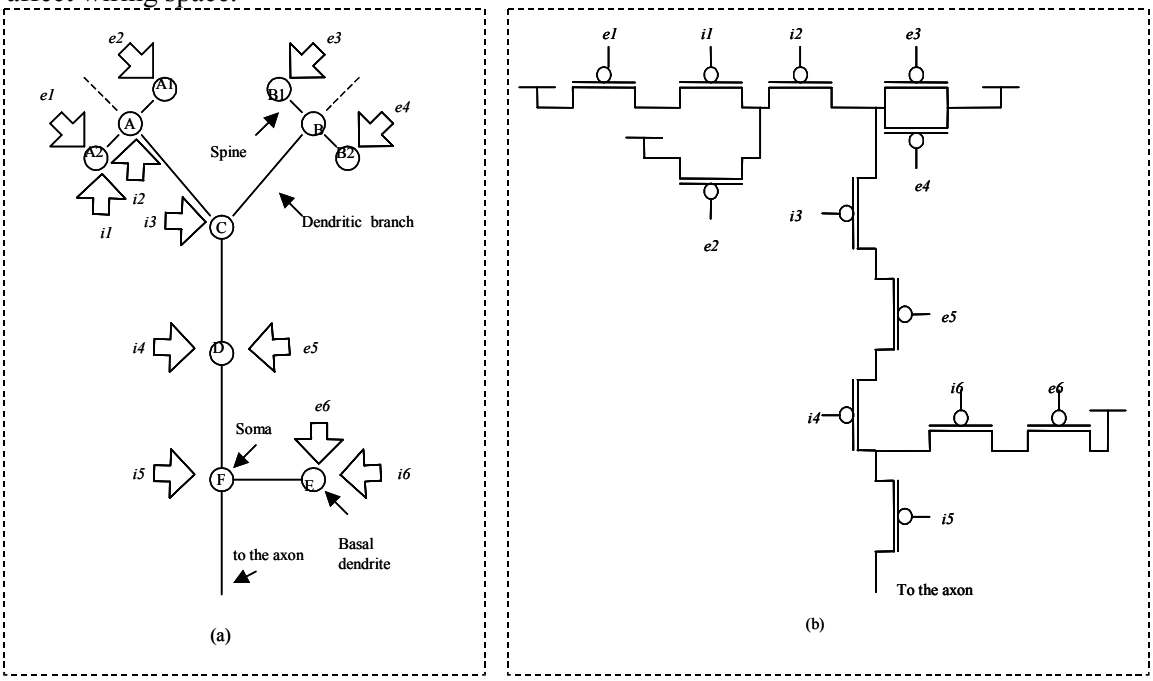


Figure 3: (a) Abstraction of the dendrite tree of a cortical pyramidal neuron [5]. (b) Possible CMOS implementation of the dendrite tree.

5.1.1 Estimating the number of transistors in a synapse

The simplest possible synaptic structure is a single MOS transistor. However, additional transistors would be required to model memory, learning, neurotransmitters, the refactory period, and other phenomena that complicate the behavior of the synapse. While the biological mechanisms underlying the synaptic behavior are complex and not completely understood, we are making an assumption that charge storage can be used to model a wide range of phenomena, like the influence of neurotransmitter concentrations and the impact of learning. Additional elements can be used to model dynamic behavior such as the refactory period. Therefore, our estimate of synaptic complexity at this point is a simple order of magnitude more complex than

the simplest synapse, or 10 transistors. The electronic synapses in the literature fall well under this estimate.

5.1.2 Estimating Space Occupied by Neurons

We are assuming that the number of transistors on a chip doubles every 18 months, assuming Moore's law holds until 2021. Gordon Moore himself predicts it will hold until 2013 [44]. We are assuming present day that there are 1 billion transistors per chip. Intel has reported on a dual core chip, Montecito, with 1.67 billion transistors and dimensions 27.72 mm x 21.5 mm, for a total area of 596 mm^2 in 90 nm technology [45]. Since the chips we are proposing are custom logic rather than processor/cache configurations, we make a more conservative estimate of 1 x 10^9 transistors/chip. The reasoning here is that custom neurons could be more difficult to lay out, and therefore a die of identical size would contain less transistors. We are assuming constant die size of 2.5 cm x 2.5 cm over the 15 years.

Let ψ_{init} be the number of transistors per chip initially [45], and ψ_{fin} be the number of transistors to be predicted for 2021, 15 years from now. $\psi_{init}=10^9$ transistors per chip and,

$$\psi_{fin} = \psi_{init} * \chi(M) \text{ transistors/chip,}$$

where χ is the multiplicative factor, and M is the number of months between initial and final estimates.

$$Y = M/R \text{ so } \chi(M) = 2^{(M/R)},$$

where Y is the number of times the transistor count doubles, and R is the number of months elapsed for the transistor count to double. Assuming doubling in 18 months, and assuming we look 180 months (15 years) into the future, 2021, the transistor counts should have doubled 10 times.

$$\chi(180) = 2^{(180/10)} = 2^{10} = 1024$$
$$\psi_{fin} = \psi_{init} * \chi(180)$$
$$\psi_{fin} = (10^9 \, transistors/chip) * 1024$$
$$\psi_{fin} \approx 10^{12} \, transistors/chip$$

Therefore, we predict that in 15 years (180 months) there would be 10^{12} transistors/chip.

We assume that on the average, each cortical neuron has 10,000 synapses [5, 4]. If each synapse contains 10 transistors, plus the few transistors representing the axon, the total number of transistors in a neuron is 1 x 10^5. Therefore, the number of neurons per chip can be estimated to be

$$\frac{(10^{12} \text{ transistors/chip})}{(10^5 \text{ transistors/neuron})} \approx 10^7 \text{ neurons/chip}.$$

Thus, in 15 years, we will require 1 x 10^4 chips (integrated circuits) to construct an artificial brain with a 100 billion (1 x 10^{11}) neurons. Boahen predicts biomimetic chip densities will be within a factor of 10 of biological neuron density within the decade [20]; however, his estimates regarding the number of transistors per synapse differ from ours.

There is more uncertainty when estimating system size because the future of multi-chip modules (MCMs) is

less predictable. However, we will assume MCM's occupy 30% of board space. On each MCM, we assume dies occupy 70% of the space. We assume that 6 12" x 12" boards fit vertically in a 1 ft^3 of space. Then the system could hold 180 chips/ ft^3. Based on these assumptions, in the year 2021, we require a space of 55.5 ft^3 to house our neural circuits, absent any interconnections. If the racks are 8 ft. tall, then we require 6.9 sq. feet of space for the equipment. Allowing for air space around the equipment, we estimate the neurons in the artificial brain to occupy 14 sq. ft. This is approximately the total free space available in the first author's university office.

5.2 Predicting Interconnections

The massive interconnectivity challenge with on- and off-chip networks to connect neural circuits bundled into on-chip cores will now be examined, and interconnection feasibility predicted. If 1 x 10^7 neurons/chip were possible in 15 years, the question is how these neurons could be organized so that interconnectivity could be possible. This problem is sufficiently complex that an immediate answer is not possible. However, some speculation is indeed possible.

Estimates of chip size assume that there will be 100 cores on a chip in approximately 6 years [46], so we can safely assume that we will be able to build chips with 100 cores in 15 years. (Each core on a conventional CMOS integrated circuit is a collection of circuits that comprise an independent executing environment, like a processor.) This implies that ***each core would contain 1 x 10^5 neurons***. Based on this model, assuming that the core size is 2.5 mm on a side, we can assume that 100 cores will fit on a 2.5 cm x 2.5 cm die. Therefore, each neuron, of the 10^5 in each core, occupies 62.5 μ^2 of die area.

Massive interconnections characterize brain structures, particularly in the cerebral cortex. Using a pyramidal windowing structure we designed (Figures 4 and 5) we can make a predictable number of connections to an area of silicon that is a function of the silicon area and the available layers of metal. The number of window layers depends on the number of available metal interconnect layers. The number of connections possible to a core H by H in area can be computed with Equation 1 [47]:

$$C_{total} = \sum_{i=\alpha}^{\omega} C_i = floor\left(\sum_{i=\alpha}^{\omega} 4 \left(\frac{(H - 2(E + D_{\alpha-1}) - \sum_{j=\alpha}^{i-1} 2L_j - D_i}{D_i + S_i} \right) \right), \quad (1)$$

where ω-α is the number of available metal interconnect layers, E is the required spacing from the edge of the core to the lowest (outer) layer of metal contacts, D_i is the required width of contacts on layer i, and S_i is the spacing required between contacts terminating on layer i and layer $i + 1$. S_i is the contact spacing on the ith layer. The variable α represents the first layer of metal allowed for inter-neural connections.

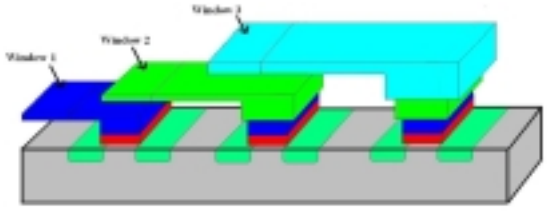


Figure 4. The pyramidal window structure [47]

For our predicted technology in 2021, we assume all metal layers have the same design rules as metal 1 (the

metal layer closest to the substrate). Therefore, in 2019, D_i= D =23.3 nm[1], L_i=14 nm and S_i=14 nm [42]. A rudimentary calculation, assuming all metal layers have the same parameters as the first metal layer, shows that less than 6000 connections can emerge from the projected neuron area, 62.5 μ^2. If each neuron has 10,000 synapses, and a fan out of 10,000, this projected neural area cannot support the number of necessary point-to-point connections. Therefore, alternative means of neural communication are necessary. Boahen gives a firing estimate of 10^{16} "activations" a second. While electronics can certainly match the speed of neurons, and their communication with other neurons in real time, scaling up emulation to millions or billions of neurons poses a communications problem. Point-to-point connections are impractical for all but the closest neurons. If conventional CMOS is employed, some sharing of the interconnection media must occur.

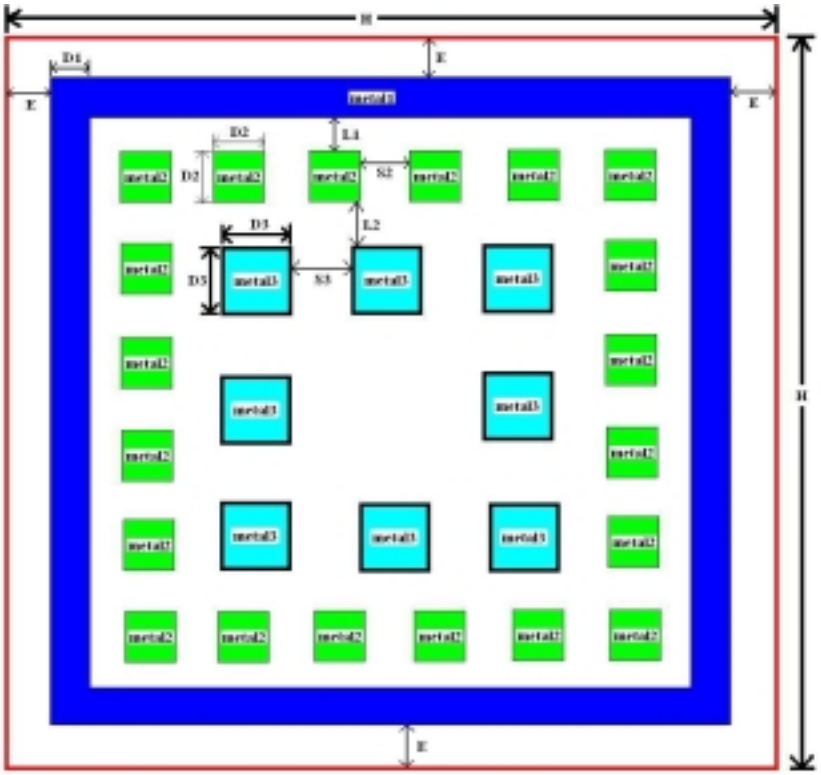


Figure 5. The arrangement of contacts for maximum connections to a given area [47].

For our estimations of communication complexity, we can assume that the axon of a neuron firing fans out to about 10,000 synaptic connections [5]. The "distance" these connections are from the soma can be modeled as an exponential distribution, with the axon more likely to have presynaptic terminals close to the soma. Experimental results [48] support this assumption. The complication is that Singh's experimental data is for the 3-dimensional human brain structure, and we would like to assume that the exponential nature of the axon length data holds as we map biological neurons in 3D space onto the substrates of chips, modules and boards. Unfortunately, this differential geometric mapping is a difficult one to describe mathematically, so further study is required.

With this assumption, we are going to further assume that connections within the core are point-to-point.

[1] MOSIS deep submicron rules: assuming 14 nm is 3 λ, deep submicron rules state that vias are 3 λ x 3 λ wide and there is a 1 λ border of conducting material around the via. Therefore D is 5* λ.

There are 10^5 neurons/core, and 10^4 synaptic connections each axon makes. So there are 10^9 connections per core. Assuming there are going to be 14 metal layers in 2021 [42], we propose allocating layers one through five to intra-neuron connections, including power and ground wiring, layers six through 12 to intracore connections and the final two layers to inter-core connections. Assuming wire lengths can be modeled using a Poisson distribution [49], most of these connections are short, and hence we assume most neural connections will be within a single core.

Inter-core communications will probably of necessity be shared. One could make the same arguments here for networks on chip (NOC) that DeMicheli has made for conventional CMOS integrated circuits [50]. Such networks are typically packet switching networks. and have been reported to occupy about 6.6% of the layout area of example integrated circuits [51]. However, due to the decreasing traffic as distance increases between neurons, Such a packet-switched network could become unnecessary for long connections and occupy more substrate area than necessary. A lower-bandwidth network might be appropriate for more distant on-chip connections, creating a hierarchy of interconnection schemes, as proposed by Raghavan [9].

A leading candidate for such a network is a modified torus topology with token ring protocol that we have been researching, as shown in Figure 6 [9]. For our token ring design, the chip area occupied by the token ring is estimated to be ~ .5% of the total chip area [52]. Some trivial modifications of the token ring NOC to reflect smaller payloads will be researched. One possibility is to make every synaptic connection addressable, resulting in a large address space, subdivided into intra-core, inter-core, inter-chip, inter-module and inter-board fields. For very large systems, inter-board and inter-rack fields could also be used to extend the address space. Spike trains entering the network would be encoded digitally with firing duration, duty cycle and spike frequency fields, and decoded to analog spike trains at each addressable synaptic connection.

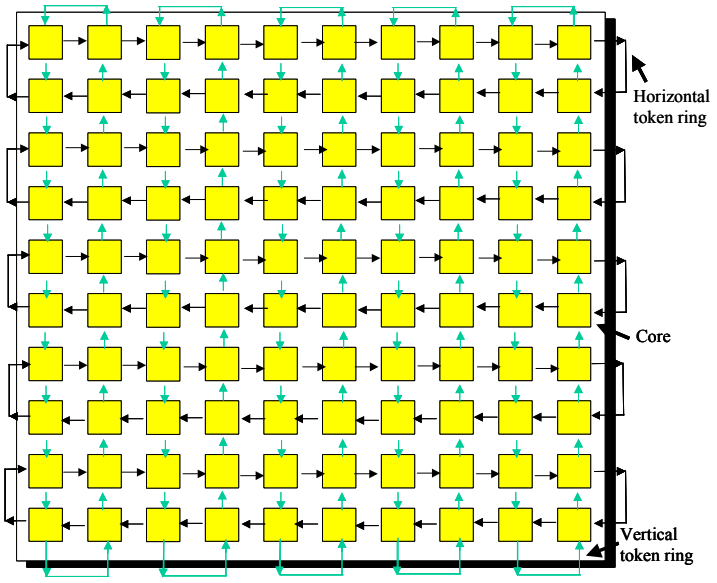


Figure 6. Modified token ring network with neural cores.

6. Conclusions

Our biomimetic neural size estimates have barely begun to incorporate interconnections; however, it is obvious that interconnection issues in the artificial brain pose a vast technological challenge. Research on potential interconnection strategies is continuing within the USC artificial brain project and in other research groups.

At the beginning of this paper, we posed a question as to when technology will support the construction of an artificial brain. The predictions made in this paper, looking 15 years into the future, are highly uncertain and many issues are not likely to be resolved even in the next decade, as we continue to investigate this question. This paper represents a beginning and not a final answer. The hope is that researchers will be inspired by the questions raised in this paper and will join us in this exploration.

7. Acknowledgements

The authors are indebted to the many directed research students who assisted in the research described in this paper, in particular Ko-Chung Tseng, Nattapak Atichartakarn, Aman Khan, Sangil Kim, Sim Ji Lee, and Pocholo Pasicolan.

8. References

1. Bota, M., Dong, H-W, and Swanson, L., "From gene networks to brain networks," *nature neuroscience*, Vol. 6, No. 8, Aug., 2003.

2. Mead, C, *Analog VLSI and Neural Systems,* 1989, Addison-Wesley Longman, Boston.

3. Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne, http://bmi.epfl.ch/

4. Hawkins, J., and Blakeslee, S., *On Intelligence*, Times Books, New York, 2004.

5. Shepherd, G. "Introduction to Synaptic Circuits," in *The Synaptic Organization of the Brain*, edited by Gordon Shepherd, 5th edition, Oxford University Press, 2004.

6. Calvin, W.H. and Graubard, K., "Styles of Neuronal Computation" in: *The Neurosciences Fourth Study Program*, edited by O. Schmitt and F.G. Worden, Cambridge: MIT Press, 1979 (Table II page 521).

7. Tansley, Brian, "Is the Computer Cognitive Science's Lamppost?" http://http-server.carleton.ca/~btansley/aboutme/mypointofview.html.

8. Mel, B and Schiller, J., "On the fight between excitation and inhibition: Location is everything," *Science's STKE*, Sept. 2004.

9. Raghavan, D., *Extending the Design Space for Networks on Chip*, MS Thesis, University of Southern California, May, 2004.

10. Boahen, K., "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process*., Vol. 47, no. 5, pp. 416-434, May, 2000.

11. Schüffny, R. et al., "Hardware for Neural Networks, " *4th International Workshop Neural Networks in Applications NN '99*, March 1999.

12. De Garis, H., "CAM-Brain ATR's Billion Neuron Artificial Brain Project – A Three Year Progress Report," *International Conf. on Evolutionary Computation*, 1996, pp. 886-891.

13. Furber, S., http://www.cs.manchester.ac.uk/apt/people/sfurber/

14. Reyneri, L. "On the Performance of Pulsed and Spiking Neurons", *Analog Integrated Circuits and Signal Processing*, Kluwer Academic Publishers, The Netherlands, 2002, pp. 30, 101-119.

15. Elias, J.G., Chu, H-H, and Meshreki, S.M., "Silicon implementation of an artificial dendritic tree," *IJCNN, International Joint Conference on Neural Networks, 1992,* Volume 1, pp. 154-159.

16. Murray, A., Corso, D., Tarassenko, L., "Pulse-stream VLSI Neural Networks Mixing Analog and Digital Techniques", *IEEE Transactions on Neural Networks*, March 1991

17. Liu, B., and Frenzel, J.F., A CMOS neuron for VLSI implementation of pulsed neural networks, *Proc. 28th Ann. Conf. Ind. Electron. IECON0*2, Nov. 5-8, Sevilla, Spain, pp. 3182-3185, 2002.

18. Farquhar, E., and Hasler, P. "A Bio-Physically Inspired Silicon Neuron," *IEEE Transactions on Circuits and Systems*, Vol. 52, No. 3, March 2005, pp. 477-488.

19. Mahowald, M., "VLSI analogs of neuronal visual processing: a synthesis of form and function," Ph.D. dissertation, California Institute of Technology, Pasadena, 1992.

20. Boahen, K., Neuromorphic microchips, *Scientific American*, May, 2005, pp. 56-63.

21. Pan, Dong and Wilamowski, B.M., "A VLSI implementation of mixed-signal mode bipolar neuron circuitry," *Proceedings of the International Joint Conference on Neural Networks*, vol.2, pp. 971-976 vol.2, 20-24 July 2003

22. Chua, I. L. O, and Yang, I., "Cellular Neural Networks: Theory" and "Cellular Neural Networks: Applications" *IEEE Trans. on CAS* vol. 35, pp. 1257-95 Oct. 1988.

23. Chua, L.O., "CNN Chips Crank up the Computing Power," *IEEE Circuits and Devices*, July 1996, vol. 12, no. 4, pp. 18-27.

24. Chua L.O. and Roska, T. "The CNN paradigm," *IEEE Trans. CAS*, vol. 40, no. 3, 1993, pp. 147-156.

25. Chiju, C., et al., "Analysis and Performance of a Versatile CMOS Neural Circuit based on Multi-Nested Approach,"*7th IEEE International Symposium on Signals Circuits and Systems*, July, 2005, pp. 417-420.

26. Sato, S.; Nemoto, K.; Akimoto, S.; Kinjo, M.; Nakajima, K., "Implementation of a new neurochip using stochastic logic," *IEEE Transactions on Neural Networks*, vol.14, no.5, pp. 1122- 1127, Sept. 2003.

27. Chen, Lu and Shi, Bingxue, "Building blocks for PWM VLSI neural network," *5th International Conference on Signal Processing Proceedings WCCC-ICSP 2000*, vol.1, no.pp.563-566, 2000.

28. Linares-Barranco, B., Sanchez-Sinencio, E., Rodriguez-Vazquez, A., and Huertas, J.L., "A CMOS implementation of FitzHugh-Nagumo neuron model," *IEEE Journal of Solid-State Circuits*, vol.26, no.7, pp.956-965, July 1991.

29. Fu, C. et al., "A Novel Technology for Fabricating Customizable VLSI Artificial Neural Network Chips," *IJCNN,* 1992.

30. Chao, C. "Incorporation of Learning within the CMOS Neuron", University of Southern California M.S. Thesis, July, 1990.

31. Liu, J. and Brooke, M., "Fully parallel on-chip learning hardware neural network for real-time control," *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems ISCAS '99*, vol.5, pp.371-374, 1999.

32. Koosh, V.F., and Goodman, R., "VLSI neural network with digital weights and analog multipliers," *Proceedings of the 2001 IEEE International Symposium on Circuits and Systems ISCAS 2001*, vol.3, pp.233-236 vol. 2, 6-9 May 2001

33. Gerstner, W., and Kistler, W. *Spiking Neuron Models*, Cambridge University Press. 2002.

34. Haykin, S., *Neural Networks, A Comprehensive Foundation*, 2 Ed, Prentice-Hall, Inc, New Jersey, 1999.

35. Mass, W., and Bishop, C., *Pulsed Neural Networks*, The MIT Press, Massachusetts, 1999.

36. Perez-Uribe, A., Structure-*adaptable digital neural networks*, PhD thesis, 1999, EPFL.

37. Ros, E., Agis, R., Carrillo, R., and Ortigosa, E., "Post-synaptic Time-Dependent Conductance in Spiking Neurons: FPGA Implementation of a Flexible Cell Model," *Proceedings of IWANN'03: LNCS 2687*, pp 145-152, Springer, Berlin, 2003.

38. Upegui, A., Pena-Reyes, C.A., and Sanchez, E., "A methodology for evolving spiking neural network topologies on line using partial dynamic reconfiguration". Submitted to *International Congress on Computational Intelligence (CIIC'03)*. Medellín, Colombia.

39. Yao, X. "Evolving artificial neural networks". *Proceedings of the IEEE*, 87(9), 1423-1447, September 1999.

40. Wells, Richard B., "Preliminary Discussion of the Design of a Large-Scale General-Purpose Neurocomputer," MRC Institute, The University of Idaho, Nov. 14, 2003 http://www.mrc.uidaho.edu/~rwells/techdocs/Preliminary%20Discussion%20of%20the%20Design%20of%20a%20GP%20Neurocomputer.pdf

41. Moravec, Hans, "*When will computer hardware match the human brain?" Journal of Transhumanism*, vol. 1, March 1998.

42. International Technology Roadmap for Semiconductor, 2005 http://www.itrs.net/Common/2005ITRS/Interconnect2005.pdf

43. Cohen, E., and Sterling, P. (1991). "Microcircuitry related to the receptive field center of the ON-beta ganglion cell." *J. Neurophysiol.* Vol. 65, pp. 352-359.

44. Kanellos, Michael, "Moore's Law to roll on for another decade," CNET News.com February 10, 2003, 2:27 PM PST http://news.com.com/2100-1001-984051.html

45. Intel Corporation, http://www.intel.com/pressroom/kits/events/idfspr_2006/20060313_multicore_fact_sheet_decoder.pdf

46. Gain, B., "Faster Chips, Kill, Kill, Kill," Wired, Feb. 24, 2006, http://www.wired.com/news/technology/0,70285-0.html?tw=wn_index_11

47. Tseng, C-C, "Estimation of maximum connections for CMOS neuron chip design," http://eve.usc.edu/tseng.pdf, Directed Research Report, Electrical Engineering Systems Department, University of Southern California, 2005.

48. Singh, M., Hwang, D., Sungkarat, W., and Veera, K., Evaluation of MRI DTI-tractography by tract-length histogram, *Progress in Biomedical Optics and Imaging: Physiology, Function and Structure from Medical Images*, 5746(1): 138-147, 2005.

49. Kurdahi, F, and Parker, A., "Techniques for Area Estimation of VLSI Layouts," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, January 1989, pp. 81-92.

50. Benini, L. and De Micheli, G., "Networks on chip: a new paradigm for systems on chip design*", In Proceedings of Conference on Design, Automation and Test in Europe*, pages 418-419, 2002.

51. Dally, W., and Towles, B., "Route Packets, Not Wires: On-Chip Interconnection Networks," *Design and Automation Conference*, Las Vegas, USA, 2001.

52. Raghavan, D., Patel, J., Mittal, R., and Parker, A., "Exploring the Design Space for Networks on Chip," working paper, http://eve.usc.edu/raghavan.pdf