or event); the cerebellum should not do this sort of completion, because a "partial" cue corresponds to a different physical situation in which different motor actions might be appropriate. The hippocampus is thought to exhibit one-trial learning (necessary for episodic memory), while the cerebellum is thought to be an iterative learner.

An important feature of the hippocampus that is missing from the cerebellum is the recurrent connections in CA3, which aid pattern completion by iterative refinement. There is no corresponding recurrent pathway in the cerebellum. On the other hand, the cerebellum appears to have a direct teaching input: the climbing fibers that make powerful synapses on Purkinje cells. Hippocampus does not have such an input.

The dentate gyrus in the hippocampus is thought to orthogonalize input patterns by projecting subsets of active input lines (codons) into a higher dimensional space. A similar role is played by the granule cells in the cerebellum, which project mossy fiber input patterns into a higher dimensional space in the parallel fibers.

10. *We studied three different representations for vectors in the nervous system: maps, cosine-tuned cells, and basis functions. Briefly describe each representation and one model that uses it.*

MAPS: Each possible vector (x,y) is represented by a cell c. In the distributed approximation, each cell c covers a small receptive field in the 2D space. These cells may be topographically organized or not. The superior colliculus models (Arai et al., van Opstal and Kappen) used topographically organized maps, while the place cell models (Burgess et al., Redish and Touretzky) used non-topographically organized maps.

COSINE TUNED NEURONS: Each cell has a cosine tuning to ϕ and a linear tuning to r for each vector (r,ϕ) in polar coordinates. The Redish and Touretzky model used cosine tuned neurons to represent vectors in a motor task.

BASIS FUNCTIONS: Each cell firing activity is a function of (x,y) coordinates of the vector such that the population of all cells forms a basis set. Typical basis sets are gaussians and sigmoids. A map in which each cell has a gaussian tuning to (x,y) coordinates is a special case of a basis function representation. The main advantage of a basis function representation is that any function of (x,y) can be represented as a linear combination of basis functions. The Pouget and Sejnowski model of coordinate transformations uses basis function representations of vectors.

tems make use of similar principles in their computations.

On the other hand, the systems are different in that computations in the fly visual system are fully grounded in neuroanatomy: there are only a few types of cells involved: photoreceptors and a couple of layers of interneurons, and the connectivity of the necessary circuitry is known. In contrast, the mammalian system can only be modeled at a more abstract level. There are many more layers of circuitry between the photoreceptors and motion-sensitive cells in visual cortex. The anatomy and physiology are not fully known; even the functioning of simple cells (edge detectors) is not fully understood at present. Therefore, the Adelson & Bergen model is less well grounded than that of Franceschini et al., an unavoidable consequence of our limited anatomical knowledge at the present time.

8. *Real brains probably don't use the backpropagation learning procedure, but we looked at several models of biological systems that did. Describe two such models and explain why they are relevant even though backpropagation was used, or criticize them if you think the use of backprop was detrimental.*

Backprop is only useful for neuroscience as a mechanism for generating a neural net. A neural net generated with backprop can be explored as an instance of a transformation, but one cannot look at the learning involved. Two models that used backprop were the Zipser and Andersen model of parietal cortex and the Lehky and Sejnowski model of shape from shading.

The Zipser and Andersen model trained a three layer neural net with backprop to transform a retinotopic representation (an array of gaussian tuned (retinotopic) neurons) and a eye-position representation (piecewise linear eye-position-sensisitive neurons) into a representation of cranio-topic position. They make two claims based on this work: (1) that the neurons in the hidden layer show similar firing patterns to neurons in monkey parietal area 7a and (2) that area 7a neurons include enough information to represent cranio-topic position even though there is no topographic map of craniotopic coordinates.

Backprop was not well used by Zipser and Andersen because their comparison method for claim 1 was subjective. Because of this, they could not make any claims about real neurons. If they had used real neural firing patterns (which they had recorded from 7a) as inputs to a neural net and then showed that it could be trained to produce craniotopic coordinates, then they would have proved claim 2.

Lehky and Sejnowski used backprop to train a three layer net to perform shape from shading and found that the hidden units looked (again qualitatively) like edge detectors. Their use of backprop was justified because shape from shading is still unsolved analytically. Therefore, they had to use backprop (or another learning mechanism) just to get a working algorithm. The only claim made was that edge detectors might not be what they seem, that their true function may be dependent on their outputs (as well as their inputs).

Their claim would have been stronger made if they had used the same hidden units to do edge detection and showed that the same units could indeed be used for two processes given two different output projections.

9. *Both the hippocampus and the cerebellum have been described as associative memories. In what ways do these memories differ in their operating characteristics? Which broad architectural features of the hippocampus are missing from the cerebellum, and vice versa? What cerebellar structures play a similar computational role to the dentate gyrus? Justify your response.*

The hippocampus is believed to store memories briefly, while awaiting transfer to cortex; the cerebellum is a long-term, permanent memory. The hippocampus is supposed to permit associative retrieval based on partial cues (e.g., an odor might cause us to recall a particular place

(A-2) Place cells could be special cases of general sensory-configuration cells. Pyramidal cells in hippocampus also respond to non-place sesnory information, such as odors (Eichenbaum et al. 1987). This theory says that place cells are an epiphenomenon.

(A-3) Place fields in the dark could be driven by completion of patterns partially instantiated by other cues.

(A-4) In most instances of this theory, the hippocampus is a temporary memory store, used to train the long term memory store in cortex. Anterograde amnesia can be explained by an inability to store the memory in the temporary memory store.

(B-1) Recurrent connections in CA3 can be used to complete a representation of a place when some cues are missing.

(B-3) Some instantiations of this theory say that place fields in the dark are driven by pattern completion driven by non-visual cues, but other instantiations suggest that place cells in the dark imply that the place code receives information from a path integrator.

(B-4) This theory has trouble with the anterograde amnesia data, but anterograde amnesia could be explained in the following way: Every memory must be localized in space. Without a cognitive map to place the memory in space, the memory cannot be stored.

6. *Compare the Granger & Lynch model of piriform cortex with the Wilson & Bower model. What does each accomplish? How do the approaches and goals of these modeling efforts differ?*

The goal of the Granger & Lynch model is to promote a functional hypothesis: that piriform cortex performs hierarchical pattern discrimination. They demonstrate this in a computer simulation that draws upon the anatomy and physiology of the piriform cortex. But they also ignore many important features of the real system, and introduce assumptions (such as a 1000 msec refractory period) that are difficult to justify.

Wilson & Bower's goal is to accurately model the behavior of the piriform cortex, and explain the origin of the oscillations observed there. They take a more bottom-up approach than Granger & Lynch, using much more realistic neuronal models and sticking closely to the neurophysiological data. They do not start with a functional hypothesis, and in the end have little to say about the function of piriform cortex beyond suggesting some associative memory role. However, they do shed some light on the oscillation question, suggesting that oscillations are an intrinsic property of the cortical circuitry and are not necessarily involved in encoding information for computational purposes.

7. *Compare the circuitry for direction-selective motion detectors in the insect visual system with some model of motion processing in the mammalian visual system. Are the two systems fundamentally similar, or different? Justify your answer. (Either choice can lead to an acceptable answer.)*

We'll compare the Franceschini et al. model of motion detection in the fly with the Adelson & Bergen model of motion processing in mammals. The two systems are similar in that both use low-level filters to provide a rough estimate of local velocity. In the case of the fly, motion detection is done by combining direct input from one sensory cell with delayed inhibitory input from a second cell. Motion in the preferred direction will result in excitation arriving before inhibition, causing the downstream cell to fire. This gives a one-dimensional direction-selective motion detector. The mammalian model uses the same idea but with a more complex filter, involving differences of Gaussians in the spatial domain as well as excitation followed by inhibition in the time domain. A pair of these filters ''in quadrature'' (separated by 1/4 phase in space and time) can be used to construct directionally-sensitive motion detectors. The two sys-

3. *One of the benefits of computer simulations is that they can reveal the weak points in a computational theory and suggest remedies. Defend this claim using the work of Gluck & Thompson (1987) as an example.*

Hawkins and Kandel (1984) developed a theory of the neural underpinnings of associative learning in Aplysia. In their theory, habituation, sensitization, and classical conditioning are the building blocks from which more complex learning phenomena such as higher-order conditioning differential conditioning, and blocking are constructed. Gluck and Thompson (1987) simulated this model and showed that it failed to produce blocking.

The problem with the Hawkins and Kandel model was that an interneuron needed to continue firing in order to maintain a learned association (by protecting certain weights from habituating), but it needed to *not* fire (due to being in a refactory state) in order to prevent acquisition of a new association (hence, blocking.) They discussed two solutions to this problem. One was to change the learning rule to use an S-shaped learning curve, so that the interneuron need be only slightly active: enough to preserve the already learned weights but not enough to permit learning of novel associations. However, this change was not justified by physiological observations. The second proposed solution postulated another class of interneuron that functioned only to preserve already learned associations.

4. *The Hasselmo and Schnell (1994) and O'Reilly and McClelland (1994) articles both suggest mechanisms to perform pattern separation and pattern completion in the hippocampus. Are these two models compatible or incompatible? Justify your answer.*

Both models suggest that pattern completion happens in the axons leaving CA3. In the O'Reilly and McClelland model, pattern separation occurs in the Dentate Gyrus (DG) and pattern completion in CA3. Pattern separation is achieved by a massive divergence of input from a small population in entorhinal cortex to a very large, sparsly active population in DG. Pattern completion is achieved by the recurrent connections in CA3. CA1 then cleans up the CA3 pattern of activity.

Hasselmo and Schnell's model depends crucially on the fact that acetylcholine inhibits synaptic transmission from the schaffer collaterals (CA3->CA1 connections) much more than from the direct EC->CA1 connection. Because the schaffer collaterals are the same axons that form the recurrent collaterals in CA3, it is reasonable to assume that the acetylcholine also inhibits those recurrent connections. In the language of O'Reilly and McClelland, we can say that the acetylcholine turns off pattern completion. This is exactly what would be required for a novel pattern.

Both models require two pathways from EC to CA3/1, a direct and an indirect connection. They both also detail mechanisms to prefer use of one or the other of the direct connection during pattern completion and the indirect during pattern separation. The mechanisms are not logically incompatible and complement each other nicely.

5. *Briefly describe two very different theories of the function of the hippocampus. How does each theory address (or fail to address) each of the following: (1) recurrent connections in CA3, (2) the existence of place cells, (3) the persistence of place fields in the dark, and (4) anterograde amnesia following hippocampal damage.*

The two main theories of hippocampus are (A) that it is involved in declarative memory (Marr, Rolls et al., O'Reilly and McClelland) or (B) that it stores a cognitive map (O'Keefe et al., McNaughton et al., Burgess et al., Redish and Touretzky).

(A-1) Recurrent connections in CA3 help complete incomplete patterns.

Answer Key for Final Exam: Computational Models of Neural Systems Course 15-880(B): Spring 1995

David S. Touretzky

1. *Many of the models we studied include a variant of an associative or matrix memory. Describe three mechanisms for improving the function of a simple matrix memory and give an example of a model that uses each one. (Be sure to explain how the memory is improved.)*

   A simple matrix memory takes a vector as input, performs a matrix-vector multiplication, and produces a vector as output. It produces perfect responses only when the input patterns are complete, uncorrupted, and orthognal. The number of memories that can be stored is bounded by the dimensionality of the input vector.

   (a) If the output is limited to binary patterns, thresholding can improve the performance of the memory by making it less sensitive to noisy inputs. It also becomes more tolerant of not-quite-orthogonal memories. Incomplete input patterns can be accomodated by using a threshold that varies in proportion to the magnitude of the input vector. Marr's model used this method.

   (b) The capacity of an associative memory can be improved by making the input patterns more orthogonal. This is done by projecting the input space into a higher dimensional space by looking at many possible *conjunctions* of input lines. Marr called these conjunctions "codons" in his hippocampus paper. A similar trick is performed by granule cells in his cerebellar model.

   (c) Memory performance can be further improved by organizing the memory as an auto-associator (meaning the input patterns are identical to the output patterns, modulo incompleteness and noise on the input side) and adding recurrent connections. Since the output pattern will be a somewhat "cleaned up" version of the input (due to thresholding), using recurrent connections to run the output through the network again, perhaps several times, should produce further cleanup. The current connections in CA3 are thought to perform this function in the McNaughton and Morris model.

2. *Some of the work we examined treated neurons as continuous functions, while others used more realistic models with spiking neurons. List one advantage and one disadvantage of each of approach. (Don't say "more biological" with no further explanation. And don't just list the advantages of one approach as the disadvantages of the other; pick some other point to discuss.)*

   CONTINUOUS NEURONS: Advantage: faster, more efficient simulations, since fewer units can be used and we aren't forced to deal explicitly with time. Also, more amenable to theoretical analysis. Disadvantage: limited expressive power; can't model spike coincidence detection, refractory periods, and similar temporal effects that might be important in real brains.

   SPIKING NEURONS: Advantage: allows more direct comparison with experimental data and increases the chance that the model could demonstrate unexpected properties that might be shared by real brains. Disadvantage: requires many more parameter values, some of which may not be known for the biological system being modeled.