Marr's Theory of the Hippocampus: Part I

Computational Models of Neural Systems

Lecture 3.3

David S. Touretzky

September, 2021

David Marr: 1945-1980

David Marr
1970 – Cambridge, England

Marr and Computational Neuroscience

- In 1969-1970, Marr wrote three major papers on theories of the cortex:
  - *A Theory of Cerebellar Cortex*
  - *A Theory for Cerebral Neocortex*
  - *Simple Memory: A Theory for Archicortex*

- *A fourth paper, on the input/output relations between cortex and hippocampus, was promised but never completed.*
- Subsequently he went on to work in computational vision.
- His vision work includes a theory of lightness computation in retina, and the Marr-Poggio stereo algorithm.

Introduction to Marr's Archicortex Theory

- The hippocampus is in the “relatively simple and primitive” part of the cerebrum: the archicortex.
  - The *piriform* (olfactory) cortex is also part of archicortex.
- Why is archicortex considered simpler than neocortex?
  - Evolutionarily, it's an earlier part of the brain.
  - Fewer cell layers (3 vs. 6)
  - Other reasons? [connectivity?]
- Marr claims that neocortex can learn to classify inputs (category formation), whereas archicortex can only do associative recall.
  - Was this conclusion justified by the anatomy?

What Does Marr's Hippocampus Do?

- Stores patterns immediately and efficiently, without further analysis.
- Later the neocortex can pick out the important features and memorize those.
- It may take a while for cortex to decide which features are important.
  - Transfer is not immediate.
- Hippocampus is thus a kind of medium-term memory used to train the neocortex.

An Animal's Limited History

- If 10 fibers out of 1000 can be active at once, that gives C(1000,10) possible combinations = 2.6×10^{23}.
- Assume a new pattern every 1 ms.
  - Enough combinations to go for 10^{12} years.
- So: assume patterns will not repeat during the lifetime of the animal.
- Very few of the many possible events (patterns) will actually be encountered.
- So events will be well-separated in pattern space, not close together.

Numerical Contraints

Marr defined a set of numerical constraints to determine the shape of simple memory theory:

1. Capacity requirements
2. Number of inputs
3. Number of outputs
4. Number of synapse states = 2 (binary synapses)
5. Number of synapses made on a cell
6. Pattern of connectivity
7. Level of activity (sparseness)
8. Size of retrieval cue

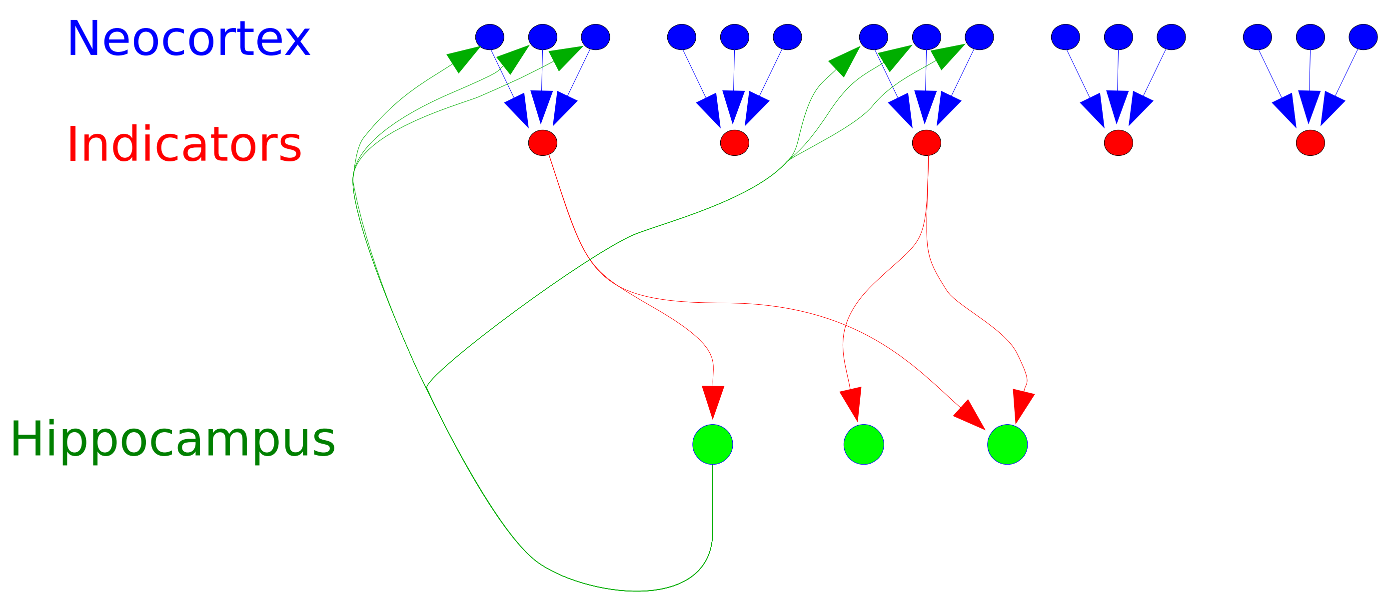
N1. Capacity Requirements

- A simple memory only needs to store one day's worth of experiences.
- They will be transferred to neocortex at night, during sleep.
- There are 86,400 seconds in a day.
- A reasonable upper bound on memories stored is:

100,000 events per day

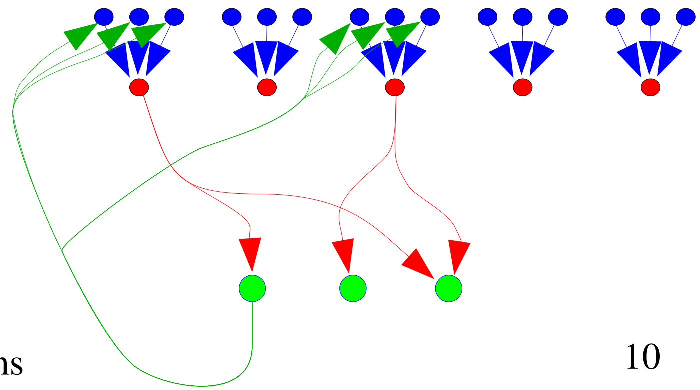
N2. Number of Inputs

- Too many cortical pyramids (10^8): can't all have direct contact with the hippocampus.
- Solution: introduce <u>indicator cells</u> as markers of activity in each local cortical region, about 0.03 mm^2.
- Indicator cells funnel activity into the hippocampal system.



Indicator Cells

- Indicator cells funnel information into hippocampus.
- Don't we lose information?
  - Yes, but the loss is recoverable if the input patterns aren't too similar (low overlap).
- The return connections from hippocampus to cortex must be direct to all the cortical pyramids, not to the indicator cells.
- But that's okay because there are far fewer hippocampal axons than cortical axons (so there's room for all the wiring), and each axon can make many synapses.



How Many Input Fibers?

- Roughly 30 indicator cells per mm^2 of cortex.
- Roughly 1300 cm^2 in one hemisphere of human cortex, of which about 400 cm^2 needs direct access to simple memory. Thus,

About 10^6 afferent fibers enter simple memory.

- This seems a reasonable number.

N3. Number of Outputs

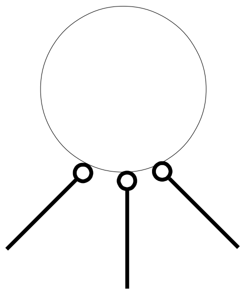
- Assume neocortical pyramidal cells have fewer than 10^5 afferent synapses.
- Assume only about 10^4 synaptic sites available on the pyramidal cell for receiving output from simple memory.
- Hence, if every hippocampal cell must contact every cortical cell, there can be at most 10^4 hippocampal cells in the memory. Too few!
  - If 100,000 memories stored, each memory could only have 10 cells active (based on the constraint that each cell participates in at most 100 memories.) Too few cells for accurate recall.
- Later this constraint was changed to permit 10^5 cells in the simple memory.

N4. Binary Synapses

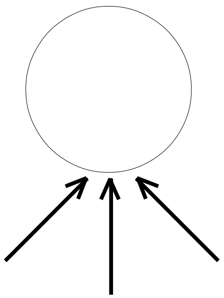
- Marr assumed a synapse is either on or off (1 or 0).
- Real-valued synapses aren't required for his associative memory model to work.
  - But they could increase the memory capacity.
- Assuming binary synapses simplifies the capacity analysis to follow.

Types of Synapses

- Hebb synapses are binary: *on or off.*
- Brindley synapses have a fixed component in addition to the modifiable component.



Hebb synapses



Brindley synapses

- Synapses are switched to the *on* state by simultaneous activity in the pre- and post-synaptic cells.
- This is known as the Hebb learning rule.

N5. Number of Synapses

- The number of synapses onto a cell is assumed to be high, but bounded.
- Anatomy suggests no more than 60,000.
- In most calculations he uses a value of 10^5.

N6. Pattern of Connectivity

- Some layers are subdivided into blocks, mirroring the structure of projections in cortex, and from cortex to hippocampus.
- Projections between such layers are only between corresponding blocks.
- Within blocks, the projection is random.

N7. Level of Activity

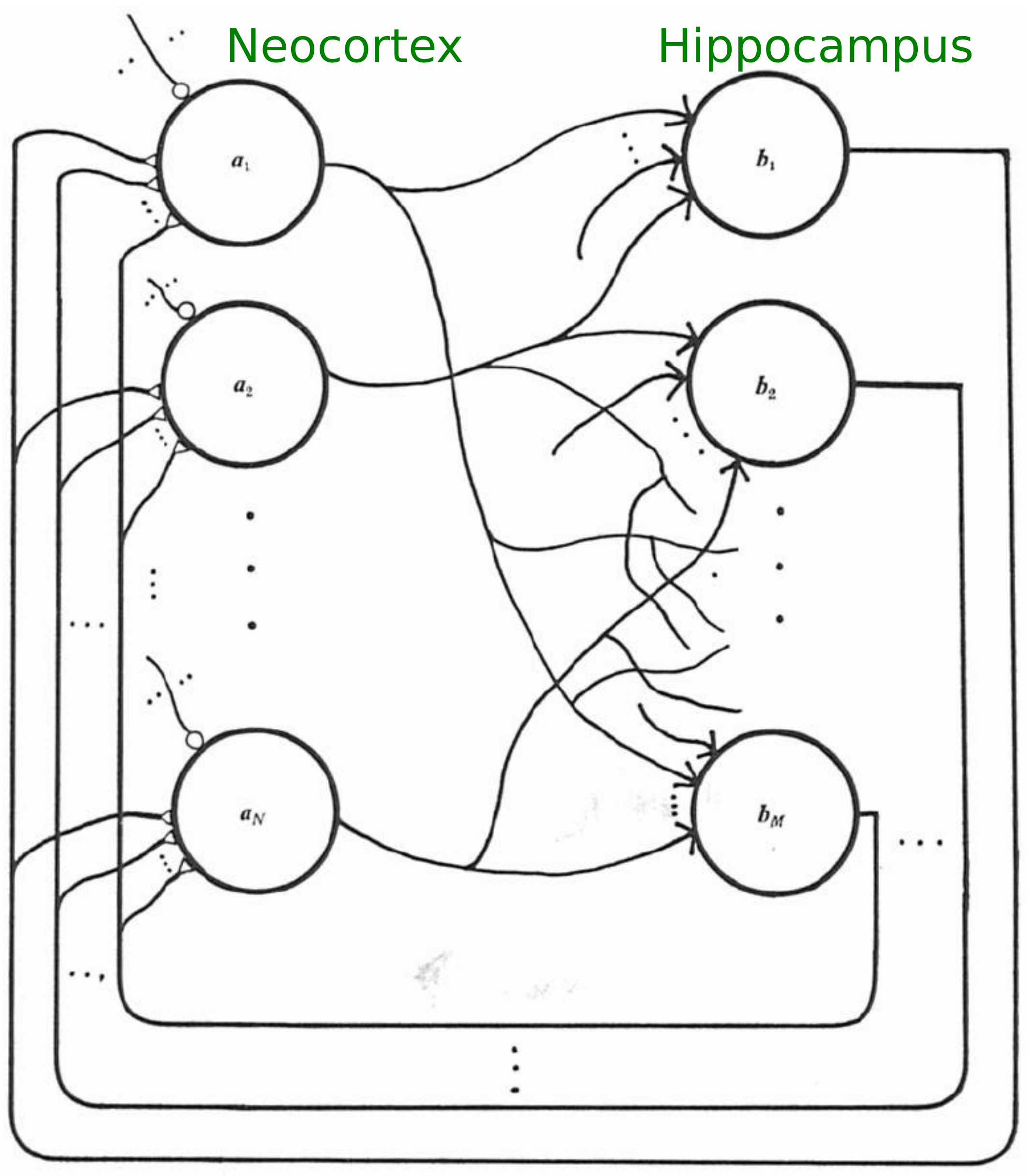
- Activity level (percentage of active units) should be low so that patterns will be sparse and many events can be stored.
- Inhibition is used to keep the number of active cells constant.
- Activity level must not be too low, because inhibition depends on an accurate sampling of the activity level.
- Assume at least 1 cell in 1000 is active.
- That is, $\alpha > 0.001$.

N8. Size of Retrieval Cue

- Fraction of a previously stored event required to successfully retrieve the full event.
- Marr sets this to 1/10.
- This constitutes the minimum acceptable cue size.
- If the minimum cue size is increased, more memories could be stored with the same level of accuracy.

Marr's Two-Layer Model

- Event E is on cells $a_1 \ldots a_N$ (the cortical cells)
- Codon formation on $b_1 \ldots b_M$ (evidence cells in HC)
- Inputs to the b_j use Brindley synapses
- Codon formation is a type of competitive learning (anticipates Grossberg, Kohonen)
- Recurrent connections to the a_i use Hebb synapses



Simple Representations

- Only a small number of afferent synapses are available at neocortical pyramids for the simple memory function; the rest are needed for cortical computation.
- In order to recall an event E from a subevent X:
  - Most of the work will have to be done within the simple memory itself.
  - Little work can be done by the feedback connections to cortex.
- No fancy transformation from **b** back to **a**.
- Thus, for subevent X to recall an event E, they should both activate the same set of **b** cells.

Recalling An Event

- How to tell if a partial input pattern is a cue for recalling a learned event, or a new event to be stored?
- Assume that events E to be stored are always much larger (more active units) than cues X used for recall.
- Smaller pattern means not enough dendritic activation to trigger synaptic modification, so only recall takes place.

Codon Formation

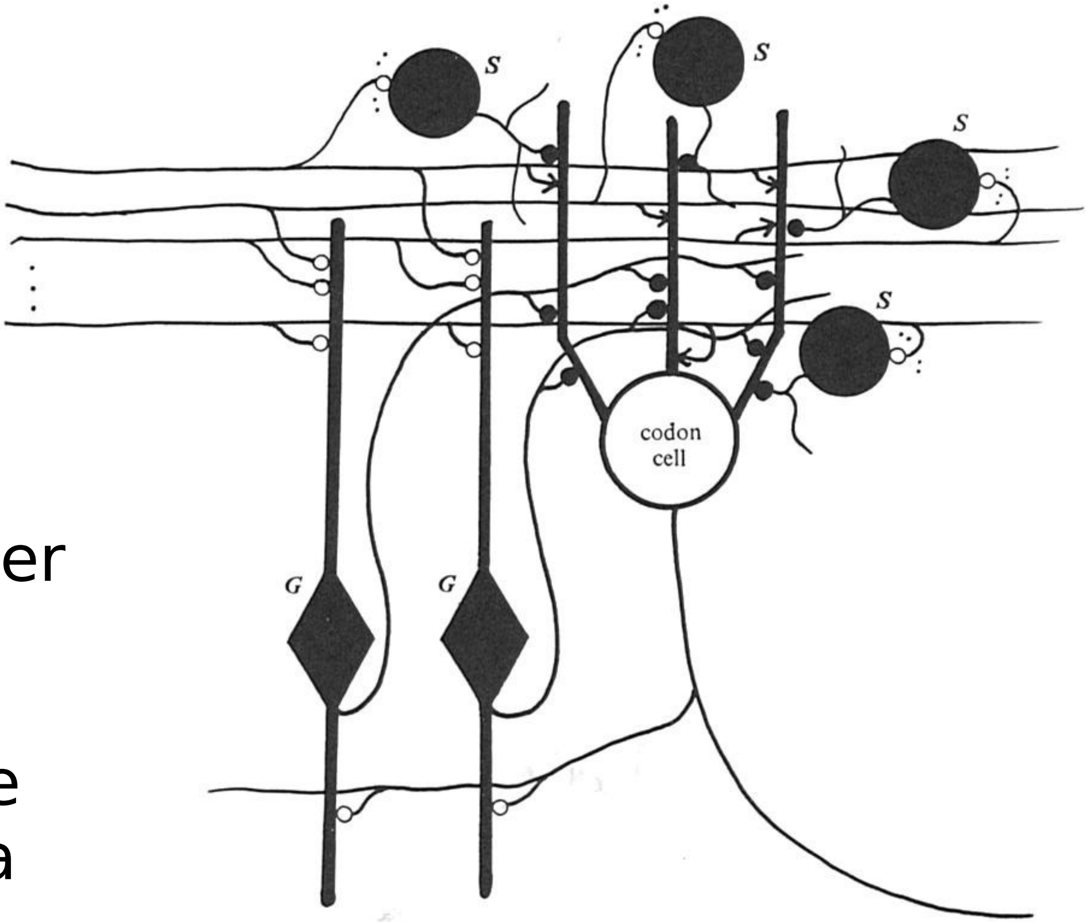
- Memory performance can be improved by orthogonalizing the set of key vectors.
  - The **b** cells do this. How?
- Project the vector space into a <u>higher dimensional</u> space.
- Each output dimension is a conjunction of a random *k*-tuple of input dimensions (so non-linear).
- In cerebellum this was assumed to use fixed wiring. In cortex it's done by a learning algorithm.
- Observation from McNaughton concerning rats:
  - Entorhinal cortex contains about 10^5 projection cells.
  - Dentate gyrus contains 10^6 granule cells.
  - Hence, EC projects to a higher dimensional space in DG.

Codon Formation

- For each input event E, different **b** cells will receive different amounts of activation.
- Activation level depends on which **a** cells connect to that **b** cell.
- We want the pattern size L to be roughly the same for all events.
- Solution: choose only the L most highly activated **b** cells as the simple representations for E.
- How to do this?
  - Adjust the thresholds of the **b** cells so that only L remain active.

Inhibition to Control Pattern Size

- S and G cells are inhibitory interneurons.
- S cells sample the input lines and supply feed-forward inhibition to the codon cells.
- G cells' modifiable synapses track the number of patterns learned so far, and raise the inhibition accordingly. They sample the codon cell's output via an axon collateral.



Threshold Setting

- Two factors cause the activation levels of **b** cells to vary:
  1) Amount of activity in the **a** cells (not all patterns are of the same size, due to partial cues)
  2) Number of potentiated synapses from **a** cells onto the **b** cell. This value gradually increases as more patterns are stored.
     - More cells can become active as more weights are set.

- Solution:
  1) S-cells driven by codon cell afferents compute an inhibition term based on the total activity in the a_i fibers.

     Assumes no synapses have been modified.
  2) G-cells driven by codon cell axon collaterals use negative feedback to compensate for effects of weight increases.

- Together, S and G cells provide <u>subtractive inhibition</u> to maintain a pattern size of L over the **b** units.

Recall From a Subevent

- If subevent X is fully contained in E, the best retrieval strategy is to lower the codon threshold until roughly L of the **b** cells are active.
- But if X only partially overlaps with E, some spurious input units will have synapses onto codon units. A better strategy is for codon cells to take into account the fraction *f* of their A active synapses that have been <u>modified</u> by learning (meaning they are part of some previously-stored pattern).
- Unmodified synapses that are active during recall can only be a source of noise.
- Thus, a **b** cell should only fire if a sufficient proportion *f* of its active synapses have been <u>modified</u>, meaning they are part of at least one stored pattern — perhaps the correct one, E.

Recall From a Subevent

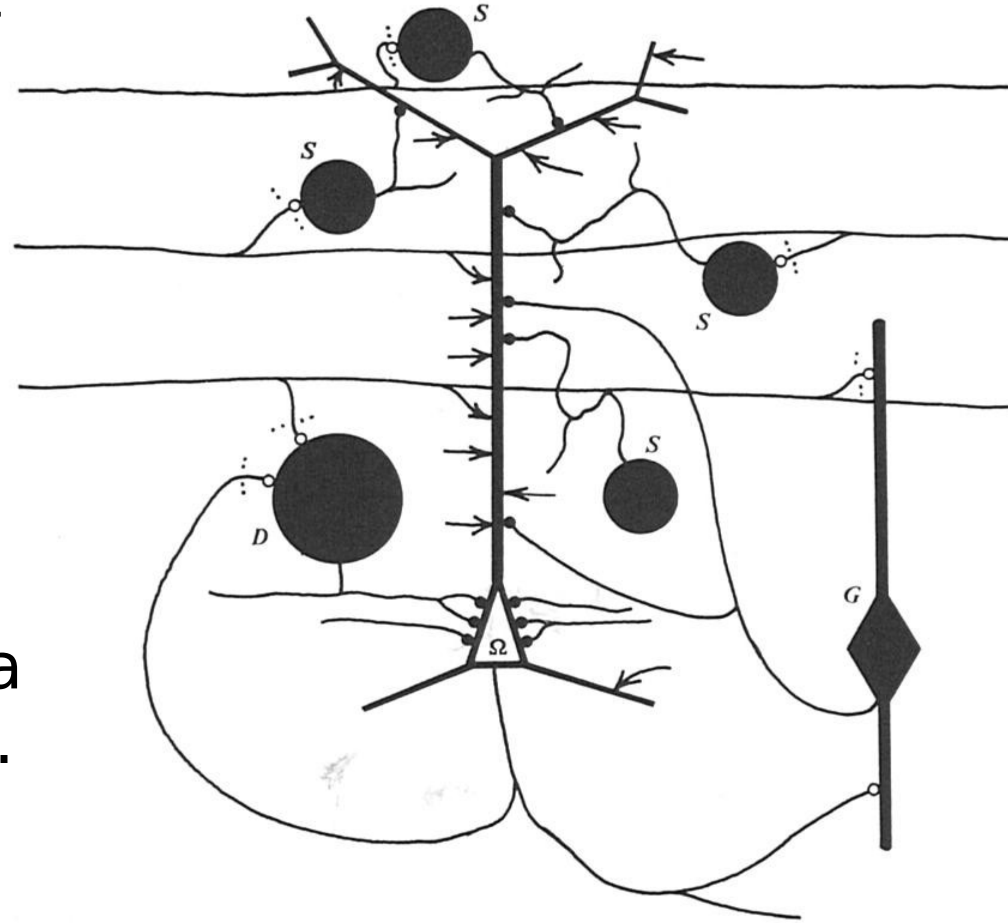
- A cell should only fire if it's being driven by enough modified synapses.
- *A* = number of active synapses.
- *f* = fraction of synapses that have been modified.
- The cell's division threshold is equal to *fA*.
- Let *S* be the summed activation of the cell:

$$S = \sum_i a_i w_i$$

- The cell should fire if $S > fA$, or $S / (fA) > 1$.

D-Cells

- D cells compute *fA* and pass it as an inhibitory input to the pyramidal cells.
- D cells apply their inhibition directly to the cell body, like basket cells in hippocampus.
- This type of inhibition causes a division instead of subtraction.
- McNaughton: division can be achieved by shunting inhibition, e.g., the chloride-dependent $GABA_A$ channel.



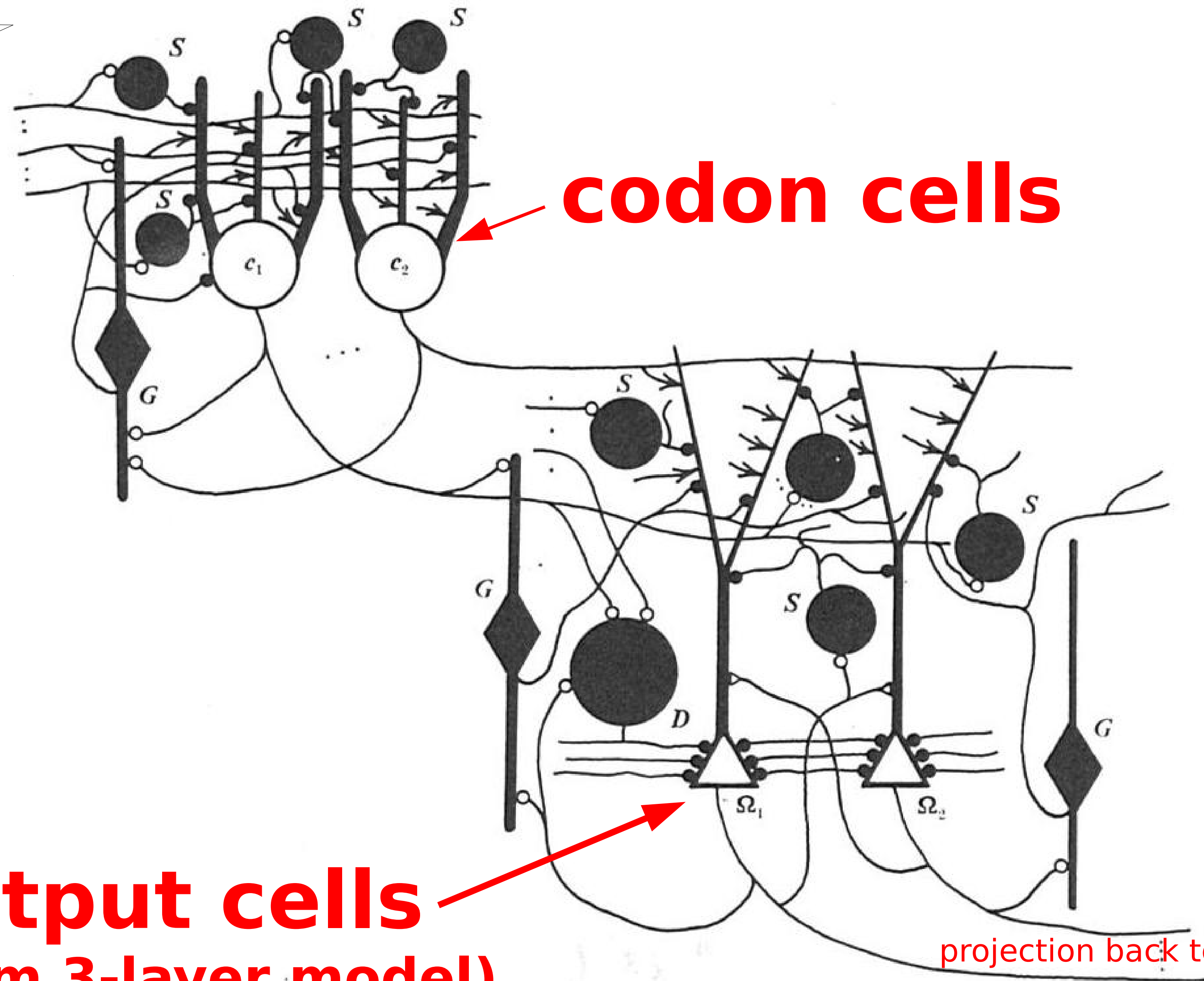
Dual Thresholds

- Cells have two separate thresholds:
  - The absolute threshold T, controlled by inhibition from S and G cells, should be close to the pattern size L, but must be reduced when given a partial cue.
  - The division threshold *fA*, controlled by inhibition from D cells.
- Marr's calculations show that both types of thresholding are necessary for best performance of the memory.
- How to set these thresholds? No procedure is given.
  - Willshaw & Buckingham try several methods, e.g., *staircase strategy*: start with small *f* and large *T*. Gradually reduce T until enough cells are active, then raise *f* slightly and repeat.

3 Layer Model: A Simple Memory With Output Cells

A cells

codon cells



Output cells
(from 3-layer model)

Inadequacy of the Simple Model

- Assume that $N = 10^6$ a_i afferents.
- Assume each neocortical pyramid can accept 10^4 synapses from the b_j cells.
- Assume upper bound of 200 learned events per cell, due to limitation on number of afferent synapses. (Marr derived this from looking at Purkinje cells in cerebellum.)
  - Use 100 events/cell as a conservative value.
- If capacity $n = 10^5$ events, and each **b** cell participates in 100 of them, then activity $\alpha = 10^{-3}$. With 10^4 **b** cells, only 10 can be active per event.
  - Too few for reliable representation. Threshold setting would be too difficult with such a small sample size.

What's Wrong With This Argument?

- The simple model is inadequate because the activity level is too low: only 10 active units per stored event.
- But this is because Marr assumes only 10^4 evidence (codon) cells. Why?
  - Limited room for afferent synapses back to the cortical cells.
- This is based on the notion that every evidence (codon) cell must connect back to *every* cortical cell.
- Later in the paper he relaxes this restriction and switches to 10^5 evidence cells.

Combinatorics 1: Permutations

- How many ways to order 3 items: A, B, C?

- Three choices for the first slot.
- Two choices left for the second.
- One choice left for the third.

B A C

- Total choices = 3 x 2 x 1 = 3! = 6.

Combinatorics 2: Choices

- How many ways to choose 2 items from a set of 5?

In formal notation, what is the value of $\binom{5}{2} = C(5,2)$?

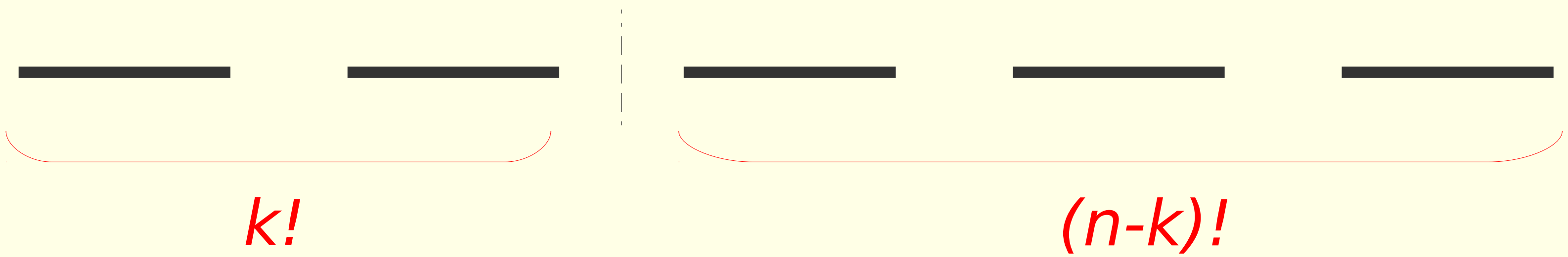
- Five choices for first item. Four choices for the second.
- Permutations of the chosen item are equivalent: combination B,E is the same as combination E,B

- So total ways to choose two items is (5 x 4)/(2!) = 10.
- Since 5! = 5 x 4 x 3 x 2 x 1, we can get 5x4 from 5!/3!

$$\binom{5}{2} = \frac{5!}{3!}/2! = \frac{5!}{3!\cdot 2!}$$

Choices (continued)

- How many ways to choose *k*=2 items from *n*=5 ?
- Allocate 5 slots giving *n!* = 120 permutations:



- All permutations of the *k* chosen items are equivalent, so divide by *k!* = 2.
- All permutations of the *(n-k)* unchosen items are equivalent, so divide by *(n-k)!* = 6.

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Review of Probability

- Suppose a coin has a probability *z* of coming up heads.
- The probability of tails is (1-*z*).
- What are the chances of seeing *h* heads in a row?

$$z^h$$

- What are the chances of seeing exactly *h* heads in a row, followed by exactly *t* tails?

$$z^h \cdot (1-z)^t$$

- What about seeing exactly *h* heads total in *N* tosses?

$$\binom{N}{h} \cdot z^h \cdot (1-z)^{(N-h)}$$

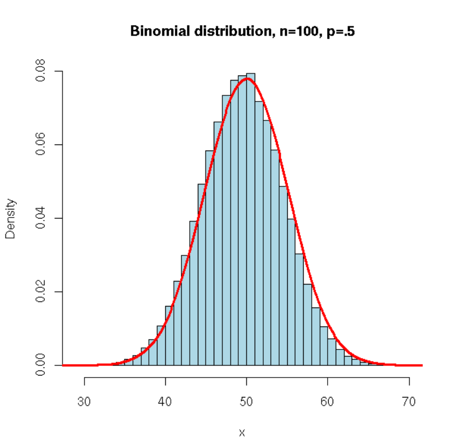
Binomial Distribution

- How many heads should we expect in N=100 tosses of a biased (z=0.2) coin?
  - Expected value is E<*h*> = *Nz* = 20.
- What is the probability of a particular sequence of tosses containing exactly *h* heads?

$$P\left[\langle t_{1,} t_{2,} \ldots, t_N \rangle\right] = z^h \cdot (1-z)^{N-h}$$

- The probability of getting exactly *h* heads in any order follows a binomial distribution:

$$\text{Binomial}(N;z)[h] = \binom{N}{h} \cdot z^h \cdot (1-z)^{N-h}$$



Marr's Notation

P_i	Population of cells.
N_i	Number of cells in population P_i
L_i	Number of active cells for a pattern in P_i
α_i	Fraction of active cells: L_i / N_i
R_i	Threshold of cells in P_i
S_i	Number of afferent synapses of a cell in P_i
Z_i	Contact probability: likelihood of synapse from cell in P_{i-1} to P_i
Π_i	Probability that a particular synapse in P_i has been modified
$E\langle x \rangle$	Expected (mean) value of x
n	Number of stored memories

Response to an Input Event

- Assume afferents to P_i distribute uniformly with probability Z_i.
- L_{i-1} = number of active afferents.
- What is the expected pattern size in this population?

$$E\langle L_i \rangle = N_i \sum_{r=R_i}^{L_{i-1}} \binom{L_{i-1}}{r} \cdot (Z_i)^r \cdot (1-Z_i)^{L_{i-1}-r}$$

- What do the terms in this formula mean?

Response to an Input Event

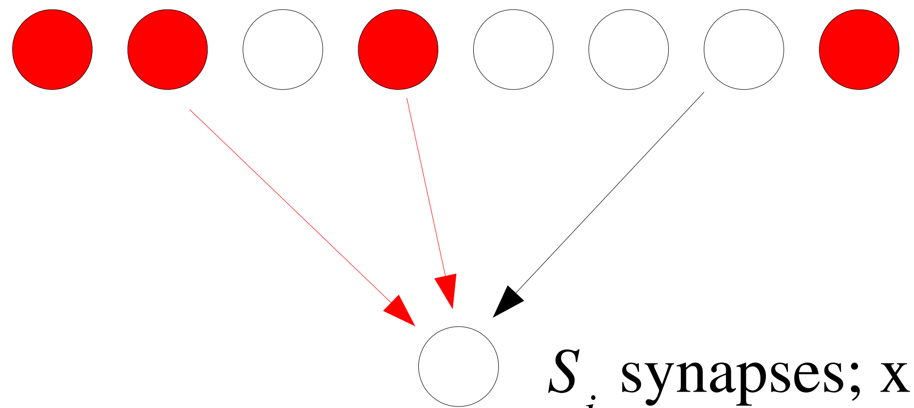
$$E\langle L_i \rangle = N_i \sum_{r=R_i}^{L_{i-1}} \binom{L_{i-1}}{r} \cdot (Z_i)^r \cdot (1-Z_i)^{L_{i-1}-r}$$

probability a unit has EXACTLY r active input fibers

probability a unit has AT LEAST R_i active input fibers (so is active)

- One term of the summation is the probability that a cell will receive an input of size exactly r, given L_{i-1} active fibers in the preceding layer.
- r is number of active fibers; R_i is the threshold.
- Must have $r \geq R_i$ in order for the layer i cell to fire. Also, $r \leq L_{i-1}$, the pattern size for layer i-1.
- Large R_i keeps us on the tail of the binomial distribution.
- The value of $\alpha_i = L_i / N_i$ will be small.

Counting Active Synapses



N_{i-1} cells; L_{i-1} are active

$$\alpha_{i-1} = L_{i-1} / N_{i-1}$$

S_i synapses; x are active

Number of active synapses x is binomially distributed.

$$P(x) = \binom{S_i}{x} \cdot (\alpha_{i-1})^x \cdot (1 - \alpha_{i-1})^{S_i - x}$$

$$E\langle x \rangle = \alpha_{i-1} S_i$$

Constraint on Modifiable Synapses

Activity $\alpha_{i-1} = L_{i-1}/N_{i-1}$.

Proportion of synapses active at each active cell of P_i is at least equal to the mean α_{i-1} because the active cells are on the tail of the distribution.

The amount by which it exceeds this decreases as $S_i \alpha_{i-1}$ grows.

Probability that a (pre,post)-synaptic pair of cells is simultaneously active is $\alpha_{i-1}\alpha_i$.

After n events, probability that a particular synapse of P_i is facilitated is:

$$\Pi_i = 1-(1-\alpha_{i-1}\alpha_i)^n$$

If α_{i-1} is small, then $\alpha_{i-1}\alpha_i$ is smaller, so this gives roughly

$$\Pi_i \approx 1-\exp(-n\alpha_{i-1}\alpha_i)$$

because for small ϵ, $(1-\epsilon)^n \approx \exp(-n\epsilon)$

Constraint on Modifiable Synapses

- For modifiable synapses to be useful, not all should be modified after *n* events are stored.
  - Otherwise we could just make all of them fixed.

- Suppose we want at most 1 – (1/e) of them to be modified, which is about 63%.

$$\Pi_i \leq 1 - (1/e)$$
$$= 1 - \exp(-1)$$
$$\approx 1 - \exp(-n \alpha_{i-1} \alpha_i)$$

- Thus we have computational constraint C1:

$$n \alpha_{i-1} \alpha_i \leq 1$$

Condition for Full Representation

- Activity in P_i must provide an adequate representation of the input event.
- Weak criterion of adequacy: change in input fibers (active cells in P_{i-1}) should produce a change in the cells that are firing in P_i.
- Cells in P_i just above threshold $\rightarrow$ losing one input will shut off the cell.

Condition for Full Representation

Probability P that an arbitrary input fiber doesn't contact any active cell of P_i (so P_i doesn't care if it's shut off) is:

$$P = (1 - Z_i)^{L_i}$$

$$L_i = \alpha_i N_i$$

$$Z_i = S_i / N_{i-1}$$

$$(1-\epsilon)^n \approx \exp(-n\epsilon)$$

$$P \approx \exp\left(-\alpha_i N_i \cdot S_i / N_{i-1}\right)$$

Let's require $P < e^{-20}$ (about 2×10^{-9}). Then with a little bit of algebra we have computational constraint C2:

$$S_i \alpha_i N_i \geq 20 N_{i-1}$$

Summary of Constraints

- To store lots of memories, patterns must be sparse.

$$\text{Constraint C1:} \quad n \alpha_i \alpha_{i-1} < 1$$

- For the encoding to always distinguish between input patterns, outputs must change in response to any input change.
  - There must be enough units and synapses to assure this.

$$\text{Constraint C2:} \quad S_i \alpha_i N_i \geq 20 N_{i-1}$$

- Assumes output cells are just above threshold so losing 1 input fiber will turn them off. They must be on the tail of the binomial distribution for this to hold.

What's Next?

- Move to a larger, three-layer, block-structured model.
- Add recurrent connections.
- Derive conditions under which recurrent connections improve recall results.
- Map this model onto the circuitry of the hippocampus.