Marr's Theory of Hippocampus

David S. Touretzky

David Marr

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*

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 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?

Marr claims that neocortex can classify inputs (category formation), whereas archicortex, being simpler, can only do associative recall.

Was this conclusion justified by the anatomy?

What Does the Hippocampus Do?

Stores patterns efficiently, without further analysis.

Later the neocortex can pick out the important features of the patterns, and remember 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

The Pattern Space

An Animal's Limited History

If 10 fibers out of 1000 can be active at once, that gives $\binom{1000}{10}$ possible combinations.

Assume a new pattern every 1 msec.

Enough combinations to go for 10^{12} years.

So: assume patterns will not repeat in the lifetime of the animal.

Very few possible events will actually be encountered.

So events will be well-separated in the pattern space, not close together.

Numerical Constraints

N1. Capacity Requirements

A simple memory only needs to store one day's worth of experiences.

They will be transferred to the neocortex at night, during sleep.

There are 86,400 seconds in a day

A reasonable upper bound on memories stored is 10,000 events per day.

Numerical Constraints (cont.)

N2: Number of Inputs to the Memory

Too many cortical pyramids (10^8): can't all have direct contact with hippocampus.

Solution: introduce *indicator cells* as markers of activity in each local cortical region, about 0.03 mm^2.

Indicator cells funnel activity into the hippocampal system.

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, and each axon can make many synapses.

Numerical Constraints (cont.)

Roughly 30 indicator cells per mm^2.

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 entering simple memory.

Numerical Constraints (cont.)

N3. Number of Outputs from Memory.

Assume neocortical pyramidal cells have fewer than 10^5 afferent synapses.

Assume only about10^4 synaptic sites available 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.

Later this constraint was changed to permit 10^5 memory cells.

Numerical Constraints (cont.)

N4. Synaptic Modification.

Hebb synapses are either *on* or *off*.

Brindley synapses have a fixed component in addition to the modifiable component.

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.

Numerical Constraints (cont.)

N5. Number of Synapses Made on a Cell

Assumed to be high but bounded: no more than 60,000.

In most calculations he uses a value of 10^5.

Numerical Constraints (cont.)

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.

Numerical Constraints (cont.)

N7. Level of Activity

Activity level should be low so that patterns will be sparse and lots of 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.

Assumes that at least 1 cell in a thousand is active.

That is, $\alpha > 0.001$.

Numerical Constraints (cont.)

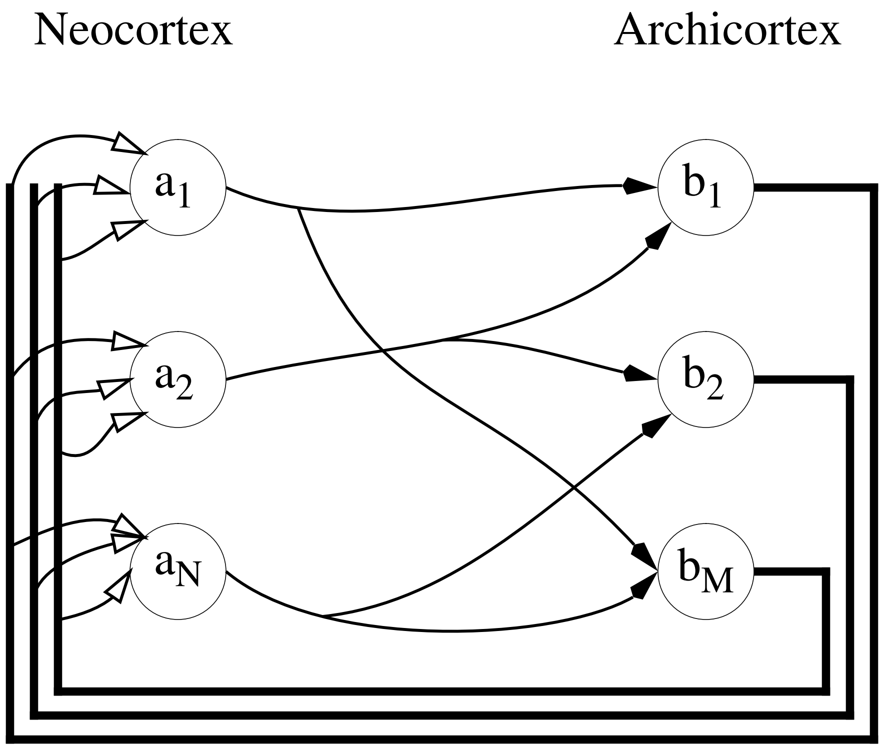
N8. Size of Cue.

Fraction of a previously stored event required to successfully retrieve the full event is set at 1/10.

This constitutes the minimum acceptable cue size.

If this minimum is increased, more memories could be stored.

A Two-Layer Model



Event E is on cells a_1 ... a_N (the cortical cells.)

Codon formation on cells b_1 ... 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.

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 transformations between **b** and **a**.

Thus, for subevent X to recall an event E they should both activate the same set of **b** cells.

The **b** cell representation is called the *simple representation* of E.

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 to be stored are always much larger (more active units) than cues 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 higher-dimensional space.

Each output dimension is a conjunction of a random *k*-tuple of input dimensions.

In cerebellum this was assumed to be hardwired. 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 thesame for all events.

Solution: choose only the L most highly activated **b** cells as the simple representation for E.

How to do this?

Adjust the thresholds of the **b** cells so that only L remain active.

Threshold Setting

Two factors cause the activation levels of **b** cells to vary:

1. Amount of activity in the **a** cells (not all patterns of 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.

Note: don't need G-cells if old weights decay at the same rate that weights for new memories are potentiated.

Recall From a Subevent

If subevent X is fully contained in E, then best strategy is to lower the codon threshold until roughly L cells are active.

But if X only partially overlaps with E, a better strategy is to take into account the fraction of the A active synapses that have been modified by learning (meaning they are part of some stored pattern)

Unmodified synapses that are active during recall are a source of noise.

A cell should only fire if a sufficient proportion f of its active synapses have been modified.

The cell's division threshold is equal to fA.

Let S be the summed activation. The cell should fire if $S > fA$,
or $S / (fA) > 1$

Division Cells

D-cells compute *f*A 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 rather than a subtraction.

McNaughton notes that division can be achieved by shunting inhibition, e.g., the chloride dependent $GABA_A$ channel.

Dual Thresholds

Cells have two separate thresholds.

There is 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 only a partial cue.

There is also the ***division threshold,*** fA, controlled by inhibition from D cells.

Marr's calculations show that both types of thresholding are necessary for the best performance of the memory.

How to set these thresholds? No explicit procedure 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.

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.

Thus $M = 10^4$ b_j cells in the simple memory, each fully connected back to cortex.

Assume upper bound of 200 learned events/cell, due to limitation on number of afferent synapses. (Derived from Purkinje cells.)

Use 100 events/cell as a conservative value.

If $n = 10^5$ events, and each **b** cell participates in 100 of them, then activity $\alpha = 10^{-3}$. With 10^4 **b** cells, only 10 active per event.

Too few for a reliable representation.

Why? Perhaps because threshold setting would be too difficult.

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 Marr assumes only 10^4 evidence cells. Why? (Limited room for afferent synapses on the cortical cells.)

This is based on the notion that every evidence cell must connect back to every cortical cell.

Later in the paper he switches to 10^5 evidence cells.

Combinatorics 1: Permutations

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

Allocate three slots:

____ ____ ____

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

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 $\binom{5}{2}$?

Five choices for the first item.

Four choices for the second item.

Permutations of the two chosen items are equivalent:
combination (B,E) = combination (E,B).

So total ways to choose two items = (5 x 4)/(2!) = 10.

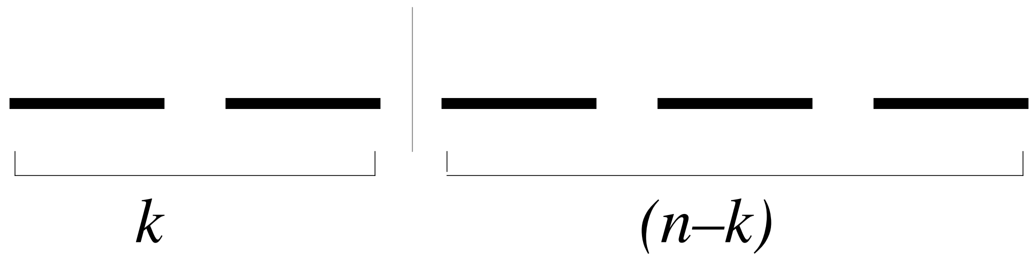
Since 5! = 5 x 4 x 3 x 2 x 1, we can get 5 x 4 from 5!/3!.

$$\binom{5}{2} = \left(\frac{5!}{3!}\right) / (2!) = \frac{5!}{3! \cdot 2!}$$

Choices (continued)

How many ways to choose k=2 items from n=5?

Allocate 5 slots giving $n! = 120$ permutations:



All permutations of the 2 chosen items are equivalent, so divide by $k! = 2$.

All permutations of the 3 unchosen items are equivalent, so divide by $(n-k)! = 6$.

Result is (120/2) / 6 = 10.

General formula for choosing k items from a set of n:

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

Review of Probability

Suppose a coin has probability r of coming up heads.

The probability of tails is $(1-r)$.

What are the chances of seeing p heads in a row?

$$r^p$$

What are the chances of seeing exactly p heads followed by exactly q tails?

$$r^p \cdot (1-r)^q$$

What are the chances of seeing exactly p heads out of n tosses?

$$\binom{n}{p} \cdot r^p \cdot (1-r)^{(n-p)}$$

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 that a given cell in P_i receives a synapse from a given cell in P_{i-1}.
Π_i	Probability that a particular modifiable 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.

What is the expected pattern size in this population?

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

Note: r is the number of active input fibers from layer i–1 that contact a particular unit in layer i.

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.

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. Value of $\alpha_i = L_i / N_i$ will be small.

Constraint on Modifiable Synapses

Activity $\alpha_i = L_i / N_i$.

Proportion of synapses active at each cell of P_i is at least α_{i-1}

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 simutaneously active is $\alpha_{i-1}\alpha_i$.

After *n* events, the probability that a particular synapse of P_i has been facilitated is:

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

If α_{i-1} is small, this gives, roughly,

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

Constraint on Modifiable Synapses (cont.)

If all synapses have been modified after n events are stored, we could just make all the synapses fixed.

For modifiable synapses to be useful, not all should be modified after n events.

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

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.

Probability P that an arbitrary input fiber *doesn't* contact any active cell of P_i is:

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

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

If we want $P < e^{-20}$, then computational constraint C2:

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

The Collateral Effect

When are recurrent collaterals useful?

Let P_i be a population of cells forming a simple representation.

Each cell can learn about 100 input events.

Population as a whole learns $n = 10^5$ events.

Hence α must be around 10^{-3}.

We require $n\ \alpha_i\ \alpha_{i-1}$ to be at most 1. Estimated value is 0.1.

Hence we can let $P_{i-1} = P_i$ and use recurrent collaterals from P_i to help clean up the representation.

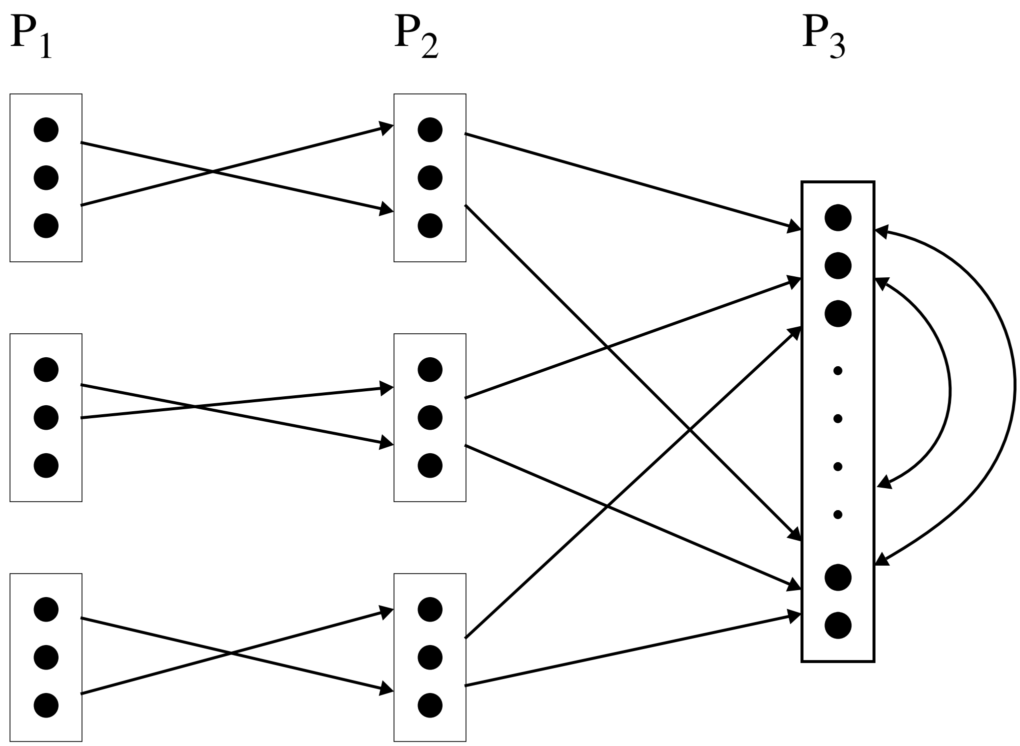
Result: an input to P_i need not be sufficient to reproduce the entire simple representation.

It can reproduce a portion, and the recurrent collaterals can then augment this to recover the entire simple representation.

The Three Layer Model

Projections: $P_1 \Rightarrow P_2 \Rightarrow P_3$.

Layers P_1 and P_2 each divided into 25 blocks.



Anatomical correlates:

- P_1 = neocortex (?)
- P_2 = entorhinal cortex and presubiculum
- P_3 = hippocampus (CA3)

Parameters for The Three Layer Model

P_1 has 1.25 x 10^6 cells, divided into 25 blocks of 50,000.

P_2 has 500,000 cells, divided into blocks of 20,000.

P_3 has a single block of 100,000 cells.

Let S_3 = 50,000.

Let x_i be number of synapses on a cell used in storing one event.

$$\Pi_i = 1 - (1 - x_i / S_i)^{n\alpha_i}$$

$$x_i = \sum_{R \geq R_i} P_i(R) \cdot R$$

$P_i(R)$ is the probablility that an active cell in layer i has exactly R active afferent synapses.

From the above, we calculate L_3 = 217, and α_3 = 0.002.

If we want useful collateral synapses in P_3, must have $n(\alpha_3)^2 < 1$.

So,with $n = 10^5$ events, we have α_3 = at most 0.003.

When Do Collateral Connections Help?

Let P_{30} be the simple representation of E_0 in P_3.

Let P_{31} be the remaining cells in P_3.

Let C_0 = the active cells in P_{30}.

Let C_1 = the active cells in P_{31} (noise.)

Note that $C_0+C_1 = L_3$.

Let C_0', C_1' be the new values after the recurrent connections have had their effect on P_3.

The ***statistical threshold*** is the ratio $C_0:C_1$ such that the effect of the collaterals is zero: $C_0:C_1 = C_0':C_1'$

Collaterals help whenever the statistical threshold is exceeded.

Calculating C_0', C_1' is a bit tricky because there is both a subtractive and a division threshold. See Marr §3.1.2 for details.

Effect of Collaterals

Marr estimates that the collaterals should have made their full contribution to recovering the event in about 3 cycles.

Additional cycles would provide no benefit.

McNaughton:

- Oscillating cycle of excitation and inhibition in hippocampus, known as the theta rhythm.
- Hippocampal cell output is phase-locked to the theta rhythm, with a period of about 140 msec.
- Assume pattern completion takes place in the 1/4 cycle where the excitation is increasing: 35 msec window.
- Conduction delay and synaptic delay totals 6–8 msec.
- This leaves room for just 4–6 cycles; very close to Marr's prediction.

Marr's Performance Estimate

Input patterns: L_1 = 2500 units (25 blocks; 100 active units each)

Output patterns: L_3 = 217 units.

With $n = 10^5$ stored events, accurate retrieval can be had with:

- 30 active fibers in one block, all of which are in E_0
- 100 active fibers in one block, of which 70 are in E_0 and 30 are noise

With $n = 10^6$ stored events, accurate retrieval requires just:

- 60 active fibers in one block, all in E_0
- 100 active fibers in one block, of which 90 are in E_0

Willshaw and Buckingham's Model

Willshaw and Buckingham implemented a simplified, one-hundredth scale model of Marr's architecture

Didn't bother partitioning P_1 and P_2.

P_1 = 8000 cells, P_2 = 4000 cells, P_3 = 1000 cells.

For two-layer version, omit P_2.

Performance was similar for both architectures.

Memory capacity was roughly 1000 events.

- Partial cue of 8% of an event gave perfect retrieval 66% of the time.
- In two-layer net, 16% cue gave perfect retrieval 99% of the time.
- In three-layer version, 25% cue gave 100% perfect retrieval.

Assessment of Marr's Theory

Strong points:

- Sparse connectivity: more biologically realistic.
- Multiple inhibitory mechanisms: subtraction, division.
- Analyzes recurrent collaterals: when do they help vs. hurt.
- Anticipated many important findings: LTP, division operations, information transfer during sleep.

Weak points:

- Not that close to the biology: ignores the trisynaptic circuit.
- Says nothing about CA1. Ignores the direct perforant path input to CA3 (and to CA1).
- Claim that three layers of cells are necessary is unjustified.
- Unanswered question: how are memories transferred to neocortex?