Microsoft Research

Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.

2013 © Microsoft Corporation. All rights reserved.
NIPS Thanks Its Sponsors
Hippocampal memory reactivation in awake and sleep states

Matthew Wilson
Departments of Brain and Cognitive Sciences and Biology
MIT

The Picower Institute for learning and memory
The luminal hypothesis revisited.

Fig. 2. The position of the hippocampal formation in the rat brain is shown in this drawing of a preparation in which the cortical surface overlying the hippocampus has been removed. The hippocampus is an elongated, C-shaped structure with the long or septotemporal axis running from the septal nucleus rostrally (S) to the temporal cortex (T) ventrocaudally. The short or transverse axis (TRANS) is oriented perpendicular to the septotemporal axis. The major fields of the hippocampal formation (except for the entorhinal cortex) are found in slices taken approximately midway along the septotemporal axis. The slice pictured at top left is a representation of the summary of the major neuronal elements and intrinsic connections of the hippocampal formation as originally illustrated by Andersen et al. (see text for details).

Abbreviations: DG, dentate gyrus; mt, mossy fibers; pp, perforant path; sc, Schaffer collaterals.

From Amaral and Witter, 1989
Hippocampus in spatial and episodic memory

- The hippocampus is involved in the formation of episodic memory as well as spatial memory used in navigation.
- Navigation - linkage of spatial locations
- Episodic memory - linkage of events
- Both may depend critically on temporal sequence encoding
Neural recording device

4-channel microwire electrode

Multiple electrode microdrive array
Spike amplitude clustering
Place Fields on Linear Tracks
Place Fields on Linear Tracks
Hippocampus online and offline

Theta rhythm

Sharp wave/ripples

walk  still

Buzsaki 1989
Hippocampal spatial representations are encoded as sequences during behavior.

Davidson, Kloosterman, and Wilson, *Neuron*, 2009
Role of Sleep in Memory

• Sleep allows examination of memory independent of behavior.

• The formation of lasting memories may involve the communication of information between brain areas during sleep.

• Broadly identify two stages of non-REM sleep – (NREM) and rapid eye movement sleep (REM).
Experimental design

- Sleep
- Run
- Sleep

- 1-2 hrs
- 15 min
- 1-2 hrs

Colors:
- Green: slow-wave sleep
- Red: REM sleep
- Black: awake behavior
Compressed Run sequences are expressed in hippocampus during nREM sleep.

4 sec

150 msec

Lee and Wilson, Neuron, 2002
Sequences are re-expressed during CA1 ripple events

Duration of low probability sequences

Correlation of low probability sequences and ripples

Example of a low probability sequence and a ripple event
Are there signatures of memory reactivation in the neocortex during hippocampal reactivation

• Simultaneously record in the hippocampus and primary and secondary visual cortex during spatial behavior.

• Look for reactivation in both structures during sleep.
Experimental Design:

| PRE (1-2hrs) | RUN (20-40mins) | POST (1-2hrs) |

B

1. Intra-maze local cues, no prominent distal cues
2. Well trained animals: alternation task
3. Recording sites: visual cortex (Occ1, Occ2) and CA1
4. Sleep states (SWS, REM, Wake, Int) classified using EMG and hippocampal EEG
Sequence memory reactivation in hippocampus and visual cortex

Ji and Wilson, Nature Neuroscience, 2007
Reactivation occurs during activity frames correlated with the slow oscillation.
Can we influence memory reactivation during sleep?

Sound L
- downward frequency sweep

Sound R
- upward frequency sweep

Sleep box (away from track)

For 2-2.5 hours
- Sound L
- Sound R
- control sounds
Behavioral task design

- Sound L: frequency over time, nasal poke.
- Sound R: frequency over time, nasal poke.
- Reward site: left to right.

Graphs showing estimated position, actual position, and unit number over time.

- 3.7 Hz
- 13.3 Hz
Do task-related sounds bias the content of future replay?

**Hypothesis:**

**Sound R** - place cells with **right-sided** place fields are more active during replay  
**Sound L** - place cells with **left-sided** place fields are more active during replay
Bias observed in individual place cell responses

Bendor and Wilson, *Nature Neuroscience*, 2012
Bias is maintained after initial cueing

A

Temporal dynamics of rate bias

- Place field
  - Blue: right-sided
  - Red: left-sided

Mean rate bias (spk/s)

- First half (0-5.4s)
- Second half (5.4-10.8s)

** P < 0.001

Bendor and Wilson, Nature Neuroscience, 2012
Rasch et al. 2007
Learning – 50 object locations

Subsequently cued  Subsequently uncued

meow  whistle

Rasch et al. 2007

Rudoy et al. 2009
Hippocampus online and offline

<table>
<thead>
<tr>
<th>Theta rhythm</th>
<th>Sharp wave/ripples</th>
</tr>
</thead>
<tbody>
<tr>
<td>walk</td>
<td>still</td>
</tr>
</tbody>
</table>

Buzsaki 1989
Hippocampal activity during quiet wakefulness

- During awake behavior, there are periods of quiet wakefulness that have EEG that is similar to NREM consisting of brief bursts of activity modulated by high frequency "ripple" oscillations.
- Is there structure to the patterns of multiple single neuron activity during this state?
What do animals think about when they stop and eat after running down a track?

a

b

STOPPED AFTER RUNNING
They think back to where they have just been.

Memory of recent experience replayed in reverse-time order

(a) Position vs. time

(b) Hippocampal place-cell activity vs. time

(c) Reverse-time sequence replay during hippocampal ripples

Foster and Wilson, *Nature*, 2006
Questions

✓ Replay in a larger environment?

✓ Replay associated with reward sites only?

✓ Replay always begins with cells that have place fields close to animal's current location?

✓ Replay in forward and reverse directions?
Long behavioral sequences on a 10m track

Davidson, Kloosterman and Wilson, Neuron, 2009
Reconstruction of extended sequence replay during quiet wakefulness

Davidson, Kloosterman and Wilson, Neuron, 2009
Forward Replay from A to B

- Multi-unit activity
- Ripples

Position estimate

Building blocks of memory?
Forward Replay from A to B

- Multi-unit activity
- Ripples

Position estimate

Linear position (m)

Time (s)
Extended replay spans multiple ripple events
Single ripple sequences are at same scale as theta sequences.

**A**
Estimated position (cm)

- 3.5 m/s
- 7.1 m/s

**B**
LFP (mV)

- 4710, 4710.5, 4711 time (s)

**C**
Behind/ahead animal (cm)

**D**
LFP (mV)

**E**
Unit firing rate (Hz)

Time relative to theta peak (s)

Davidson, Kloosterman, and Wilson, *Neuron*, 2009
Learning sequences of actions

Temporal credit assignment

Dopamine unit activity could differentially weight the content of hippocampal sequences, propagating value information from the rewarded location backwards along the incoming trajectory.

Foster and Wilson, Nature 440: 680-683, 2006
Hippocampal place-cell sequences depict future paths to remembered goals
Brad E. Pfeiffer & David J. Foster
Nature, 2013
Learning sequences of actions

Temporal credit assignment

Dopamine unit activity could differentially weight the content of hippocampal sequences, propagating value information from the rewarded location backwards along the incoming trajectory.

Foster and Wilson, Nature 440: 680-683, 2006
Hippocampal place-cell sequences depict future paths to remembered goals
Brad E. Pfeiffer & David J. Foster
Nature, 2013
Dopamine cell representations

- unexpected reward
- predictors of reward
- errors in the prediction of reward

Reward prediction error
Current reward – Expected reward

An error signal to teach target brain regions

Schultz, Dayan, Montague. Science 1997
VTA Hippocampus co-recordings

Bregma: -4.80 mm
Bregma: -5.30 mm
Bregma: -6.04 mm

Light microscopy
Anti-TH Ab
Spatial working memory task

Force Trials

Choice Trials
Task contingency associated VTA unit activity

DA

nosepoke

nonDA

C

E

Hz

seconds

Hz

seconds

Hz

seconds

Hz

seconds
VTA unit activity during RUN

Choice point

DA

Position (cm)

FR (normalized)

C-E maxFR (Hz)

p<0.05

nonDA

Trials

Position (cm)

FR (normalized)

C-E maxFR (Hz)

0 10 20 30 40 50

Position (cm)

Velocity (cm/sec)

arm1c/e arm2c/e arm3c/e
HC ripple bursts modulate DA unit activity

Single unit activity

Multiunit activity

Local field potential

Dopamine unit activity

DA spike frequency

Firing rate (Hz)

HC multiunit PSTH

Gomperts, unpublished data
Decoding hippocampal SPW-R associated multiunit bursts with spatial sequence reactivation
Replay and nonreplay SPW-R events

Reconstruction

Single unit activity

Multiunit activity

Local field potential

Position (cm)

Time (seconds)

mV
Dopamine unit modulation at hippocampal SPW-R bursts depends on replay content

replay

Non-replay

a1

b1

c1

d1

e

Replay PSTH modulation

Replay PSTH modulation
Dopamine units preferentially coordinate with replay of reward site locations on the spatial working memory task and phase lock to hippocampal theta.
Modulation of DA activity at SPW-Rs is reduced in slow wave sleep.
Summary

- DA unit activity increases during trajectories to rewards, differentially represents correct over error trials, and correlates with Q-TD prediction error in a spatial task.

- Hippocampal SPW-R bursts are associated with the modulation of DA units.

- Hippocampal theta phase-locking of DA unit activity correlates with the degree of SPW-R associated modulation.

- DA coordination with SPW-R bursts depends on replay content:
  - Replay of spatial sequences is associated with greater modulation.
  - DA units preferentially relate to replay of reward locations.
  - SPW-R modulation of DA units is reduced in slow wave sleep.
Overall summary

- Sequence memory can be encoded in the hippocampus during active behavior.
- Sequence memory is subsequently replayed during sleep in both the hippocampus and neocortex.
- The content of reactivated memory during sleep can be biased by external manipulation.
- Sequence memory replayed during quiet wakefulness is associated with reward information and may serve a different role in learning than replay during sleep.
Wilson Lab present and former

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert Lee</td>
<td>Janelia Farm</td>
<td>Non-REM replay</td>
</tr>
<tr>
<td>Daoyun Ji</td>
<td>Baylor</td>
<td>H-Visual cortex</td>
</tr>
<tr>
<td>David Foster</td>
<td>J. Hopkins</td>
<td>Awake replay</td>
</tr>
<tr>
<td>Fabian Kloosterman</td>
<td>Leuven</td>
<td>Extended awake replay</td>
</tr>
<tr>
<td>Tom Davidson</td>
<td>Stanford</td>
<td>Extended awake replay</td>
</tr>
<tr>
<td>Dan Bendor</td>
<td>UCL</td>
<td>Biased sleep replay</td>
</tr>
<tr>
<td>Steve Gomperts</td>
<td>Harvard</td>
<td>VTA and reward</td>
</tr>
</tbody>
</table>
Overall summary

- Sequence memory can be encoded in the hippocampus during active behavior.
- Sequence memory is subsequently replayed during sleep in both the hippocampus and neocortex.
- The content of reactivated memory during sleep can be biased by external manipulation.
- Sequence memory replayed during quiet wakefulness is associated reward information and may serve a different role in learning than replay during sleep.
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert Lee</td>
<td>Janelia Farm</td>
<td>Non-REM replay</td>
</tr>
<tr>
<td>Daoyun Ji</td>
<td>Baylor</td>
<td>H-Visual cortex</td>
</tr>
<tr>
<td>David Foster</td>
<td>J. Hopkins</td>
<td>Awake replay</td>
</tr>
<tr>
<td>Fabian Kloosterman</td>
<td>Leuven</td>
<td>Extended awake replay</td>
</tr>
<tr>
<td>Tom Davidson</td>
<td>Stanford</td>
<td>Extended awake replay</td>
</tr>
<tr>
<td>Dan Bendor</td>
<td>UCL</td>
<td>Biased sleep replay</td>
</tr>
<tr>
<td>Steve Gomperts</td>
<td>Harvard</td>
<td>VTA and reward</td>
</tr>
</tbody>
</table>
NIPS Thanks Its Sponsors

[Logos of sponsors such as Amazon, Microsoft, Google, Facebook, Skytree, Two Sigma, United Technologies Research Center, Yahoo! Labs, IBM Research, Xerox, D.E. Shaw & Co., Toyota, Millionshort, Criteo, PDT Partners, Springer, and Disney Research]
Correlations strike back (again): the case of associative memory retrieval

Cristina Savin
CBL, University of Cambridge, UK
IST Austria

with Peter Dayan and Máté Lengyel
neural activity

memory storage

neural dynamics

synaptic efficacies

W
Synaptic correlations in the cortex
Synaptic correlations in the cortex

\[ P(W_{ij}, W_{ik}) \]

Song et al, 2005
Synaptic correlations in the cortex

Where do they come from?
What do they mean for circuit function?

Song et al, 2005
1. Where do synaptic correlations come from?
1. Where do synaptic correlations come from?

\[ P(W) \]

\[ x^{(t)} \quad t = 1 \ldots T \]

\[ W \]
1. Where do synaptic correlations come from?

\[ P(W) \]

\[ x^{(t)} \]

\[ W \]

\[ x_j \]
\[ x_i \]
\[ x_k \]

\[ W_{ij} \]
\[ W_{ik} \]

\[ t = 1 \ldots T \]
\[ t = 1 \ldots T-1 \]
1. Where do synaptic correlations come from?
1. Where do synaptic correlations come from?

\[ P(W) \]

\[ x^{(t)} \]

\[ W \]

shared source of variability for synapses on the same cell

dependencies between synapses sharing a pre- or post- synaptic partner
1. Where do synaptic correlations come from?

Synaptic correlations are a natural consequence of synaptic plasticity.

Dependencies between synapses sharing a pre- or post-synaptic partner.
2. What do they mean for memory recall?
2. What do they mean for memory recall?

- The pattern we want to recall
- Other patterns
- Recall cue
2. What do they mean for memory recall?

memory recall: \[ P(x|W, \tilde{x}) \propto P(x)P(\tilde{x}|x)P(W|x) \]
\[ P(x|W, \tilde{x}) \propto P(x) P(\tilde{x}|x) P(W|x) \]
\[ P(x|W, \tilde{x}) \propto P(x)P(\tilde{x}|x)P(W|x) \]
\[ P(x|W, \tilde{x}) \propto P(x)P(\tilde{x}|x)P(W|x) \]

Intractable

Approximation: maximum entropy distribution

First order
\[ P(x|W, \tilde{x}) \propto P(x)P(\tilde{x}|x)P(W|x) \]

approximation: maximum entropy distribution

2nd order

1st order

synapse 2

synapse 1
Can we improve recall performance if we take into account synaptic correlations?

approximation: maximum entropy distribution
\[ P(x|W, \tilde{x}) \propto P(x)P(\tilde{x}|x)P(W|x) \]

Can we improve recall performance if we take into account synaptic correlations?

approximation: maximum entropy distribution
Can we improve recall performance if we take into account synaptic correlations?

approximation: maximum entropy distribution

\[ P(x|W, \tilde{x}) \propto P(x)P(\tilde{x}|x)P(W|x) \]
Can we improve recall performance if we take into account synaptic correlations?
plasticity
rule
plasticity rule

additive

$\Delta W_{ij} (\text{pre, post})$

bounded synapses
plasticity rule

covariance rule

additive

$\Delta W_{ij}(\text{pre, post})$

bounded synapses
<table>
<thead>
<tr>
<th>plasticity rule</th>
<th>2\textsuperscript{nd} order better than 1\textsuperscript{st}?</th>
<th>neural implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>covariance rule</td>
<td>NO</td>
<td>simple linear (Hopfield)</td>
</tr>
<tr>
<td>additive</td>
<td>error (%) graph</td>
<td></td>
</tr>
<tr>
<td>$\Delta W_{ij}$ (pre, post)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

bounded synapses
<table>
<thead>
<tr>
<th>Plasticity Rule</th>
<th>2nd Order Better than 1st?</th>
<th>Neural Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance Rule</td>
<td>NO</td>
<td>Simple linear (Hopfield)</td>
</tr>
</tbody>
</table>

Additive

\[ \Delta W_{ij} (\text{pre, post}) \]

Simple Hebb

Bounded Synapses
### Plasticity Rule

<table>
<thead>
<tr>
<th>Plasticity Rule</th>
<th>2\textsuperscript{nd} Order Better than 1\textsuperscript{st}?</th>
<th>Neural Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Additive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta W_{ij} ) (pre, post)</td>
<td>NO</td>
<td>simple linear (Hopfield)</td>
</tr>
<tr>
<td><strong>Simple Hebb</strong></td>
<td>YES</td>
<td>nonlinear inhibition</td>
</tr>
</tbody>
</table>

### Bounded Synapses
plasticity rule

**additive**

- Covariance rule
  - pre: 0, 1
  - post: 0, 1
  - $\Delta W_{ij}(\text{pre}, \text{post})$

- Simple Hebb
  - pre: 0, 1
  - post: 0, 1
  - $\Delta W_{ij}(\text{pre}, \text{post})$

- Postsyn. gated
  - pre: 0, 1
  - post: 0, 1
  - $\Delta W_{ij}(\text{pre}, \text{post})$

2\(^{nd}\) order better than 1\(^{st}\) ?

- **NO**
- **YES**

neural implementation

- Simple linear (Hopfield)
- Nonlinear inhibition
### Plasticity Rule

<table>
<thead>
<tr>
<th>Plasticity Rule</th>
<th>$2^{nd}$ Order Better than $1^{st}$?</th>
<th>Neural Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance Rule</td>
<td>NO</td>
<td>Simple linear (Hopfield)</td>
</tr>
<tr>
<td>Additive</td>
<td>YES</td>
<td>Nonlinear inhibition</td>
</tr>
<tr>
<td>$\Delta W_{ij}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postsyn. gated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Additive**

$\Delta W_{ij} (\text{pre, post})$

- Pre: 0, 1
- Post: 0, 1
- $\Delta W_{ij}$ values:
  - +1
  - -1

**Simple Hebb**

- Pre: 0, 1
- Post: 0, 1
- $\Delta W_{ij}$ values:
  - -
  - +1

**Postsyn. gated**

- Pre: 0, 1
- Post: 0, 1
- $\Delta W_{ij}$ values:
  - D
  - P

**Plots**

- $1^{st}$ order vs. $2^{nd}$ order error percentage
  - NO
  - YES
<table>
<thead>
<tr>
<th>Plasticity Rule</th>
<th>2nd Order Better than 1st?</th>
<th>Neural Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Additive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariance Rule</td>
<td><strong>NO</strong></td>
<td>Simple linear (Hopfield)</td>
</tr>
<tr>
<td>simple Hebb</td>
<td></td>
<td>Nonlinear inhibition</td>
</tr>
<tr>
<td>( \Delta W_{ij}(\text{pre, post}) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bounded Synapses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postsyn. gated</td>
<td><strong>YES</strong></td>
<td>Nonlinear dendritic integration</td>
</tr>
</tbody>
</table>

Diagram:
- Additive plasticity rule with two matrices showing the pre and post states.
- Simple Hebb rule with two matrices showing the pre and post states.
- Postsyn. gated rule with a visual representation of dendritic integration.
plasticity rule

<table>
<thead>
<tr>
<th>additive</th>
<th>2nd order better than 1st?</th>
<th>neural implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>covariance rule</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post 0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>pre</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>simple Hebb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post 0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>pre</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>error (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st order</td>
<td>2nd order</td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>simple linear (Hopfield)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonlinear inhibition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>postsyn. gated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post 0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>pre</td>
<td>D</td>
<td>P</td>
</tr>
<tr>
<td>current</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of inputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonlinear dendritic integration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasticity Rule</td>
<td>2nd Order Better than 1st?</td>
<td>Neural Implementation</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Covariance Rule</td>
<td>NO</td>
<td>Simple linear (Hopfield)</td>
</tr>
<tr>
<td>Additive</td>
<td></td>
<td>Nonlinear inhibition</td>
</tr>
<tr>
<td>$\Delta W_{ij}$ (pre, post)</td>
<td>YES</td>
<td>Nonlinear dendritic integration</td>
</tr>
</tbody>
</table>

**Cortex**

(Branco et al., 2011)
<table>
<thead>
<tr>
<th>Plasticity Rule</th>
<th>2nd Order Better Than 1st?</th>
<th>Neural Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance Rule</td>
<td>NO</td>
<td>Simple linear (Hopfield)</td>
</tr>
<tr>
<td>Additive</td>
<td></td>
<td>Nonlinear inhibition</td>
</tr>
<tr>
<td>$\Delta W_{ij}(\text{pre, post})$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple Hebb</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Postsyn. Gated</td>
<td></td>
<td>Nonlinear dendritic integration</td>
</tr>
<tr>
<td>Bounded Synapses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasticity Rule</td>
<td>2\textsuperscript{nd} order better than 1\textsuperscript{st}?</td>
<td>Neural Implementation</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>additive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariance rule</td>
<td>NO</td>
<td>simple linear (Hopfield)</td>
</tr>
<tr>
<td>Simple Hebb</td>
<td>YES</td>
<td>nonlinear inhibition</td>
</tr>
<tr>
<td>Bounded synapses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postsyn. gated</td>
<td>YES</td>
<td>nonlinear dendritic integration</td>
</tr>
<tr>
<td>Presyn. gated</td>
<td>YES</td>
<td>linear dendritic integration</td>
</tr>
</tbody>
</table>
plasticity rule

2nd order better than 1st?

neural implementation

- simple linear (Hopfield)
- nonlinear inhibition
- nonlinear dendritic integration
- linear dendritic integration
plasticity rule

2\textsuperscript{nd} order better than 1\textsuperscript{st}?  

neural implementation

- simple linear (Hopfield)
- nonlinear inhibition
- nonlinear dendritic integration
- linear dendritic integration

synaptic correlations need to be taken into account for recall
plasticity rule

2nd order better than 1st?

neural implementation

- synaptic correlations need to be taken into account for recall

- circuit nonlinearities can be understood as adaptations for approx. optimal recall
<table>
<thead>
<tr>
<th>Plasticity Rule</th>
<th>2nd Order Better than 1st?</th>
<th>Neural Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>post 0 1</td>
<td>corr</td>
<td>synaptic correlations need to be taken into account for recall</td>
</tr>
<tr>
<td>pre 0 1</td>
<td>error (%)</td>
<td>circuit nonlinearities can be understood as adaptations for approx. optimal recall</td>
</tr>
<tr>
<td>0 1</td>
<td>current</td>
<td>For technical details: poster Fri61</td>
</tr>
</tbody>
</table>
NIPS Thanks Its Sponsors
A memory frontier for complex synapses

Surya Ganguli

Dept. of Applied Physics
and, by courtesy,
Neurobiology and Electrical Engineering

Stanford University

Joint work with: Subhaneil Lahiri

a.k.a.

“Subhy”

Poster F50 Tonight
The synaptic basis for long-term memory storage

New York Times...
What is a synapse from neuron j to neuron i?

- **Theorist:** $W_{ij}$ or $J_{ij}$ ~ size of postsynaptic potential
A gulf between theory and experiment

What is a synapse from neuron j to neuron i?

- **Theorist:** \( W_{ij} \) or \( J_{ij} \) ~ size of postsynaptic potential

- **Experimentalist:** AMPA, NMDA, CAMKII, MAPK, CREB, MHC-I, second messengers, membrane protein regulation, intracellular trafficking, new protein synthesis

Coba et al. Science Signalling 2009
Memory capacity with scalar analog synapses

Consider the number of associations a neuron with \(N\) afferent synapses can store.

\[
\sigma(k) = \text{sgn} (J \cdot \xi(k) - \theta)
\]

An online learning rule to store the desired association:

\[
J(k+1) = e^{-1/\tau} J(k) + \sigma(k) \xi(k)
\]

i.e. 1) Allows analog weights to decay slightly (forget the past inputs)
2) Add in the new association to the weight (learn a new input).

**Memory capacity**: How far back into the past can synapses reliably recall previously stored associations?

**Answer**: If \(\tau\) is \(O(N)\) then the past \(O(N)\) associations can be recalled.

**Problem**: This solution relies on individual synapses to reliably maintain \(O(N)\) distinguishable analog states.
A gulf between theory and experiment

What is a synapse from neuron j to neuron i?

- **Theorist:** $W_{ij}$ or $J_{ij}$ ~ size of postsynaptic potential

- **Experimentalist:** AMPA, NMDA, CAMKII, MAPK, CREB, MHC-I, second messengers, membrane protein regulation, intracellular trafficking, new protein synthesis

Coba et al. Science Signalling 2009
Memory capacity with scalar analog synapses

Consider the number of associations a neuron with $N$ afferent synapses can store.

$$\sigma(k) = \text{sgn} \left( J \cdot \xi(k) - \theta \right)$$

An online learning rule to store the desired association:

$$J(k+1) = e^{-1/\tau} J(k) + \sigma(k) \xi(k)$$

i.e. 1) Allows analog weights to decay slightly (forget the past inputs)
2) Add in the new association to the weight (learn a new input).

**Memory capacity**: How far back into the past can synapses reliably recall previously stored associations?

**Answer**: If $\tau$ is $O(N)$ then the past $O(N)$ associations can be recalled.

**Problem**: This solution relies on individual synapses to reliably maintain $O(N)$ distinguishable analog states.
What about real synapses which can take only a finite number of distinguishable values for their strength?

For binary synapses each synapse $J_i = +1$ or $-1$. So you can no longer add an association to synaptic weights without running into boundaries.
Memory capacity with binary synapses

$J_i = -1, J_i = +1$

Potentiation

Depression

$q = \text{prob a synapse changes strength under appropriate conditions}$

$N = \text{number of synapses}$

Memory Capacity

$q = O(1)$  \quad log $N$

$q = O(N^{-1/2})$  \quad $N^{1/2}$

Quickly learn, quickly forget

Sluggish to learn, slow to forget

Fusi and Amit 92
Synaptic complexity: from scalars to dynamical systems

We must expand our theoretical conception of a synapse from that of a simple scalar value to an entire (stochastic) dynamical system in its own right.

This yields a large universe of synaptic models to explore and understand.
Framework for synaptic dynamical systems

Theoretical approach:

A synapse is an arbitrary stochastic dynamical system with M internal states.

Some internal states correspond to a strong synapse, others a weak synapse.

A candidate potentiation (depression) event induces an arbitrary stochastic transition between states.

Montgomery and Madison Neuron 2002
Ideal observer measure of memory capacity: SNR

A continuous stream of memories are stored (at poisson rate $r$) in a population of $N$ synapses with $M$ internal states.

The memory stored at time $t=0$ demands that some synapses potentiate, while others depress, yielding an ideal synaptic weight vector $w_{\text{ideal}}$.

The storage of future memories after $t=0$ changes the weight vector to $w(t)$.

An upper bound on the quality of memory retrieval of any memory readout using neural activity is given by the SNR curve:

$$\text{SNR}(t) = \frac{\langle w_{\text{ideal}} \cdot w(t) \rangle - \langle w_{\text{ideal}} \cdot w(\infty) \rangle}{\sqrt{\text{Var} (w_{\text{ideal}} \cdot w(\infty))}}$$

Each choice of $N, M, M^{\text{pot}}$ and $M^{\text{dep}}$ yields a different memory curve.

Fusi et. al. 2005, Fusi et. al. 2007, Barrett and van Rossum, 2008
Two example synaptic molecular networks

Serial Model
Leibold and Kempter 2008

Cascade Model
Fusi et. al. 2008

To elucidate the functional contribution of molecular complexity to memory, we want to not simply understand individual models, but understand the space of all possible models within this family.
Towards a general theory of synaptic complexity

How does the structure of a synaptic dynamical system ($M^{\text{pot}}$ and $M^{\text{dep}}$) determine its function, or memory curve SNR(t)?

What are the fundamental limits of achievable memory over all possible choices of synaptic dynamical systems?

What is the structural organization of synaptic dynamical systems that achieve these limits?

What theoretical principles can control combinatorial explosion in the number of possible models as $M$ increases?
Imposing a theoretical order on synaptic dynamics

As the synaptic population undergoes continuous modification, the internal state stochastically wanders around according to a forgetting process:

$$M_{\text{forget}} = f_{\text{pot}} \cdot M_{\text{pot}} + f_{\text{dep}} \cdot M_{\text{pot}}$$

This forgetting process has:

- An equilibrium probability distribution of state occupancy: $p_i^\infty$
- And a mean first passage time matrix from state $i$ to $j$: $T_{ij}$

Starting from state $i$, the average time it takes to get to the potentiated states, weighted by their equilibrium probability.

Order states from left to right in order of decreasing $\eta_i^{\text{pot}}$. 
Topological ordering from first passage times

large; takes a long time to reach potentiated states

small; takes a short time to reach potentiated states
Topological ordering from first passage times

large; takes a long time to reach potentiated states

small; takes a short time to reach potentiated states
Optimal synapses have a simple structure in this order:

Consider optimizing the area under the memory curve:

When states are placed in this order,

(a) $M^{pot}$ should only go from left to right
(b) $M^{dep}$ should only go from right to left
(c) We can remove shortcuts in both $M^{pot}$ and $M^{dep}$ while
   (1) preserving the order
   (2) preserving the equilibrium distribution
   (3) increasing the area

$\Rightarrow$ The area under the memory curve of any synaptic dynamical system is bounded by that of a chain with the same equilibrium distribution.

Also, we show that the area of a chain cannot exceed $O(N^{1/2}M)$ for any choice of transition rates along the chain.

$\Rightarrow$ The area under the memory curve of any synaptic dynamical system can never exceed $O(N^{1/2}M)$. 
A frontier beyond whose born no curve can cross

Area bound implies a maximal achievable memory at any finite time given N synapses with M internal states:

\[ \sqrt{N}e^{-rt/(M-1)} \]

\[ \sqrt{N}(M - 1)/ert \]

Chains with different transition rates come close to the frontier at late times.

Various measures of memory (area, frontier, lifetime) grow linearly with the number of internal states M, but grow only as the square root of the number of synapses N.
The dividends of understanding synaptic complexity

(Under review: cerebellar learning with complex synapses)

A framework for interpreting molecular neurobiology data

Neurobiology

A theory of complex synapses

Technology

Mathematics

The next generation of artificial neural networks?
(Spatiotemporal credit assignment)
(Learning as message passing)

New theorems about perturbations to stochastic processes.
(Tighter bounds)
Acknowledgements

Subhaneil Lahiri
a.k.a.
“Subhy”

Interesting conversations:
Larry Abbott
Stefano Fusi
Marcus Bena
Jascha Sohl-Dickstein
David Sussillo

Poster F50 Tonight

Funding

National Science Foundation
Bio-X Neuroventures
Burroughs-Wellcome
DARPA
Genentech
The Alfred P. Sloan Foundation
Stanford Mind, Brain, and Computation
The Swartz Foundation
NIPS Thanks Its Sponsors