University of BRISTOL

Recurrent neural networks in brains and machines

Seán Froudist-Walsh

Lecturer in Computational Neuroscience

1

1

Why have recurrent connections?

2

2

A popular working memory task (n-back)

- We will present a series of letter on the screen, one at a time.
- Whenever the letter on the screen is the same as the previous letter, say ‘now’.

3

3

4

Feedforward neural networks cannot do this easy task

- But do you fancy a challenge?
- Rewind the video, and now respond any time the letter on the screen is the same as the one shown *three* letters back (e.g. **A**, B, C, **A**).

5

Intended Learning Outcomes

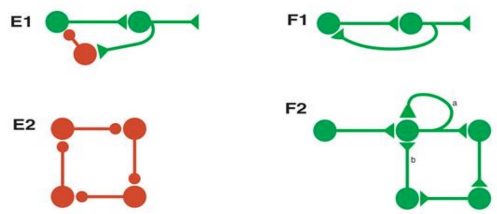
By the end of this video you will be able to

- explain why recurrent connections can be useful
- describe where most recurrent connections are in the brain
- compare biologically-realistic and trained recurrent neural networks
- design a dynamical model of the whole cortex step-by-step
- write the equation for AI-style recurrent neural networks (RNNs) and leaky RNNs
- formally compare feedforward and recurrent neural networks

6

What are recurrent connections?

- Recurrent connections are neural pathways that enable information to travel in loops, to be maintained, integrated or to re-enter the system.
- Recurrent connections can be within a brain area (local) or between brain areas (long-range).
- Recurrent connections can be excitatory or inhibitory



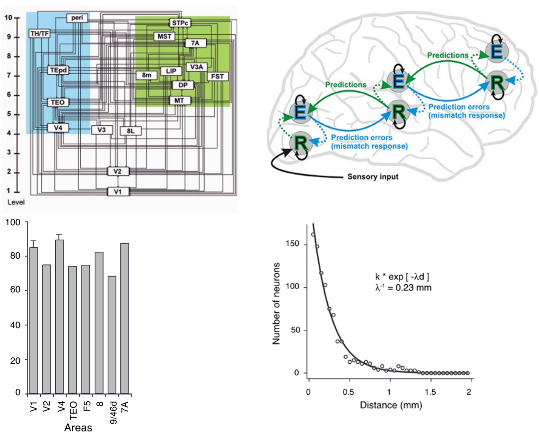
Byrne, 2023
https://nba.uth.tmc.edu/neuroscience/m/s1/introduction.html

7

7

Where are the recurrent connections in the brain?

- There are many long-distance recurrent loops across brain areas.
- In primates, approximately 2/3 of all possible connections between brain areas do exist (~97% in mice).
- However most connections (~80%) are from within the same brain area (in the cortex)



Markov et al., *Cerebral Cortex*., 2011 Markov et al., *J. Comp. Neurol.*, 2014 Stefanics et al., *Front. Hum. Neurosci, 2014*

8

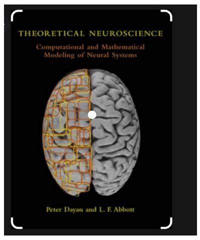
8

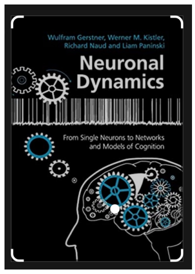
A parallel tradition of biologically-realistic models of brain functions

V1 - vision

prefrontal – working memory

hippocampus - memory





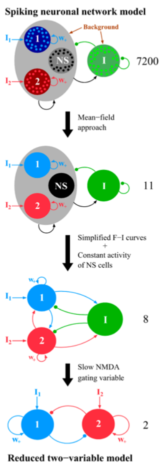
prefrontal/parietal - choice
hippocampus - memory
V1 - vision

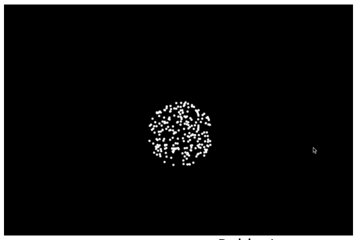
9

9

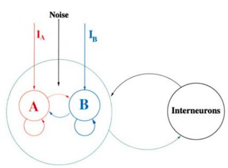
Biologically-realistic models link physiology to behaviour

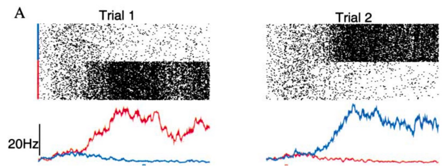
- Biologically-realistic models can be detailed models of many spiking neurons, or simplified (via mean-field techniques) to firing rate models of populations of cells
- Detailed spiking models can lead to clear experimental predictions.
- Simplified firing rate models allow mathematical analysis and deeper intuitions
- Both have been successful at explaining neural activity and behavior.
- The equations of simplified models can be similar to trained (AI-style) RNNs
- The main difference in implementation is that the parameters are not trained but taken directly from neuroscience experiments, or chosen to represent a hypothesis about how the brain computes
- Currently biologically-realistic models make more directly testable predictions for experiments, but are less capable of high-level functions compared to trained RNNs.





Bobby Ines





Amit & Brunel, *Cerebral Cortex*, 1997; Wang, *Neuron,* 2002; Wong & Wang, *J. Neurosci.*, 2006

10

10

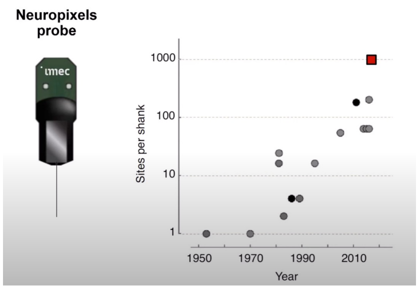
Quiz – blackboard

- pros and cons of using biologically-realistic models versus trained RNNs to simulate a cognitive task

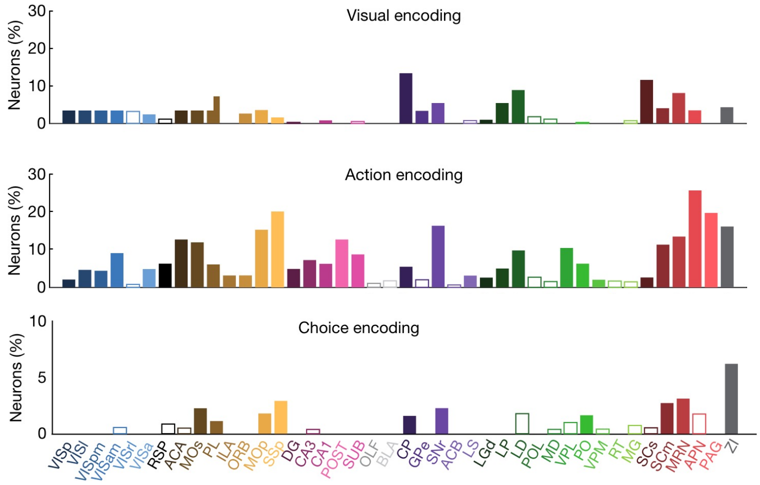
11

11

Most brain functions emerge from interactions of many areas – a challenge for local RNNs



A revolution in neural recording technology



V1 M1 hippocampus subcortical

PFC S1

image: Matteo Carandini

Jun et al., *Nature*, 2017

Steinmetz et al., *Nature*, 2019

12

12

How do we build a dynamical model of the whole cortex?

Step 1/6: building blocks – local circuit models

13

13

Canonical local cortical circuit

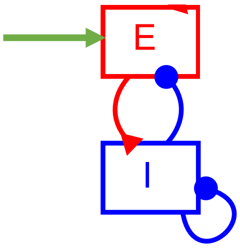
"Our view is that the rapid evolutionary expansion of neocortex has been made possible by building an 'isocortex' — a structure that uses repeats of the same basic local circuits throughout a single [cortical] sheet."

RJ Douglas and KA Martin (2012)

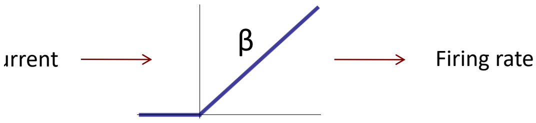
14

14

A canonical local circuit model



Excitatory and inhibitory populations



$$\tau \frac{dr_{x,E}}{dt} = -r_{x,E} + \beta_E \left[w_{E,E} r_{x,E} + G \sum_{y \in Y} C_{y \to x} r_{y,E} - w_{E,I} r_{x,I} \right]_+$$

At each step in time the firing rate of the Excitatory neurons in area x would drop down towards zero if not for synaptic inputs.

Activity is pushed up by

- positive connections from the Excitatory population to itself
- and long-range inputs from other brain areas y

Activity is pushed down by

- negative inputs from the Inhibitory population

The size of the response to input depends on the slope.

Chaudhuri et al., *Neuron*, 2015
w, β, τ set to match
Binzegger et al (2009)

The time constant determines how quickly the rate can rise and fall in response to input
Firing rates cannot be negative.

15

15

How do we build a dynamical model of the whole cortex?

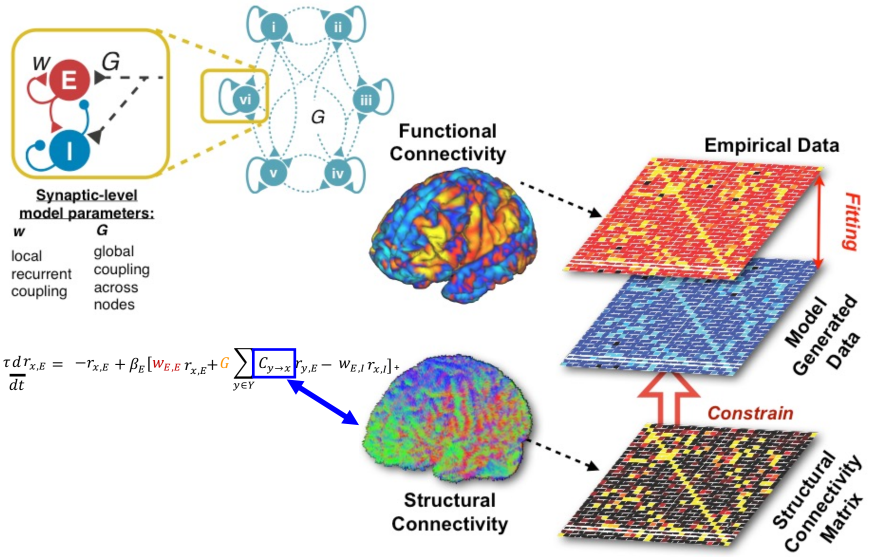
Step 1/6: building blocks – local circuit models

Step 2/6: connect the blocks – anatomical connectivity data

16

16

Modeling large-scale resting-state networks



Gustavo Deco and others

17

17

How do we build a dynamical model of the whole cortex?

Step 1/6: building blocks – local circuit models

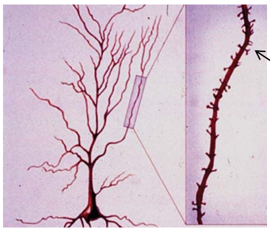
Step 2/6: connect the blocks – anatomical connectivity data

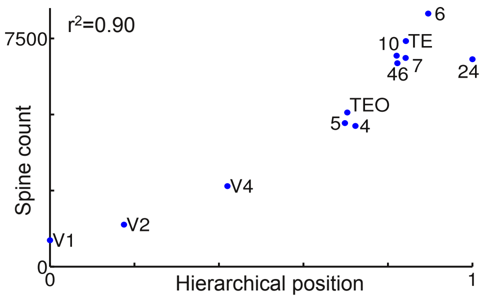
Step 3/6: (large-scale 2.0) allow local variation of circuit properties, based on data

18

18

Brain areas do not have an identical local circuitry



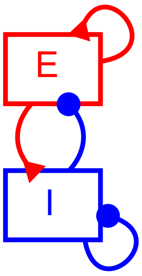


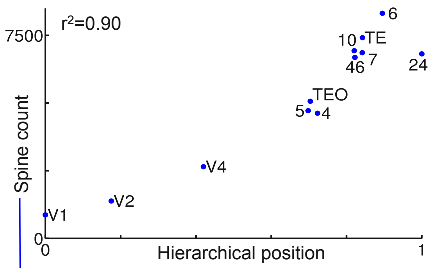
Elston, 2007

19

19

A quantitative change to parameters driven by anatomy





$$\tau \frac{d\, r_{x,E}}{dt} = -r_{x,E} + \beta_{x,E}\left[(1+bS_x)\left(w_{E,E}\, r_{x,E} + G \sum_{y\in Y} C_{y\to x}\, r_{y,E}\right) - w_{E,I}\, r_{x,I}\right]_+$$

Scale up the strength of excitatory inputs to an area based on the spine count (number of excitatory synapses per neuron)

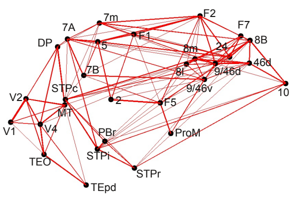
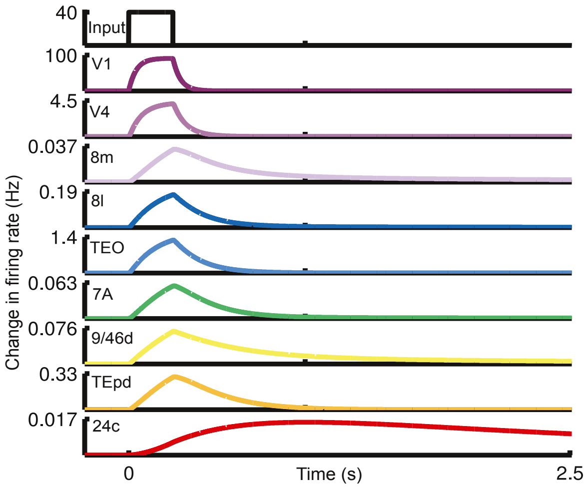
Only two free parameters

Chaudhuri et al., *Neuron*, 2015

20

20

The emergence of a hierarchy of temporal windows

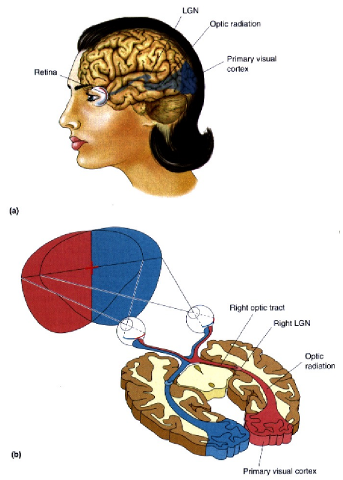



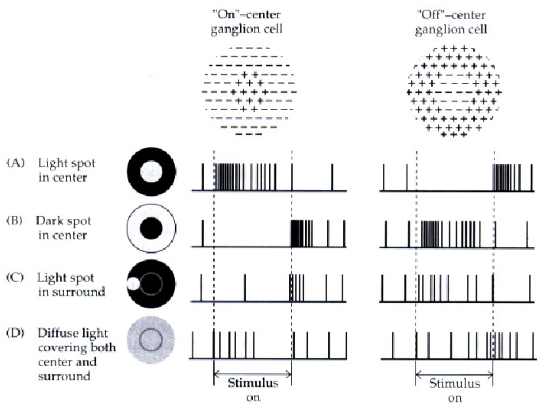
Chaudhuri et al., *Neuron*, 2015

21

21

Rapid responses in early visual areas

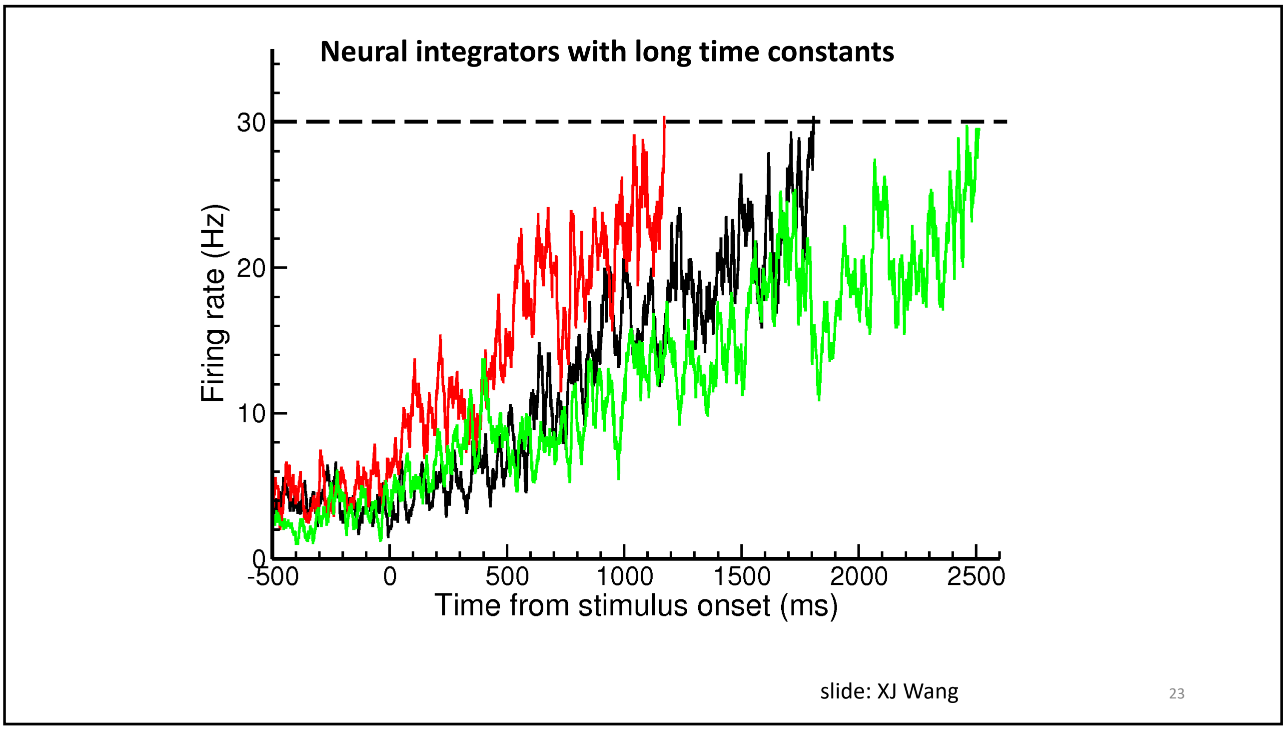




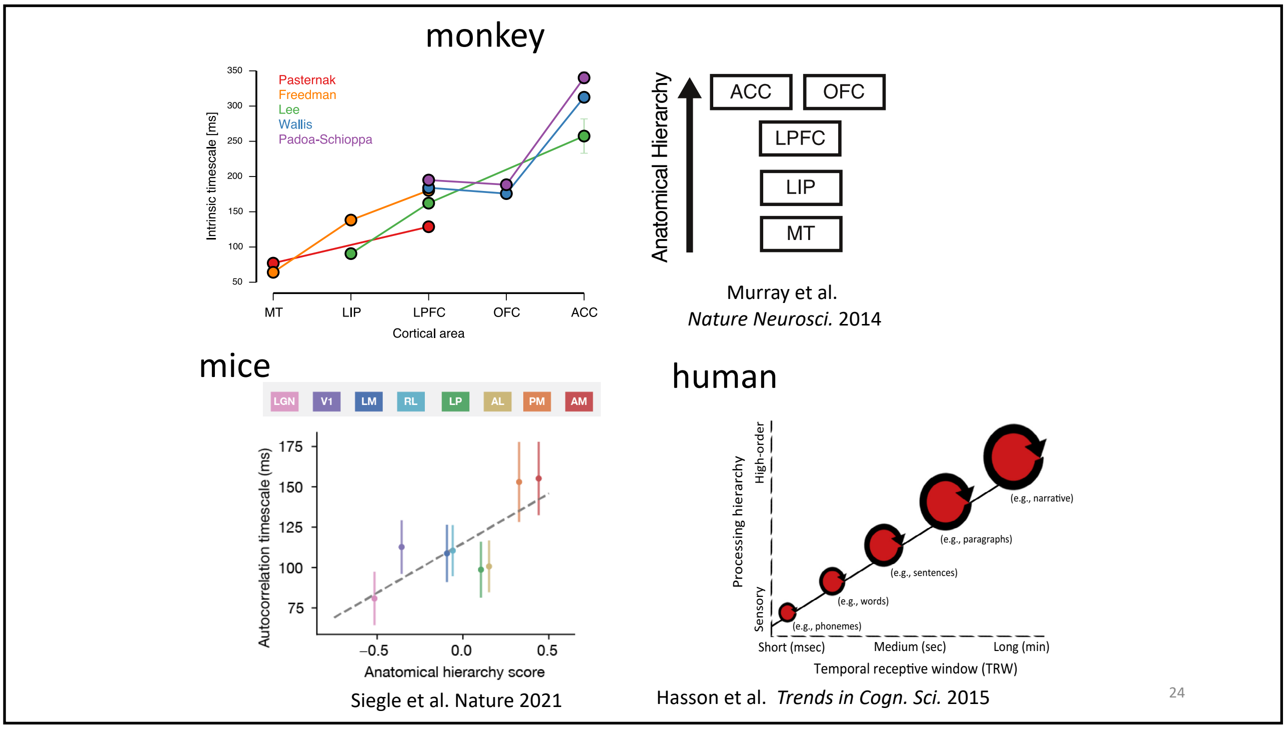
slide: XJ Wang

22

22


Neural integrators with long time constants
Firing rate (Hz)
30
20
10
0
-500
0
500
1000
1500
2000
2500
Time from stimulus onset (ms)
slide: XJ Wang
23

23

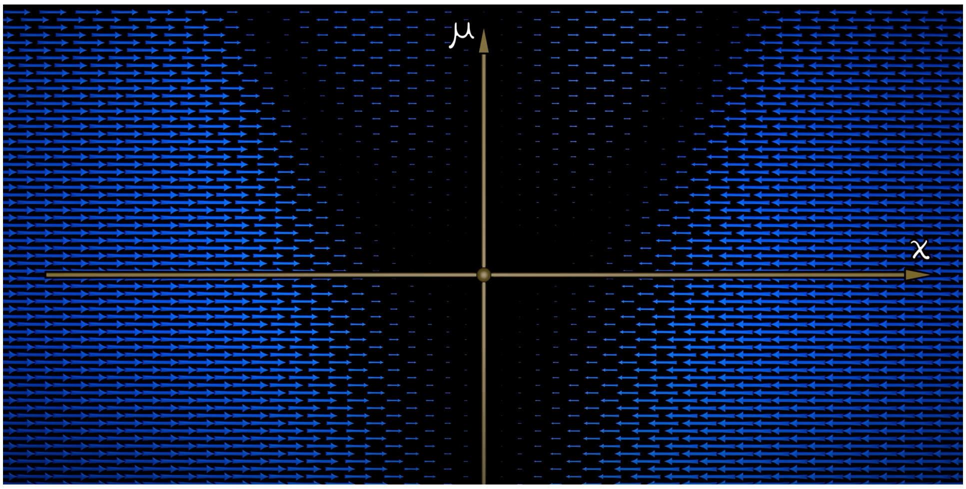

monkey
Intrinsic timescale [ms]
Pasternak
Freedman
Lee
Wallis
Padoa-Schioppa
MT
LIP
LPFC
OFC
ACC
Cortical area
Anatomical Hierarchy
ACC
OFC
LPFC
LIP
MT
Murray et al.
Nature Neurosci. 2014
mice
LGN
V1
LM
RL
LP
AL
PM
AM
Autocorrelation timescale (ms)
Anatomical hierarchy score
Siegle et al. Nature 2021
human
Processing hierarchy
High-order
Sensory
(e.g., narrative)
(e.g., paragraphs)
(e.g., sentences)
(e.g., words)
(e.g., phonemes)
Short (msec)
Medium (sec)
Long (min)
Temporal receptive window (TRW)
Hasson et al. Trends in Cogn. Sci. 2015
24

24

How do qualitatively different functions emerge across the cortex?

25

25

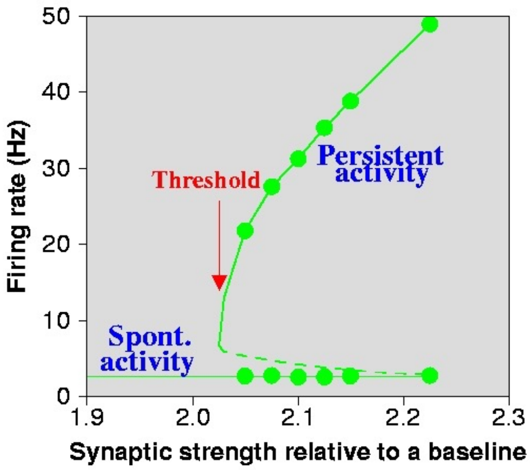


Prof. Ghrist

26

26

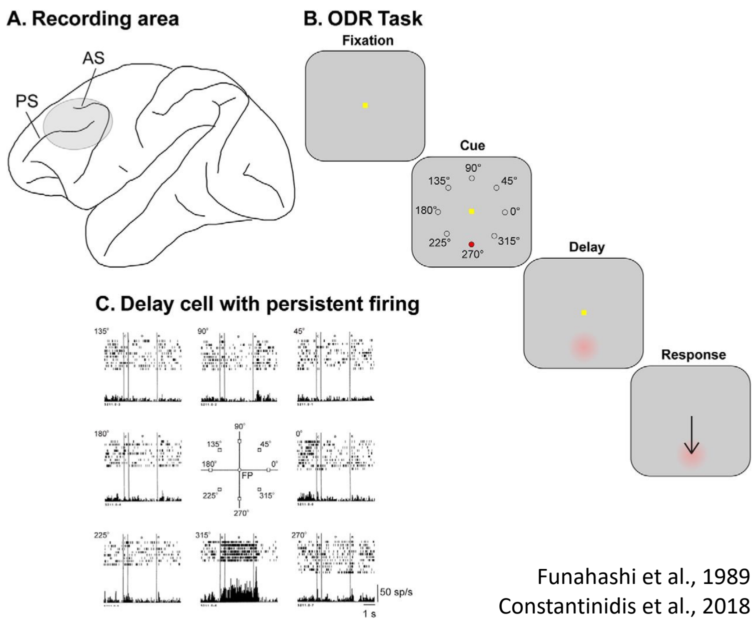
Working memory emerges with sufficiently strong recurrent connections



slide: XJ Wang

Bifurcations in *nonlinear* dynamical systems:
Graded differences give rise to qualitatively novel behavior/functions

27



Funahashi et al., 1989
Constantinidis et al., 2018

28

28

How do we build a dynamical model of the whole cortex?

Step 1/6: building blocks – local circuit models

Step 2/6: connect the blocks – anatomical connectivity data

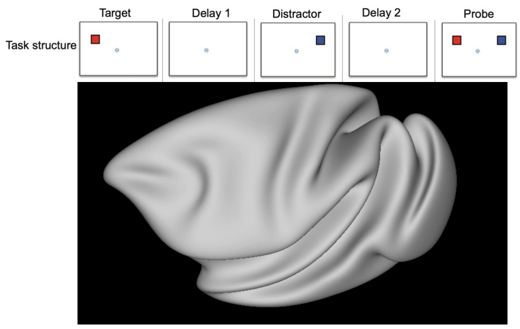
Step 3/6: (large-scale 2.0) allow local variation of circuit properties, based on data

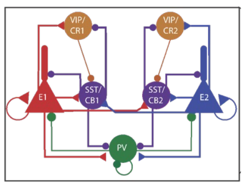
Step 4/6: (large-scale 2.0) simulate task stimuli as input & measure activity

29

29

Working memory activity across cortical areas and multiple cell types





Froudist-Walsh et al., *Neuron*, 2021
see also
Mejías & Wang, *eLife*, 2022

30

30

How do we build a dynamical model of the whole cortex?

Step 1/6: building blocks – local circuit models

Step 2/6: connect the blocks – anatomical connectivity data

Step 3/6: (large-scale 2.0) allow local variation of circuit properties, based on data

Step 4/6: (large-scale 2.0) simulate task stimuli as input & measure activity

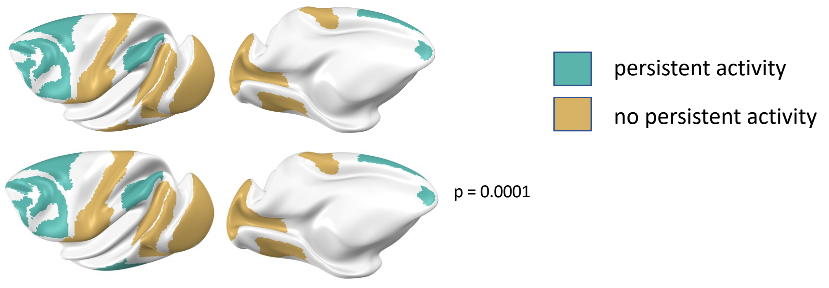
Step 5/6: (large-scale 2.0) validate model against real neural data

31

31

The model captures the persistent activity pattern of > 90 experimental studies

mega-analysis of experimental data - Leavitt et al., *TiCS*, 2017



model simulation - Froudist-Walsh et al., *Neuron*, 2021

32

32

How do we build a dynamical model of the whole cortex?

Step 1/6: building blocks – local circuit models

Step 2/6: connect the blocks – anatomical connectivity data

Step 3/6: (large-scale 2.0) allow local variation of circuit properties, based on data

Step 4/6: (large-scale 2.0) simulate task stimuli as input & measure activity

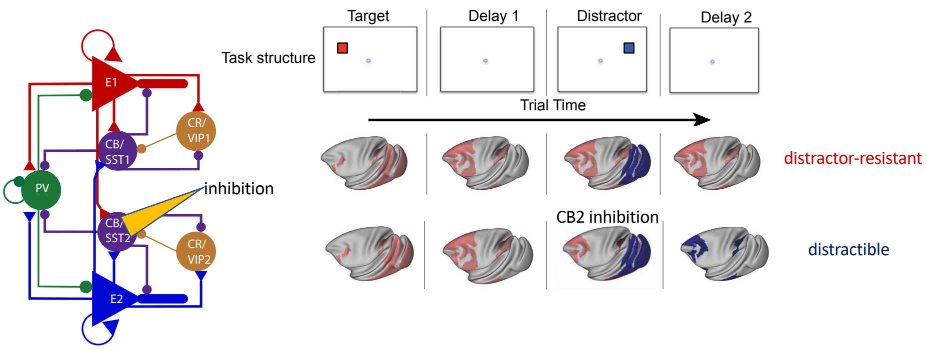
Step 5/6: (large-scale 2.0) validate model against real neural data

Step 6/6: (large-scale 2.0) make predictions for future experiments

33

33

"Zoom in" to find the cell-type responsible for distractor-resistance



Froudist-Walsh et al., *Neuron*, 2021

34

34

Now you have a dynamical model of the whole cortex and predictions for new experiments

35

35

Quiz - blackboard

What is the sequence of steps for building a whole-cortex model?

36

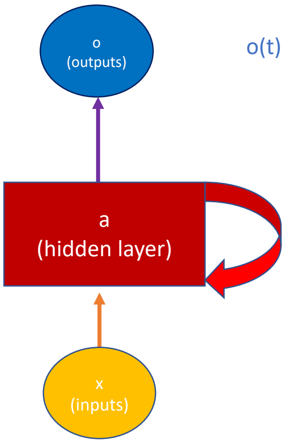
36

Recurrent neural networks (RNNs) in AI

- The basic ("vanilla") RNN

$a(t) = f(\, W_{x \to a}\, x(t) + W_{a \to a}\, a(t-1) + b_1 \,)$

$o(t) = g(\, W_{a \to o}\, a(t) + b_2 \,)$



The hidden layer activity
is a nonlinear function (e.g. ReLU, sigmoid)
of a weighted sum of the inputs
and a weighted sum
of the hidden layer activity from one timestep ago
plus a bias

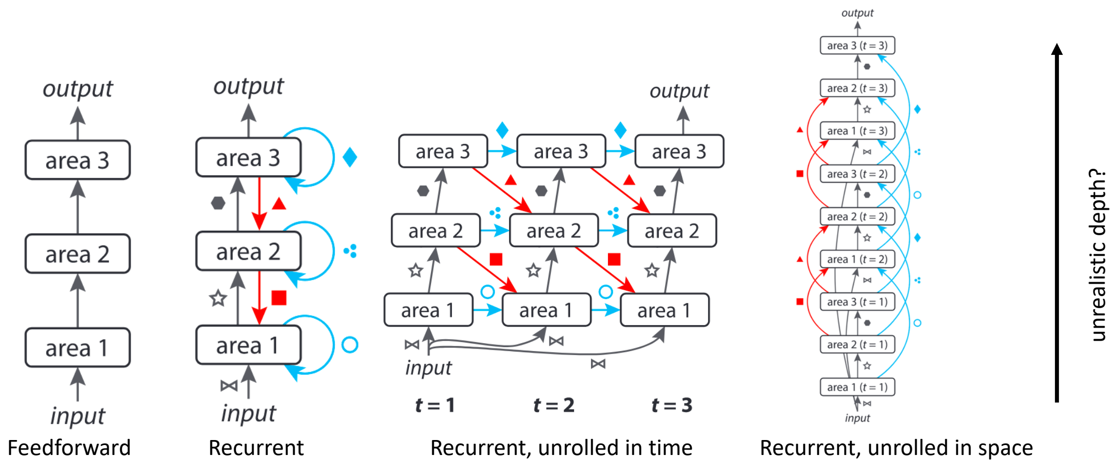
The output layer activity is a nonlinear function
is a weighted sum
of the hidden layer activity
plus a bias

Elman, *Cog. Sci.*, 1990

37

37

The mathematical equivalence of (some) feedforward and recurrent neural networks

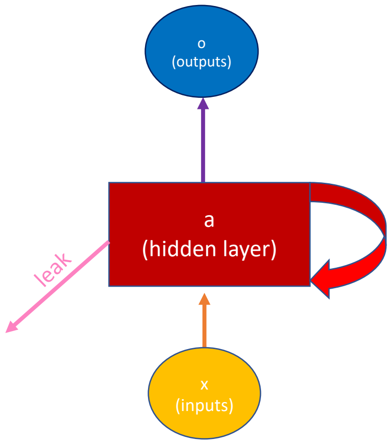


van Bergen & Kriegeskorte, *Curr. Opin. Neurobiol*, 2020

38

38

Leaky RNNs for neuro-AI



Continuous time version (written in papers)

$$\tau \frac{da}{dt} = -a(t-1) + f(\, W_{x\to a}\, x(t) + W_{a\to a}\, a(t-1) + b_1 \,)$$

Discrete time version (Euler method - written in code)

$$a(t) = a(t-1) + \frac{\Delta t}{\tau}\Big(-a(t-1) + f\,(\, W_{x\to a}\, x(t) + W_{a\to a}\, a(t-1) + b_1 \,)\Big)$$

Δt is the length of the simulation timestep. Logically, shorter timesteps lead to smaller changes in activity between timesteps.

τ is the neuronal time constant, which dictates how rapidly the activity changes in response to inputs/leak.

$$o(t) = g(\, W_{a\to o}\, a(t) + b_2 \,)$$

39

39

Quiz - blackboard

- Comparing feedforward and recurrent neural networks.

40

40

Recap

- Recurrent connections enable neural networks to use and compute on previous information, which contributes to many cognitive functions, including memory and decision-making
- The whole cortex is recurrent, but most recurrent connections are from other neurons within the same brain area.
- Biologically-realistic models make more directly testable predictions for experiments, but are less capable of high-level functions compared to trained RNNs.
- Most brain functions rely on activity across many different brain areas
- Whole cortex models are built up by connecting many local circuits and adjusting parameters across areas according to anatomy
- There is an increase in the time constants of neurons along the cortical (visual and auditory) hierarchy
- RNNs are like feedforward networks, but with inputs from the previous timestep
- Some feedforward and recurrent neural networks are mathematically equivalent
- In neuroscience, leaky RNNs are often used to allow for more biologically-realistic neural dynamics

41

41

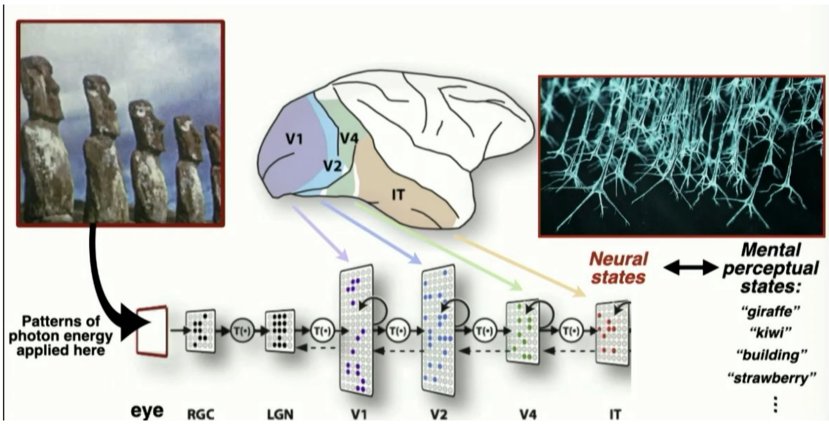
Still curious? You can dive in deeper to any of today's topics:

- Most cortical connections are from within the same brain area
  - Markov, Nikola T., P. Misery, Arnaud Falchier, C. Lamy, J. Vezoli, R. Quilodran, M. A. Gariel et al. "Weight consistency specifies regularities of macaque cortical networks." *Cerebral cortex* 21, no. 6 (2011): 1254-1272.
- An early cortex-wide dynamical model
  - Chaudhuri, Rishidev, Kenneth Knoblauch, Marie-Alice Gariel, Henry Kennedy, and Xiao-Jing Wang. "A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex." *Neuron* 88, no. 2 (2015): 419-431.
- A cortex-wide dynamical model of working memory
  - Froudist-Walsh, Sean, Daniel P. Bliss, Xingyu Ding, Lucija Rapan, Meiqi Niu, Kenneth Knoblauch, Karl Zilles, Henry Kennedy, Nicola Palomero-Gallagher, and Xiao-Jing Wang. "A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex." *Neuron* 109, no. 21 (2021): 3500-3520.
- Comparing feedforward and recurrent neural networks
  - van Bergen, Ruben S., and Nikolaus Kriegeskorte. "Going in circles is the way forward: the role of recurrence in visual inference." *Current Opinion in Neurobiology* 65 (2020): 176-193.
- The "vanilla" RNN paper
  - Elman, Jeffrey L. "Finding structure in time." *Cognitive science* 14, no. 2 (1990): 179-211.

42

42

Next: training deep neural networks and comparisons with brains



video: James DiCarlo (MIT)
Yamins et al., *PNAS*, 2014

43

43