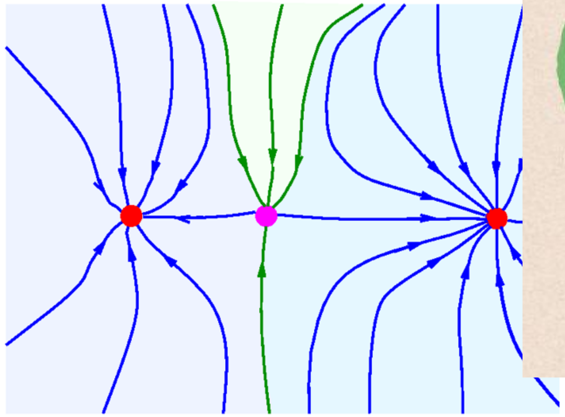


Schizophrenia, Attractors and Working Memory

Peggy Seriès, IANC
Informatics, University of Edinburgh, UK

pseries@inf.ed.ac.uk

CCN Lecture 6

Summary of previous lecture

- Schizophrenia is a very serious illness characterised by “positive” (hallucinations, delusions) and negative symptoms.
- One neurobiological correlate of the illness is impairment in working memory.

- Short-term/ working memory: Dynamic process - “Sustained” a.k.a. “Delay” or “Persistent” Activity.

- Attractor Networks as (main) model of working memory / sustained activity

- Hopfield Network as example of a point attractor model (Lab 2).

Computational Cognitive Neuroscience. Lab 2

Hopfield Networks. February 2025

Lecturer: Peggy Seriès
Teaching Assistant: Lars Werne

Tutorial Objectives

In this tutorial, you will:

- Learn to implement an associative memory system – the Hopfield network.
- Explore the pattern-completing properties of Hopfield networks.
- Implement synaptic pruning into the model, as a putative computational framework for Schizophrenia.

Introduction

In this tutorial, you will code and simulate a fundamental neuron *population* model, which we discussed in Lecture 5: the Hopfield Network. Hopfield networks are an early kind of *attractor network*, which have been finding great acclaim as models of *associative memory* in the brain. You will explore the model's ability to recall stored activity patterns from partial or noisy inputs. You will then incorporate synaptic pruning – the systematic deletion of synapses – into the network, and discuss how this process could relate to the emergence of Schizophrenia.

Towards a theory of Working Memory/ Sustained Activity

- A theory of working memory should answer:

- How it is initiated?
- Why does it persist ?
- What makes it specific?
- How does it end?

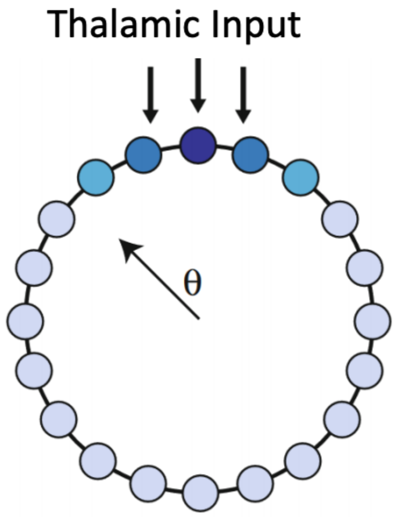
- Reason for capacity limit?
- Relationship with attention, long-term memory?

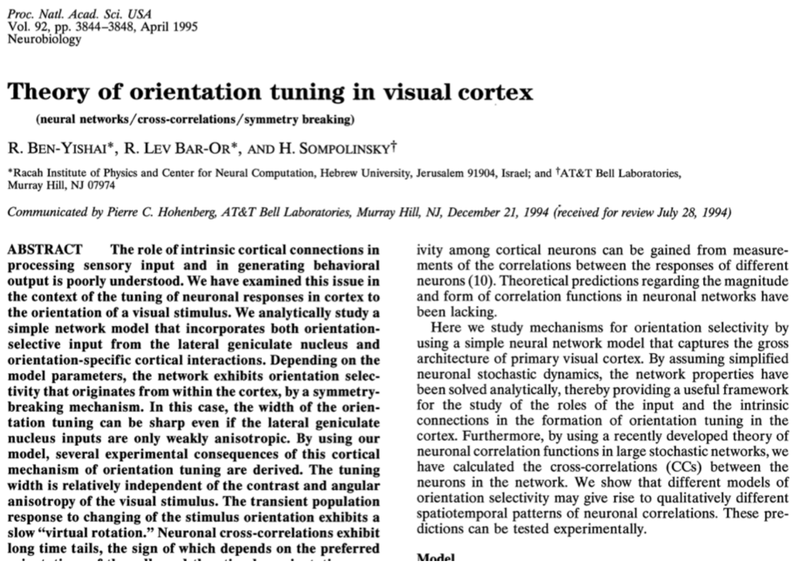
- Mechanism : **reverberations** through connections (which?), or cellular?
- Lots of experimental and theoretical work to answer these questions

How to build biologically realistic attractor models?

- Recently, effort to build biophysically plausible models of sustained activity / attractor dynamics for memory.

- Ring Model offers starting point.
- Originally model of V1 orientation selectivity, but anatomical organization of PFC also resembles a recurrent network.




Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 3844–3848, April 1995
Neurobiology

Theory of orientation tuning in visual cortex

(neural networks/cross-correlations/symmetry breaking)

R. Ben-Yishai*, R. Lev Bar-Or*, and H. Sompolinsky†

*Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel; and †AT&T Bell Laboratories, Murray Hill, NJ 07974

Communicated by Pierre C. Hohenberg, AT&T Bell Laboratories, Murray Hill, NJ, December 21, 1994 (received for review July 28, 1994)

ABSTRACT The role of intrinsic cortical connections in processing sensory input and in generating behavioral output is poorly understood. We have examined this issue in the context of the tuning of neuronal responses in cortex to the orientation of a visual stimulus. We analytically study a simple network model that incorporates both orientation-selective input from the lateral geniculate nucleus and orientation-specific cortical interactions. Depending on the model parameters, the network exhibits orientation selectivity that originates from within the cortex, by a symmetry-breaking mechanism. In this case, the width of the orientation tuning can be sharp even if the lateral geniculate nucleus inputs are only weakly anisotropic. By using our model, several experimental consequences of this cortical mechanism of orientation tuning are derived. The tuning width is relatively independent of the contrast and angular anisotropy of the visual stimulus. The transient population response to changing of the stimulus orientation exhibits a slow "virtual rotation." Neuronal cross-correlations exhibit long time tails, the sign of which depends on the preferred

ivity among cortical neurons can be gained from measurements of the correlations between the responses of different neurons (10). Theoretical predictions regarding the magnitude and form of correlation functions in neuronal networks have been lacking.

Here we study mechanisms for orientation selectivity by using a simple neural network model that captures the gross architecture of primary visual cortex. By assuming simplified neuronal stochastic dynamics, the network properties have been solved analytically, thereby providing a useful framework for the study of the roles of the input and the intrinsic connections in the formation of orientation tuning in the cortex. Furthermore, by using a recently developed theory of neuronal correlation functions in large stochastic networks, we have calculated the cross-correlations (CCs) between the neurons in the network. We show that different models of orientation selectivity may give rise to qualitatively different spatiotemporal patterns of neuronal correlations. These predictions can be tested experimentally.

Model

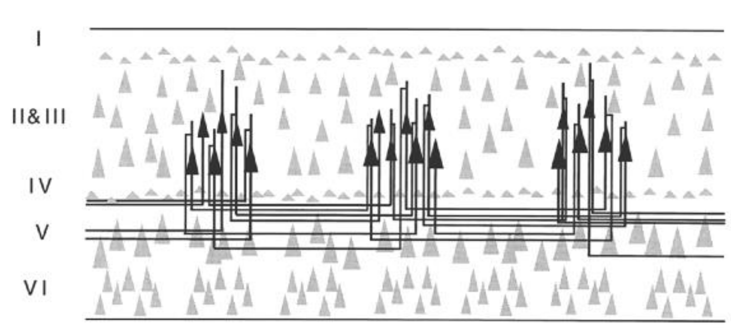
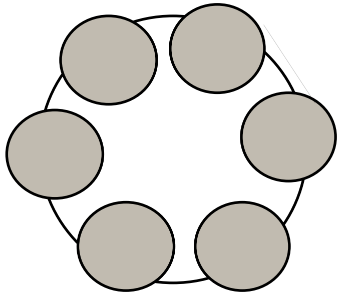


Fig. 4. Schematic diagram illustrating the pattern of connections between prefrontal neurons in the superficial layers. The figure summarizes results of anatomical tracer injection experiments and retrograde labeling. From Kritzer and Goldman-Rakic (1995), with permission.

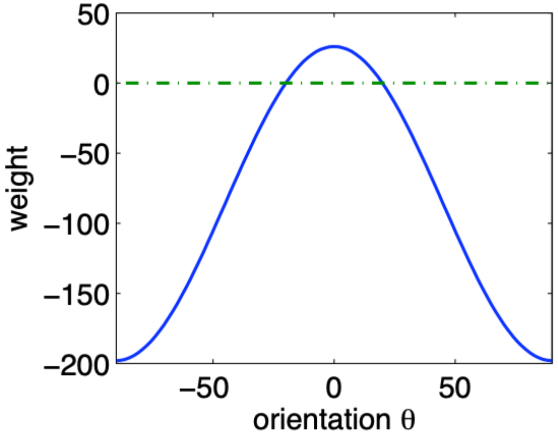
Back to the Ring Model of orientation selectivity

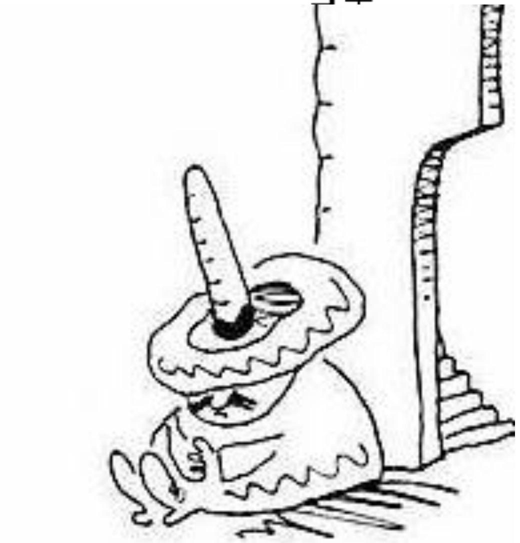


- N neurons, with preferred angle, θ_i ,evenly distributed between $-\pi/2$ and $\pi/2$
- Neurons receive thalamic inputs *h.*

\+ recurrent connections, with excitatory weights between nearby cells and inhibitory weights between cells that are further apart (mexican-hat profile)

$$\tau_r \frac{dv(\theta)}{dt} = -v(\theta) + \left[h(\theta) + \int_{-\pi/2}^{\pi/2} \frac{d\theta'}{\pi} \left(-\lambda_0 + \lambda_1 \cos(2(\theta - \theta')) \right) v(\theta') \right]_+$$

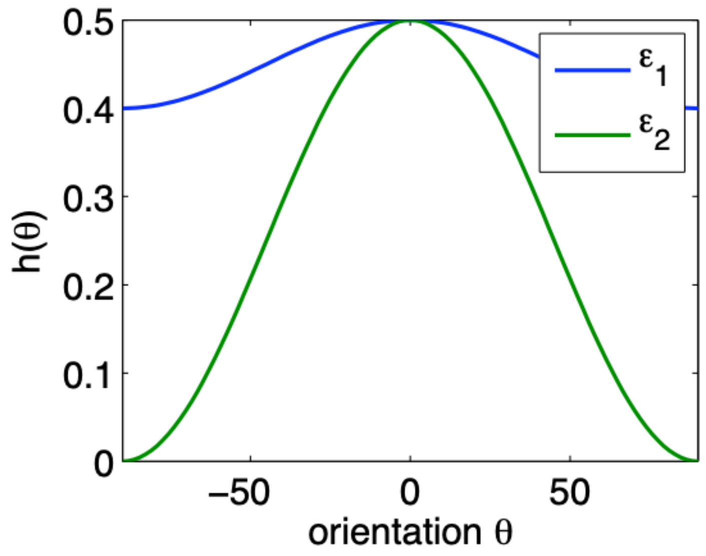




Back to the Ring Model of orientation selectivity

- h is input, can be tuned (Hubel Wiesel scenario) or very broadly tuned.

$$h(\theta) = c[1 - \epsilon + \epsilon * cos(2\theta)]$$



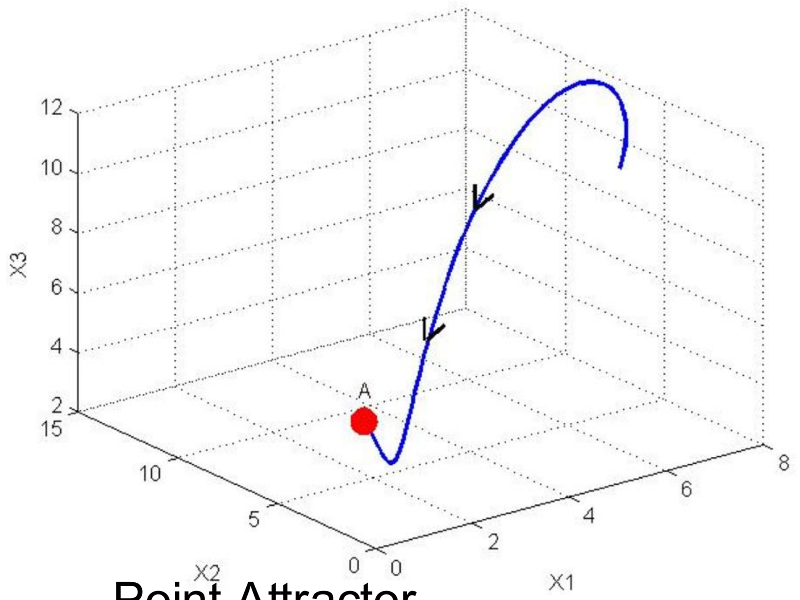
- The steady-state can be solved analytically. Model analyzed like a physical system.

- Model achieves i) orientation selectivity; ii) contrast invariance of tuning, even if input is very broad.
- The width of orientation selectivity depends on the shape of the Mexican-hat but is independent of the width of the input.
- Symmetry breaking /Attractor dynamics.

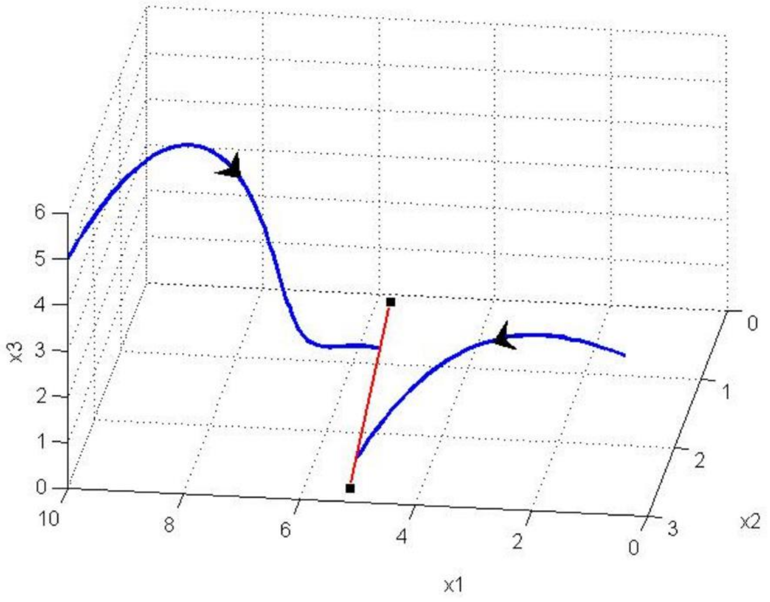
The ring model is a line attractor

- Attractor network : a network of neurons, usually recurrently connected, whose time dynamics settle to a stable pattern.
- That pattern may be stationary (fixed points), time-varying (e.g. cyclic), or even stochastic-looking (e.g., chaotic).
- The particular pattern a network settles to is called its 'attractor'.

- The ring model is called a line (or ring) attractor network. Its stable states are also sometimes referred to as 'bump attractors'.



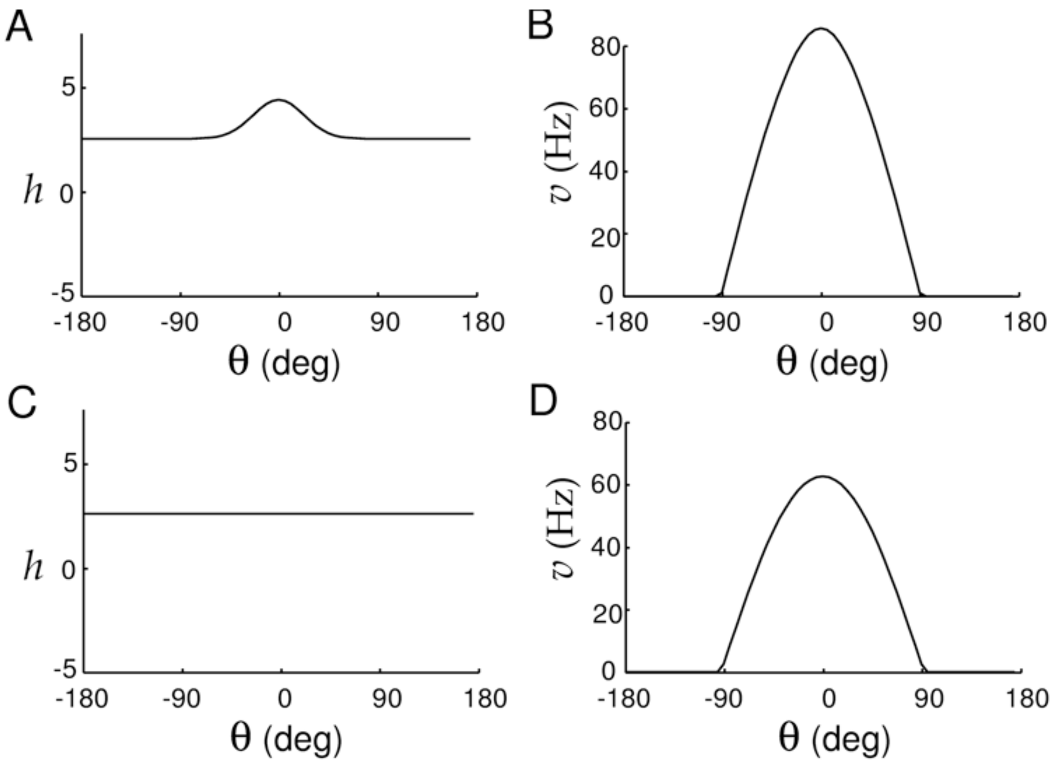
Point Attractor



Line Attractor

The ring model displays delayed activity

- If recurrent connections are strong enough, the pattern of population activity once established can become independent of the structure of the input. It can persist when input is removed.
- A model of working memory ?



More biologically realistic attractor models?

• Problems with firing rate models: difficult to relate with electrophysiological data, can't address the question of issue of spontaneous vs persistent activity, and dynamical properties of synaptic interactions are ignored.

• Can we create biophysical realistic/spiking models where recurrent networks can give rise to location-specific, persistent discharges ? (Compte et al 2000, Gutkin et al 2000, Tegner et al 2002, Renart et al 2003a, Wang et al 2004)

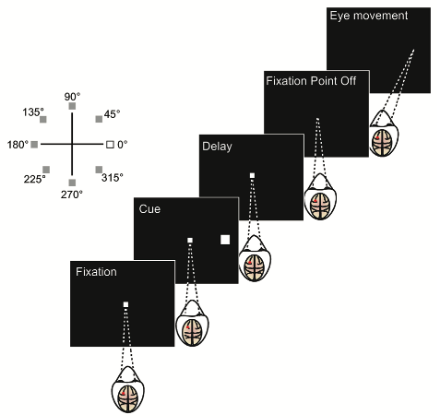
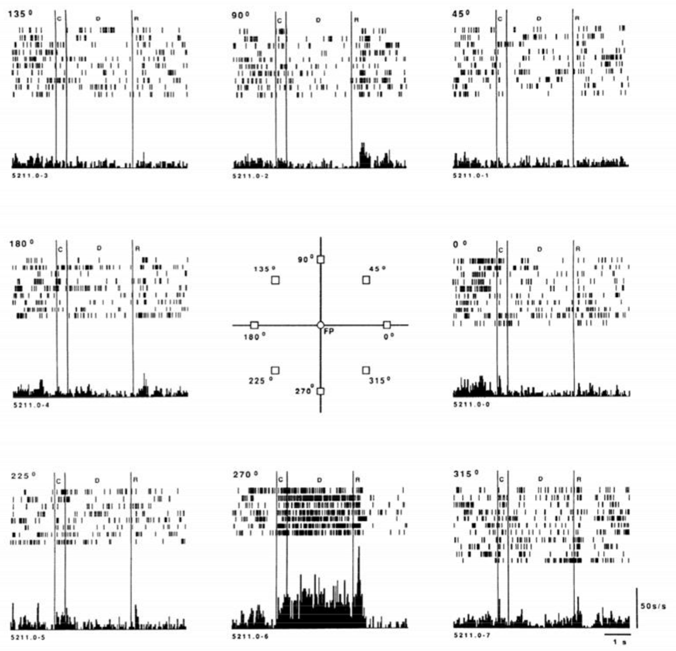


Fig. 1. Successive frames illustrate the sequence of events in the oculomotor delayed-response task. Trials begin with the appearance of a fixation point at the center of the screen, which the monkey is required to foveate throughout the trial. A spatial cue is subsequently presented, typically at one of eight locations (*left*). After a delay period of a few seconds, the fixation point is turned off and the monkey is required to indicate the location of the cue by moving his eyes accordingly on the screen.



Funahashi et al, 1989

4) Towards a Biophysical Model of Working Memory:

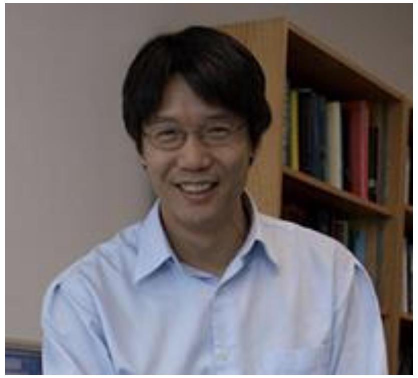
Does a ring model with spiking neurons also show delayed activity?

In spiking networks, challenges:

- Stability of delay activity
- runaway excitation
- Accounting for spontaneous activity in addition to memory state
- Oscillations can destabilise the memory activity.

Solution

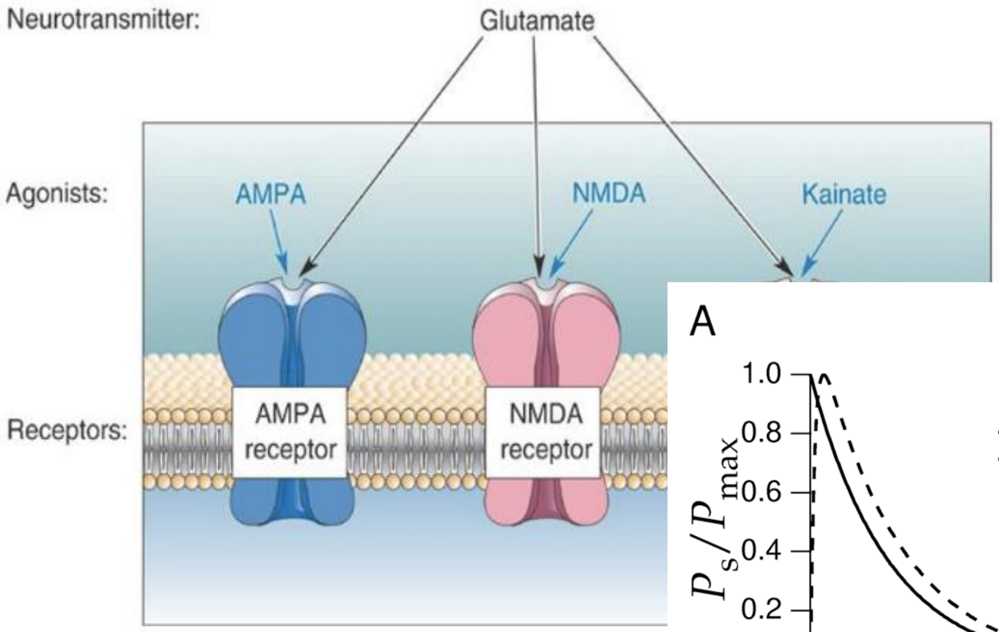
Working memory found particularly stable when excitatory reverberations are characterised by slow time course, e.g. when synaptic transmission mediated by **NMDA receptors** (prediction)



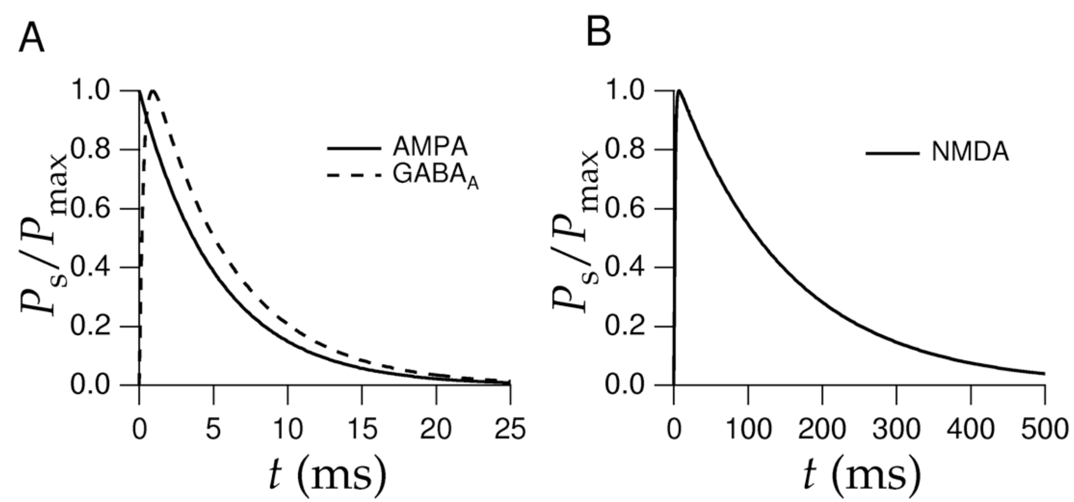
X-J Wang (NYU)

N-methyl-d-aspartate (NMDA)

- NMDA receptor is a glutamate receptor, the human brain's primary excitatory neurotransmitter. Crucial for learning, memory, and neuroplasticity
- Different synapses have different dynamics : in excitatory synapses: AMPA is fast, NMDA slow.



$$C_m \frac{dV_i(t)}{dt} = -\sum_j g_{ij}\left(t-\tau_{ij}\right)\left(V_i(t)-E_{\mathrm{EXCIT}}\right) - \sum_j g_{ij}\left(t-\tau_{ij}\right)\left(V_i(t)-E_{\mathrm{INHIB}}\right)$$
$$-g_{\mathrm{LEAK}}\left(V_i(t)-E_{\mathrm{LEAK}}\right) - g_{\mathrm{AHP}}(t)\left(V_i(t)-E_{\mathrm{AHP}}\right).$$

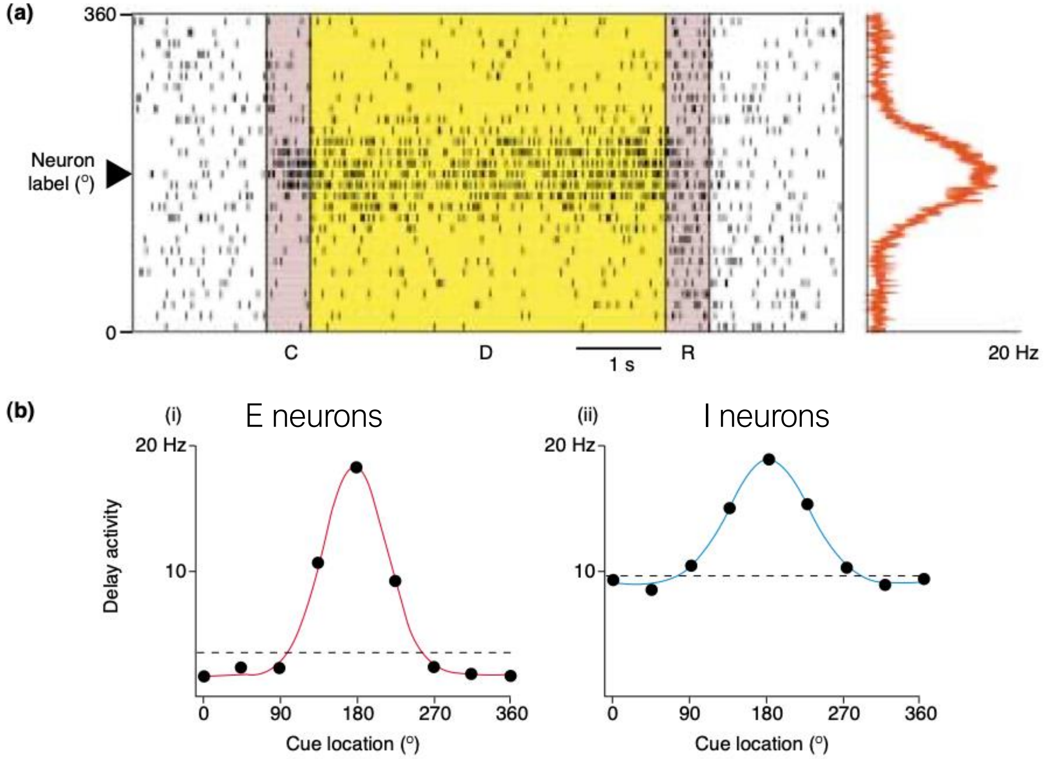


Towards a Biophysical Model of Working Memory

Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model

Compte, Brunel, Goldman-Rakic and Wang, Neuron, 2000

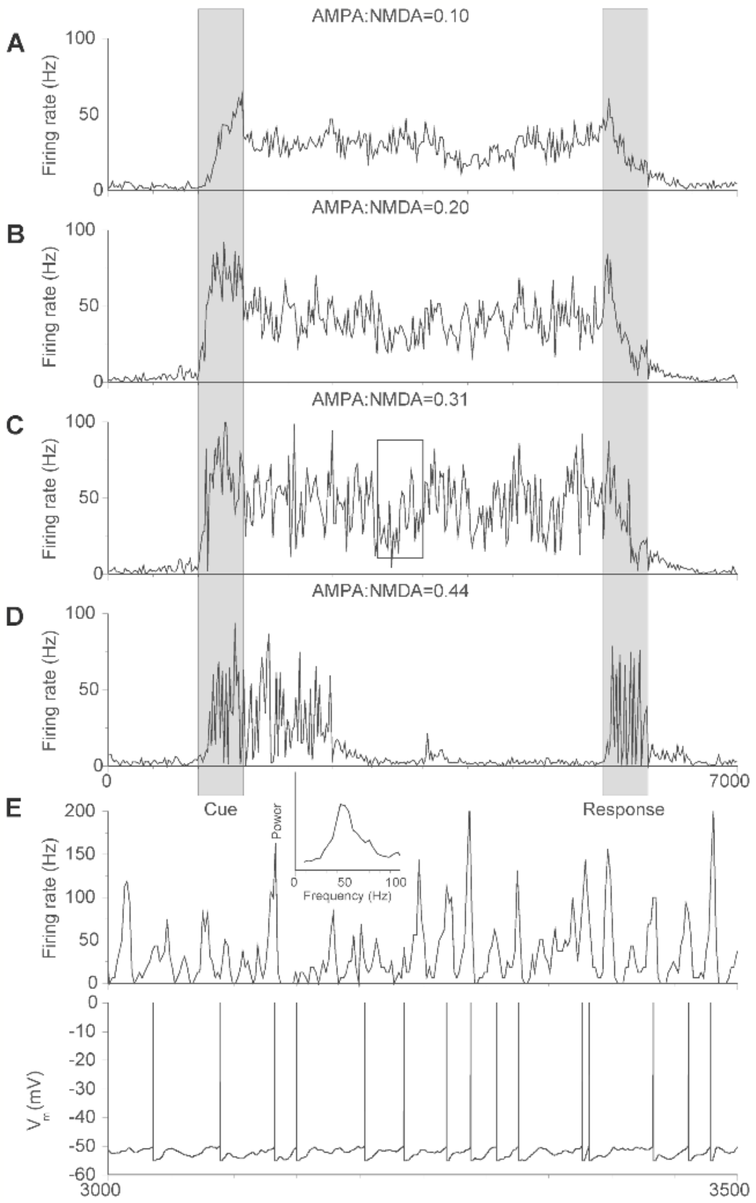
Single-neuron recordings from behaving primates have established a link between working memory processes and information-specific neuronal persistent activity in the prefrontal cortex. Using a network model endowed with a columnar architecture and based on the physiological properties of cortical neurons and synapses, we have examined the synaptic mechanisms of selective persistent activity underlying spatial working memory in the prefrontal cortex. Our model reproduces the phenomenology of the oculomotor delayed-response experiment of Funahashi *et al.* (S. Funahashi, C.J. Bruce and P.S. Goldman-Rakic, Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349, 1989). To observe stable spontaneous and persistent activity, we find that recurrent synaptic excitation should be primarily mediated by NMDA receptors, and that overall recurrent synaptic interactions should be dominated by inhibition. Iso-directional tuning of adjacent pyramidal cells and interneurons can be accounted for by a structured pyramid-to-interneuron connectivity. Robust memory storage against random drift of the tuned persistent activity and against distractors (intervening stimuli during the delay period) may be enhanced by neuromodulation of recurrent synapses. Experimentally testable predictions concerning the neural basis of working memory are discussed.



- Network of ~2500 integrate-and-fire neurons, mexican-hat connectivity, **NMDA excitation.**
- Reproduce Funahashi et al 1989.
- Selectivity of memory field, temporal drifts, robustness to distractors, co-existence with spontaneous activity .

Towards a Biophysical Model of Working Memory

Fig. 6. Stability of persistent activity as a function of the AMPA:NMDA ratio at the recurrent excitatory synapses. *A–D*, Temporal course of the average firing rate across a subpopulation of cells selective to the presented transient input, for different levels of the AMPA:NMDA ratio. As the ratio is increased, oscillations of a progressively larger amplitude develop during the delay period, which eventually destabilize the persistent activity state. *E*, Snapshot of the activity of the network in (*C*) between 3 and 3.5 seconds. *Top*, Average network activity. *Bottom*, Intracellular voltage trace of a single neuron. *Inset*, Power spectrum of the average activity of the network, showing a peak in the γ (40 Hz) frequency range. Persistent activity is stable even in the presence of synchronous oscillations. Adapted with permission from Renart, Brunel, and others (2003).

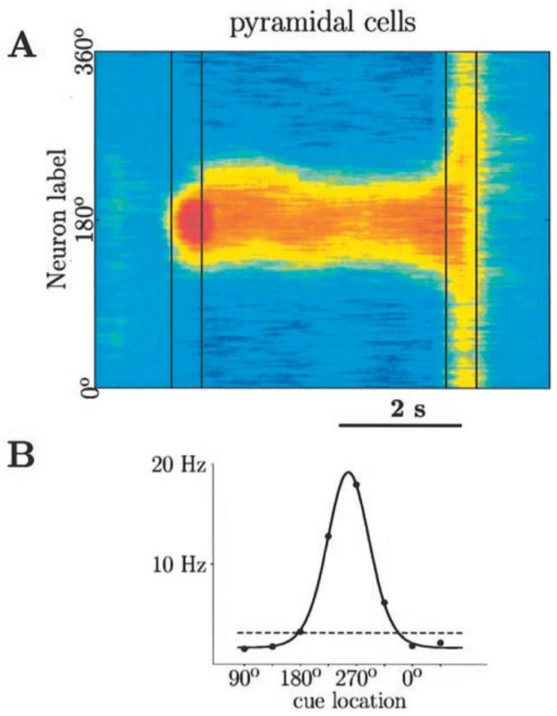


- The network dynamics and stability are sensitive to the ratio between AMPA and NMDA synapses.
- NMDA is proved crucial for persistent activity.

[Renart, Brunel, Wang , 2003]

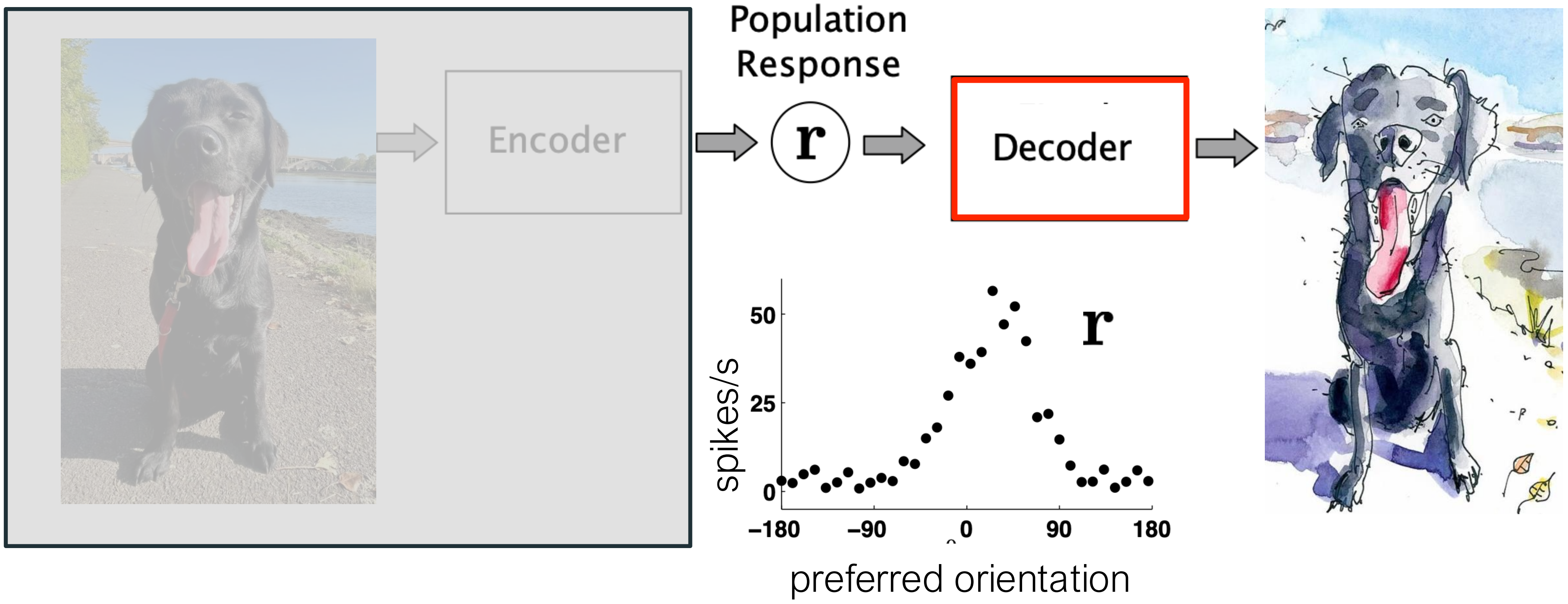
Towards a Biophysical Model of Working Memory

- A mechanism for switching the memory off: excitatory input to a large population of neurons in the network.
- Decoding can be used to infer what the memory is encoding, e.g. population vector (decoding the “center” of the memory “bump”).



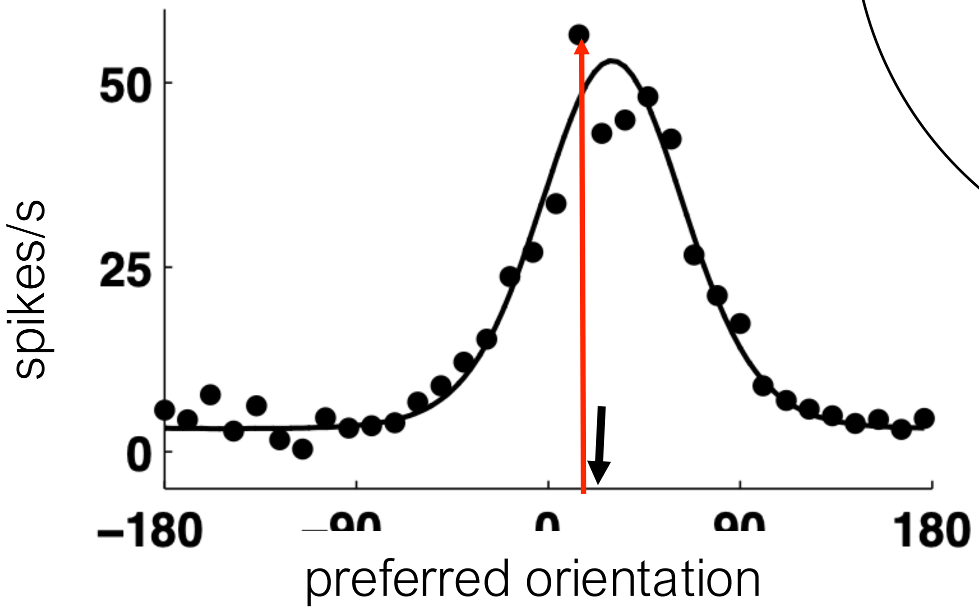
(Refresher) Decoding populations of neurons

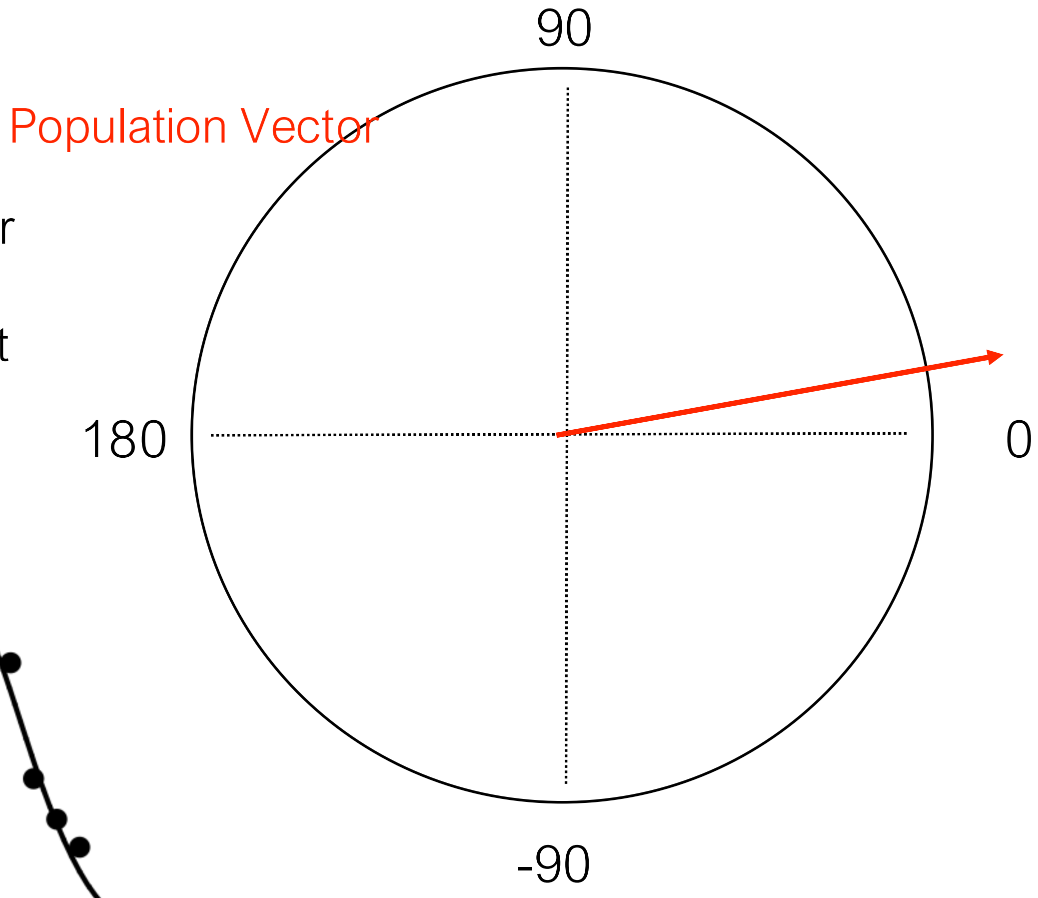
In response to a stimulus s, we observe a pattern of activity **r** (e.g. in V1).
What can we say about s given **r**?



(Refresher) A Simple Decoding Strategy: Population Vector

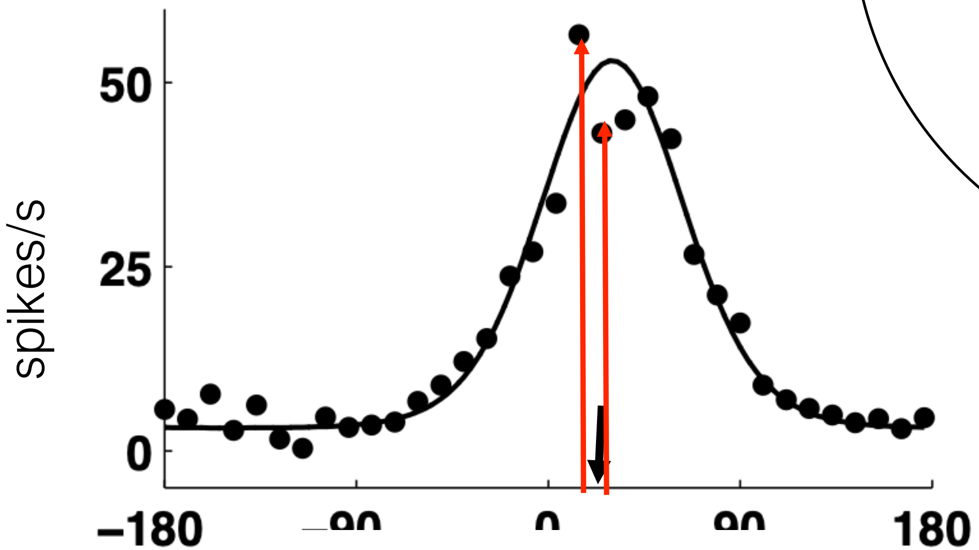
Each neuron “votes” for its preferred orientation with length given by its spike count

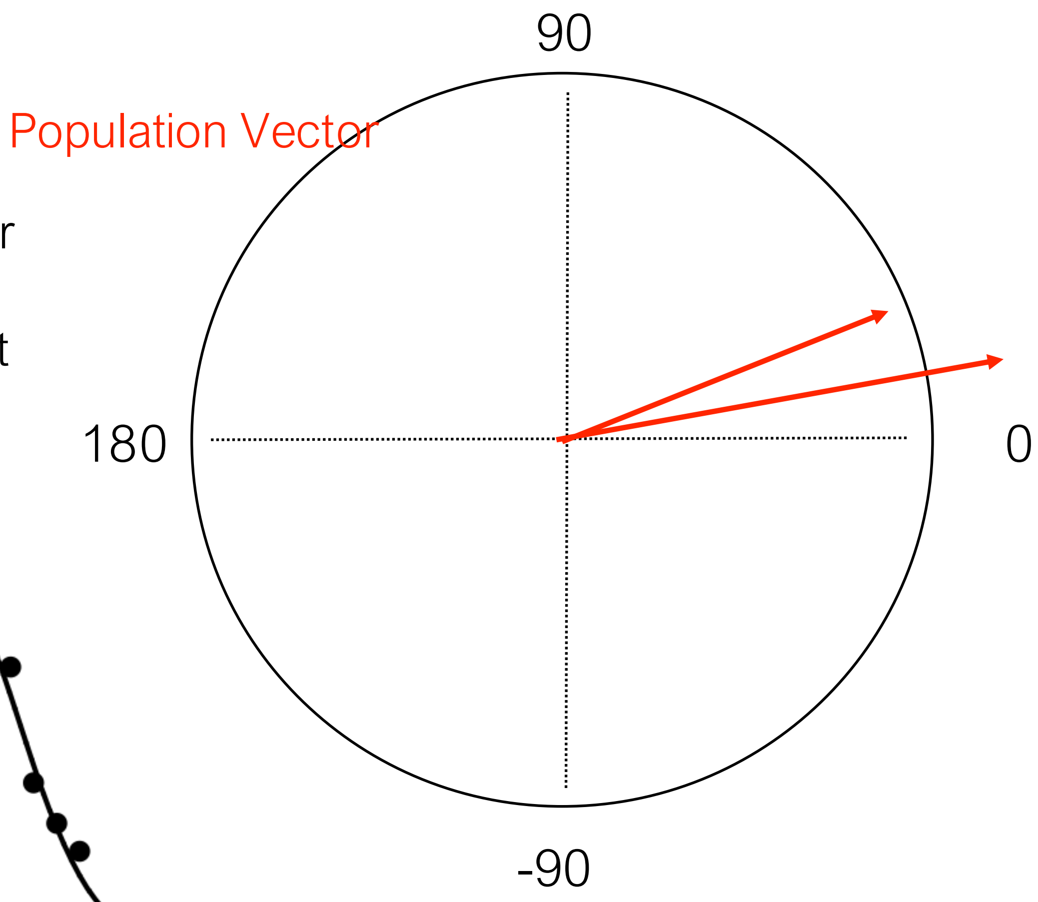




(Refresher) A Simple Decoding Strategy: Population Vector

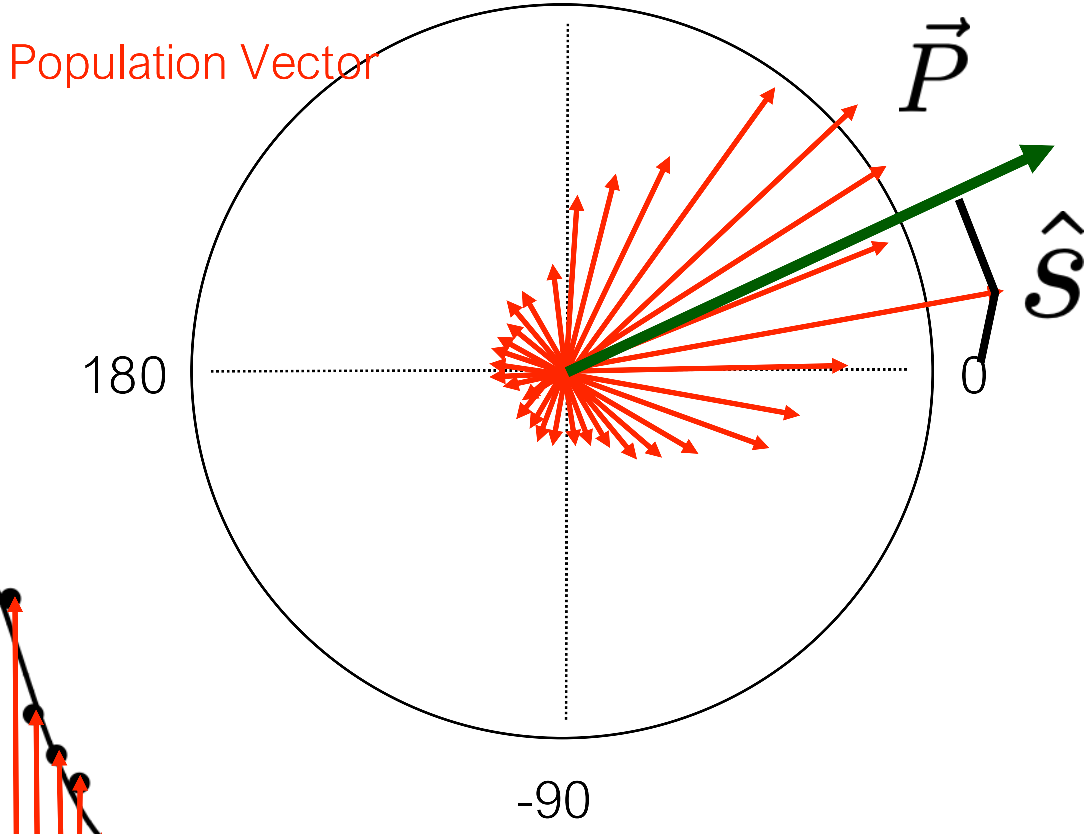
Each neuron "votes" for its preferred orientation with length given by its spike count

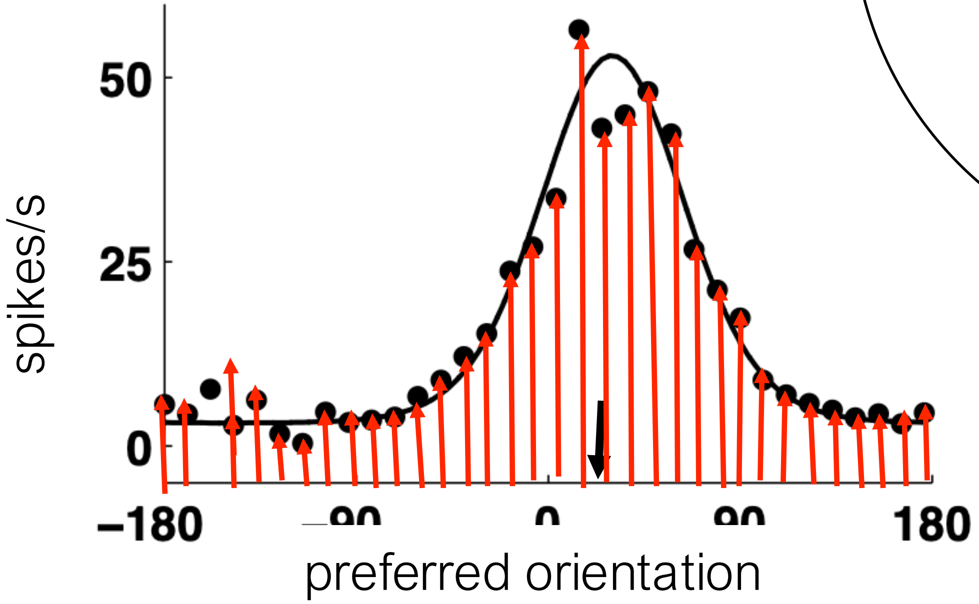




(Refresher) A Simple Decoding Strategy: Population Vector

The angle of the vector sum of all vectors is taken as the orientation estimate

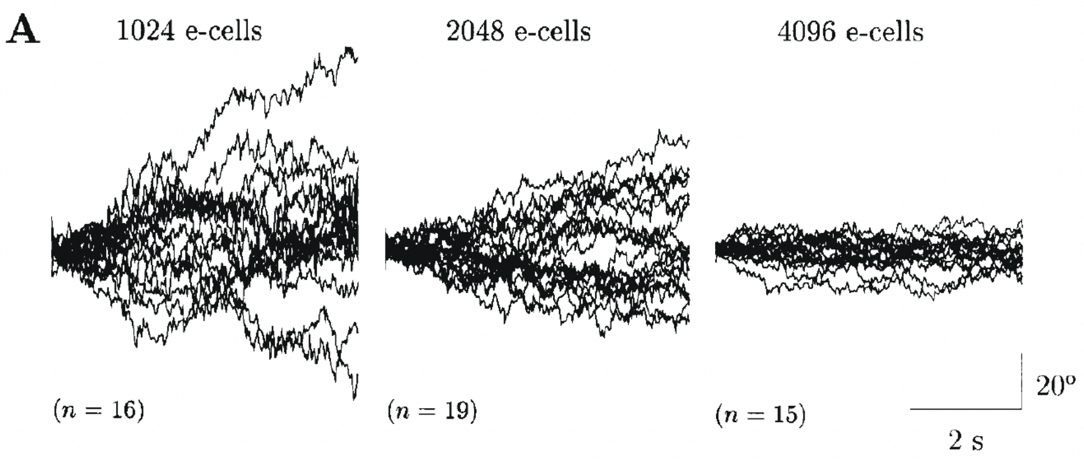




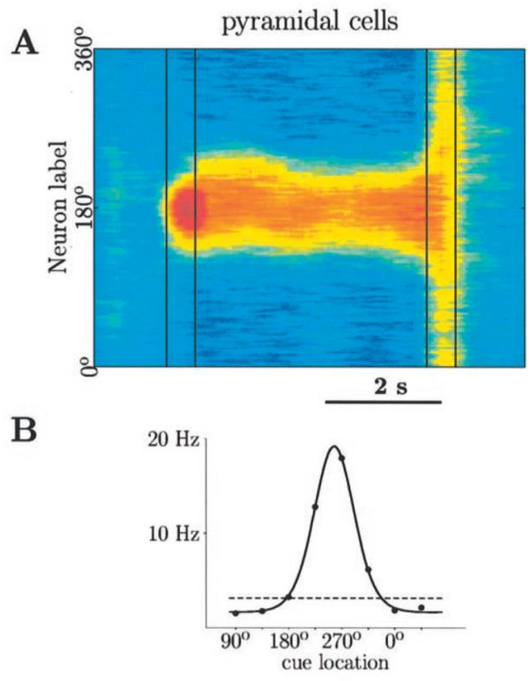
Can provide a good approximation to maximum likelihood decoding

Towards a Biophysical Model of Working Memory

- Decoding can be used to infer what the memory is encoding, e.g. population vector (decoding the "center" of the memory "bump").
- The ring model being a line attractor predicts emergence of drifts if noise is introduced, which would increase with delay time. Here, drift is found to be reduced for larger networks sizes.



Population vector position as a function of time for different runs and sizes of network

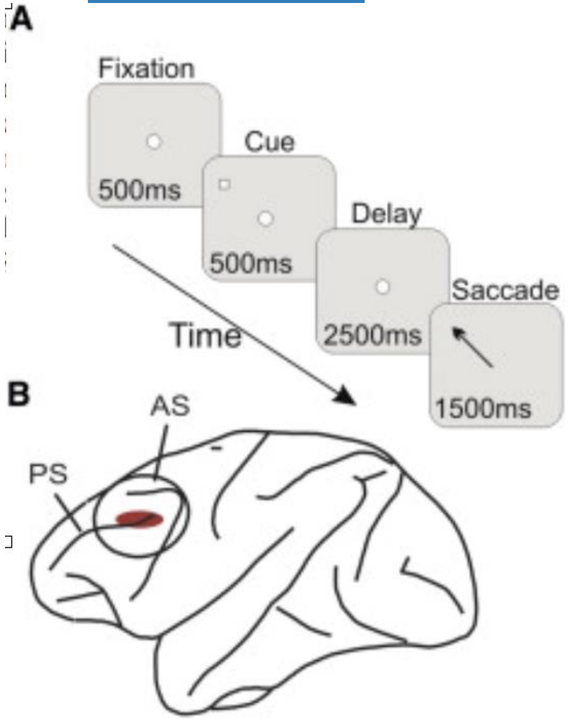


A prediction that has been verified

Neuron 2013

NMDA Receptors Subserve Persistent Neuronal Firing During Working Memory In Dorsolateral Prefrontal Cortex

Min Wang, Yang Yang, Ching-Jung Wang, Nao J. Gamo, Lu E. Jin, James A. Mazer, John H. Morrison, Xiao-Jing Wang, and **Amy F.T. Arnsten**
Dept. Neurobiology, Yale Medical School, New Haven, CT USA 06510



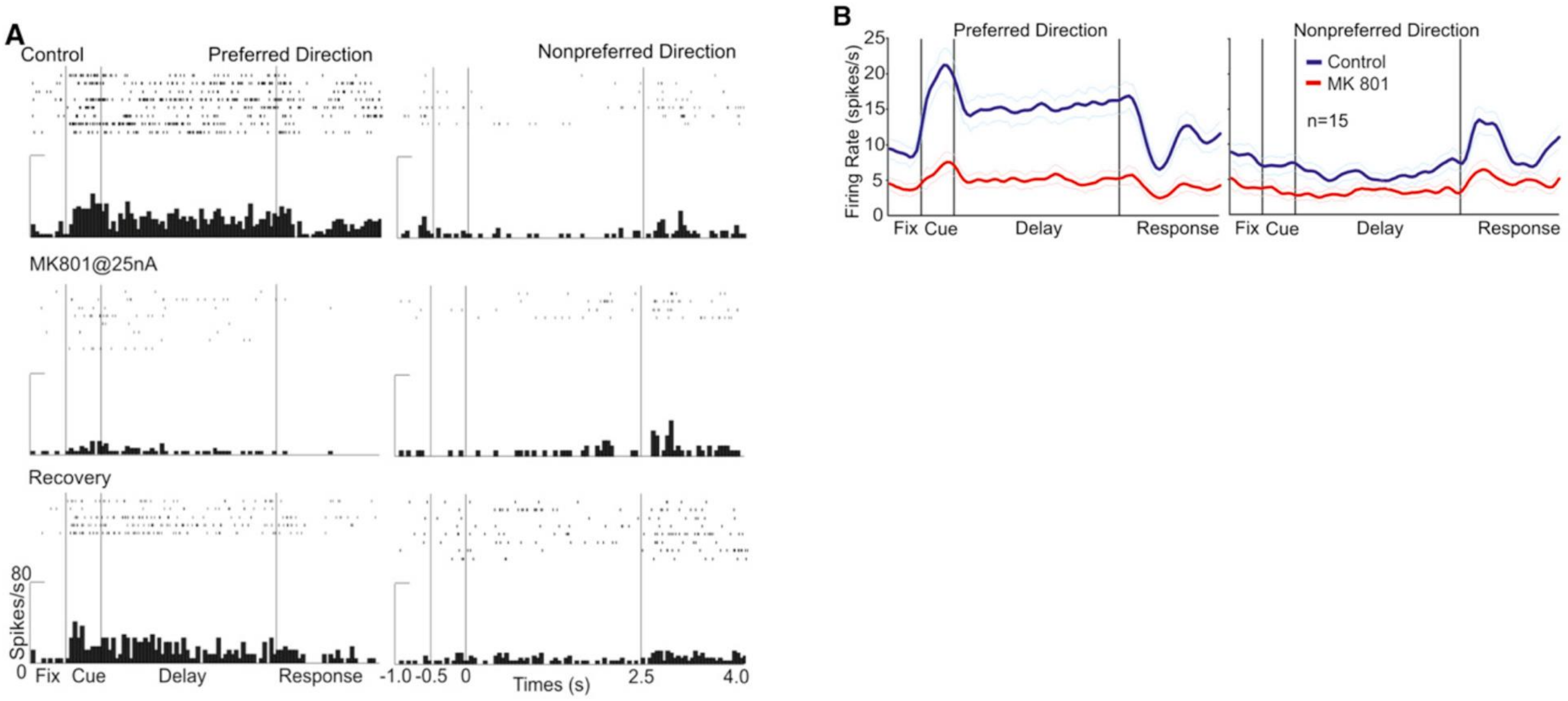
Summary

Neurons in the primate dorsolateral prefrontal cortex (dlPFC) generate persistent firing in the absence of sensory stimulation, the foundation of mental representation. Persistent firing arises from recurrent excitation within a network of pyramidal Delay cells. Here, we examined glutamate receptor influences underlying persistent firing in primate dlPFC during a spatial working memory task. Computational models predicted dependence on NMDA receptor (NMDAR) NR2B stimulation, and Delay cell persistent firing was abolished by local NR2B NMDAR blockade or by systemic ketamine administration. AMPA receptors (AMPAR) contributed background depolarization to sustain network firing. In contrast, many Response cells -which likely predominate in rodent PFC- were sensitive to AMPAR blockade and increased firing following systemic ketamine, indicating that models of ketamine actions should be refined to reflect neuronal heterogeneity. The reliance of Delay cells on NMDAR may explain why insults to NMDARs in schizophrenia or Alzheimer's Disease profoundly impair cognition.

- Delayed oculomotor task in 2 rhesus monkeys.
- Local Blockage of NMDA using iontophoresis.
- Systemic Ketamine administration (NMDA antagonist).
- Electrophysiological recordings of neurons in dlPFC.

A prediction that has been verified

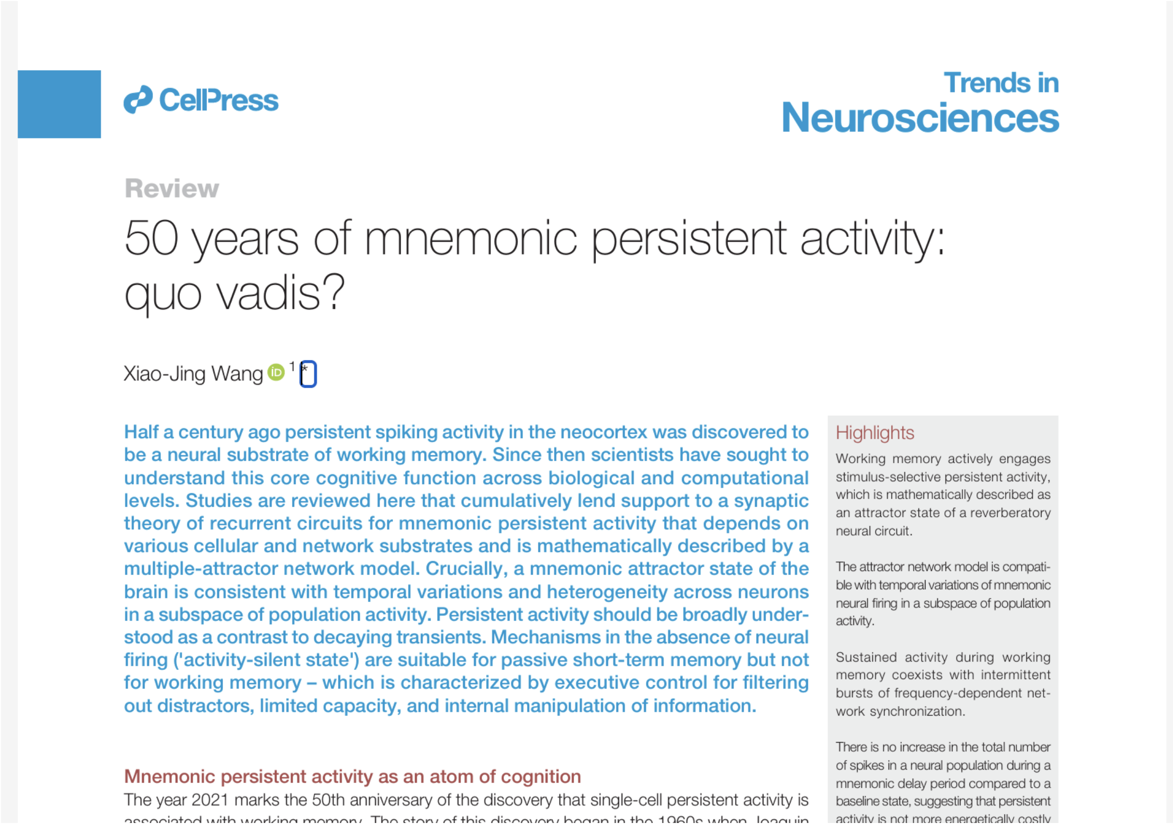
- Blocking NMDA, but not AMPA receptors, markedly reduced Delay cell firing
- Systemic ketamine also reduced Delay cell firing but increased Response cell firing
- Any cognitive operation relying on dlPFC recurrent firing would be compromised by insults to NMDAR transmission.



50 years of research

- Although considerable support has been gathered for the attractor / NMDA persistent activity model, more recent results nuance this model (dynamic patterns, additional role of short-term plasticity, dopamine modulation)

2021



CellPress

Trends in Neurosciences

Review

50 years of mnemonic persistent activity: quo vadis?

Xiao-Jing Wang[1,*]

Half a century ago persistent spiking activity in the neocortex was discovered to be a neural substrate of working memory. Since then scientists have sought to understand this core cognitive function across biological and computational levels. Studies are reviewed here that cumulatively lend support to a synaptic theory of recurrent circuits for mnemonic persistent activity that depends on various cellular and network substrates and is mathematically described by a multiple-attractor network model. Crucially, a mnemonic attractor state of the brain is consistent with temporal variations and heterogeneity across neurons in a subspace of population activity. Persistent activity should be broadly understood as a contrast to decaying transients. Mechanisms in the absence of neural firing ('activity-silent state') are suitable for passive short-term memory but not for working memory – which is characterized by executive control for filtering out distractors, limited capacity, and internal manipulation of information.

Highlights

Working memory actively engages stimulus-selective persistent activity, which is mathematically described as an attractor state of a reverberatory neural circuit.

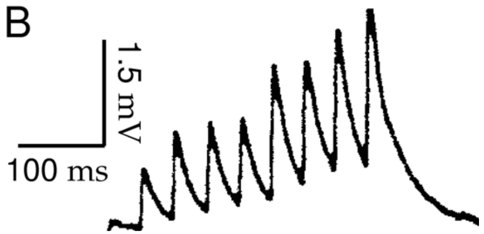
The attractor network model is compatible with temporal variations of mnemonic neural firing in a subspace of population activity.

Sustained activity during working memory coexists with intermittent bursts of frequency-dependent network synchronization.

There is no increase in the total number of spikes in a neural population during a mnemonic delay period compared to a baseline state, suggesting that persistent

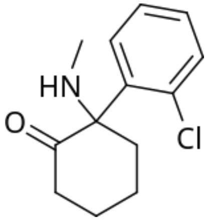
Mnemonic persistent activity as an atom of cognition

The year 2021 marks the 50th anniversary of the discovery that single-cell persistent activity is



- Short term synaptic facilitation contribute to passive maintenance of information over briefs intervals

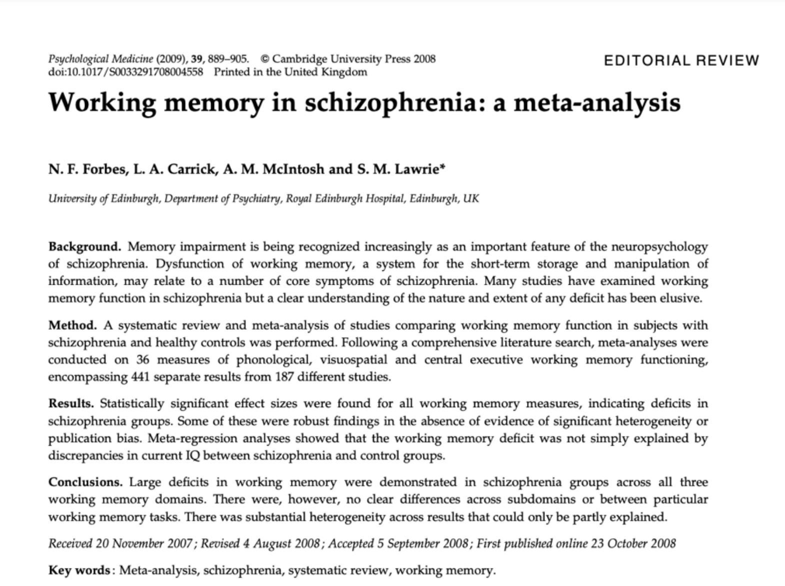
5) Application to Schizophrenia - NMDA hypothesis



- Schizophrenia associated with impairment of WM
- Reduced function of NMDA receptors in PFC

- Ketamine originally developed as an anaesthetic, blocks NMDA

- Ketamine can produce hallucinations and delusions - a model of psychosis used experimentally



Psychological Medicine (2009), 39, 889–905. © Cambridge University Press 2008
doi:10.1017/S0033291708004558 Printed in the United Kingdom

EDITORIAL REVIEW

Working memory in schizophrenia: a meta-analysis

N. F. Forbes, L. A. Carrick, A. M. McIntosh and S. M. Lawrie*

University of Edinburgh, Department of Psychiatry, Royal Edinburgh Hospital, Edinburgh, UK

Background. Memory impairment is being recognized increasingly as an important feature of the neuropsychology of schizophrenia. Dysfunction of working memory, a system for the short-term storage and manipulation of information, may relate to a number of core symptoms of schizophrenia. Many studies have examined working memory function in schizophrenia but a clear understanding of the nature and extent of any deficit has been elusive.

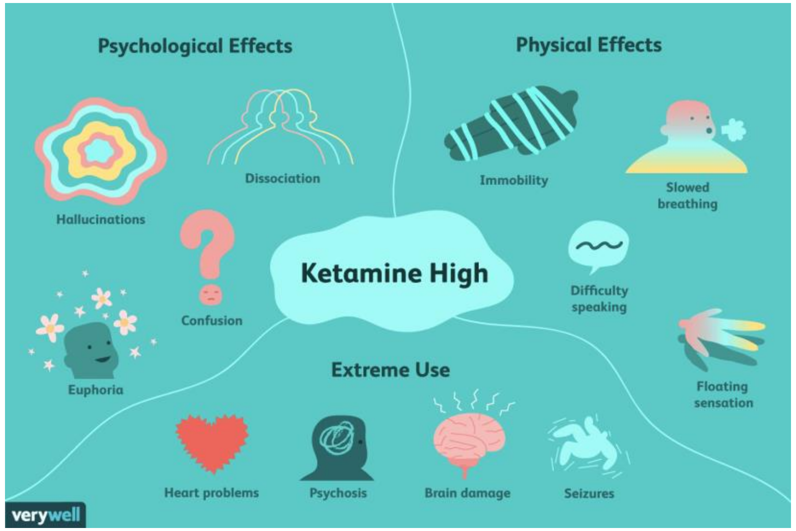
Method. A systematic review and meta-analysis of studies comparing working memory function in subjects with schizophrenia and healthy controls was performed. Following a comprehensive literature search, meta-analyses were conducted on 36 measures of phonological, visuospatial and central executive working memory functioning, encompassing 441 separate results from 187 different studies.

Results. Statistically significant effect sizes were found for all working memory measures, indicating deficits in schizophrenia groups. Some of these were robust findings in the absence of evidence of significant heterogeneity or publication bias. Meta-regression analyses showed that the working memory deficit was not simply explained by discrepancies in current IQ between schizophrenia and control groups.

Conclusions. Large deficits in working memory were demonstrated in schizophrenia groups across all three working memory domains. There were, however, no clear differences across subdomains or between particular working memory tasks. There was substantial heterogeneity across results that could only be partly explained.

Received 20 November 2007; Revised 4 August 2008; Accepted 5 September 2008; First published online 23 October 2008

Key words: Meta-analysis, schizophrenia, systematic review, working memory.



Application to Schizophrenia - NMDA hypothesis

- Schizophrenia associated with reduced function of NMDA receptor
- Instability of attractor states, shallower basins of attraction, reduced signal/noise
- spontaneous attractors

Rolls and Deco (2011) review how such impaired dynamics could explain positive symptoms (hallucinations, delusions), cognitive symptoms (working memory) and even negative symptoms (through reduced activity), and onset (excessive synaptic pruning, and reduction in grey matter volume).

Review

A computational neuroscience approach to schizophrenia and its onset

Edmund T. Rolls[a,*], Gustavo Deco[b,c]

[a] Oxford Centre for Computational Neuroscience, Oxford, UK
[b] Computational Neuroscience, Universitat Pompeu Fabra, Barcelona, Spain
[c] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

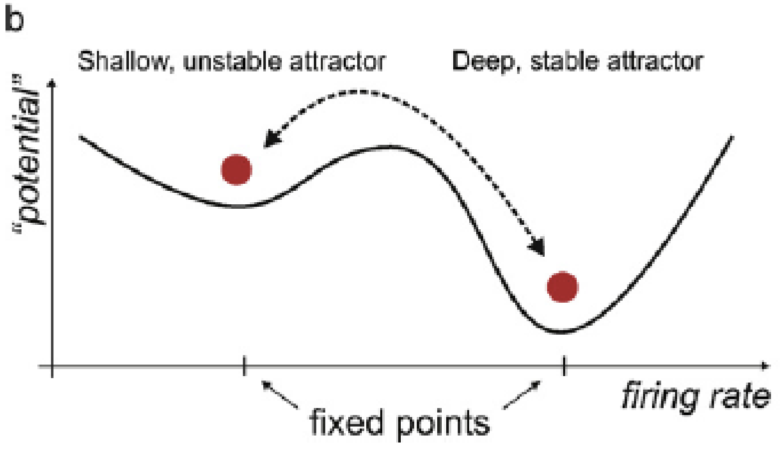
A R T I C L E I N F O

Article history:
Received 8 April 2010
Received in revised form 12 August 2010
Accepted 2 September 2010

Keywords:
Schizophrenia

A B S T R A C T

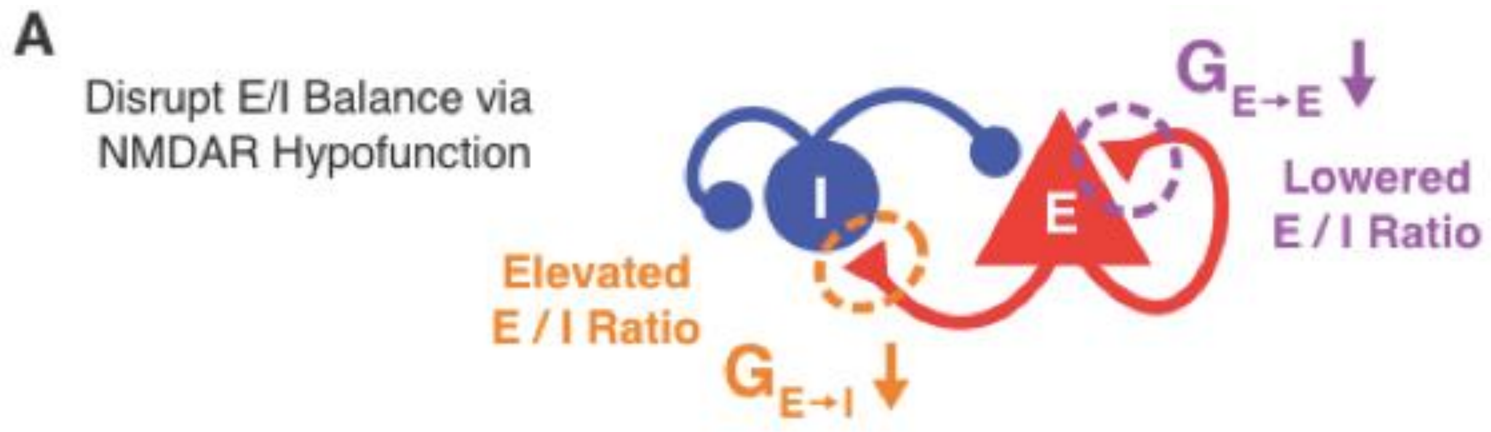
Computational neuroscience integ
factors that alter the stability of c
spiking times. A reduction of the
receptor function (present in schiz
that normally implement short-te
symptoms of schizophrenia. Redu
sion (present in schizophrenia) car



Application to Schizophrenia - NMDA hypothesis

NMDA impairment can be applied to connections on

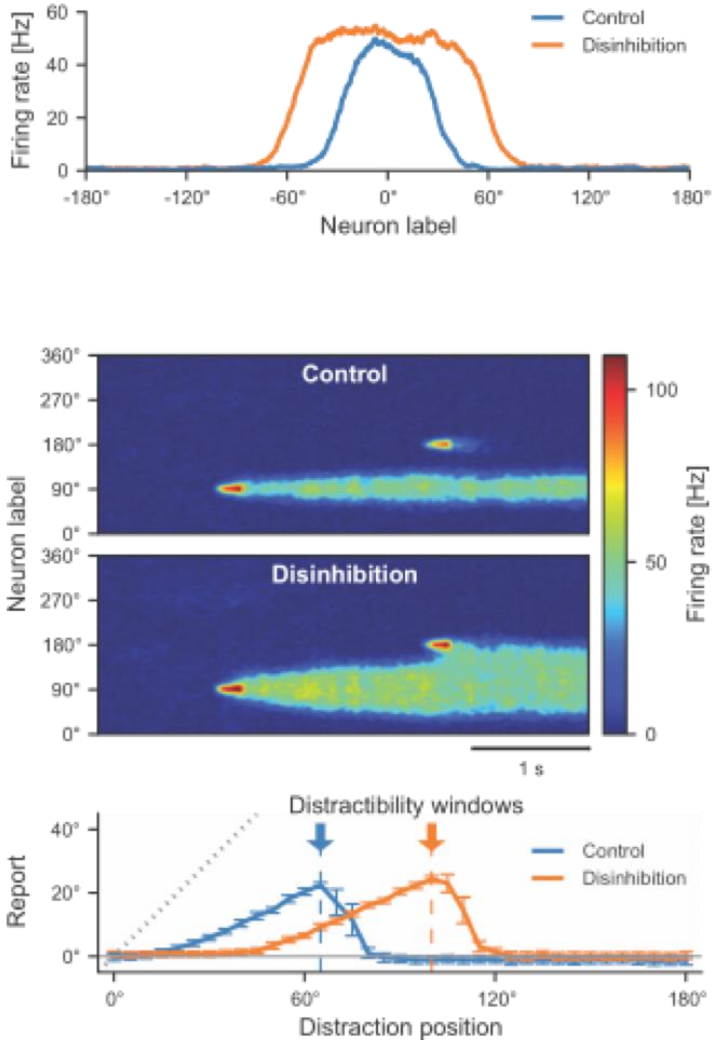
- inhibitory interneurons, which elevates E/I ratio via disinhibition;
- excitatory neurons, which on the contrary lowers E/I ratio



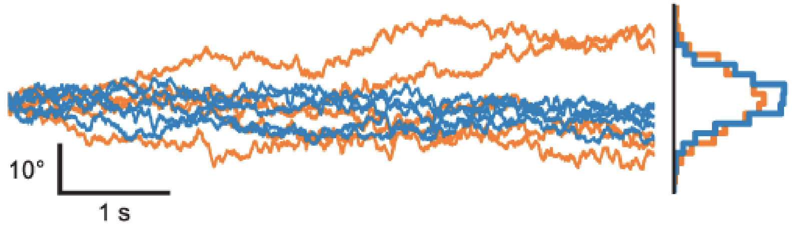
Which synapses matter most for the maintenance of working memory ?

Linking Microcircuit Dysfunction to Cognitive Impairment: Effects of Disinhibition Associated with Schizophrenia in a Cortical Working Memory Model

John D. Murray[1,2], Alan Anticevic[3,4,5], Mark Gancsos[3], Megan Ichinose[3], Philip R. Corlett[3,5], John H. Krystal[3,4,5,6,7] and Xiao-Jing Wang[2,8]



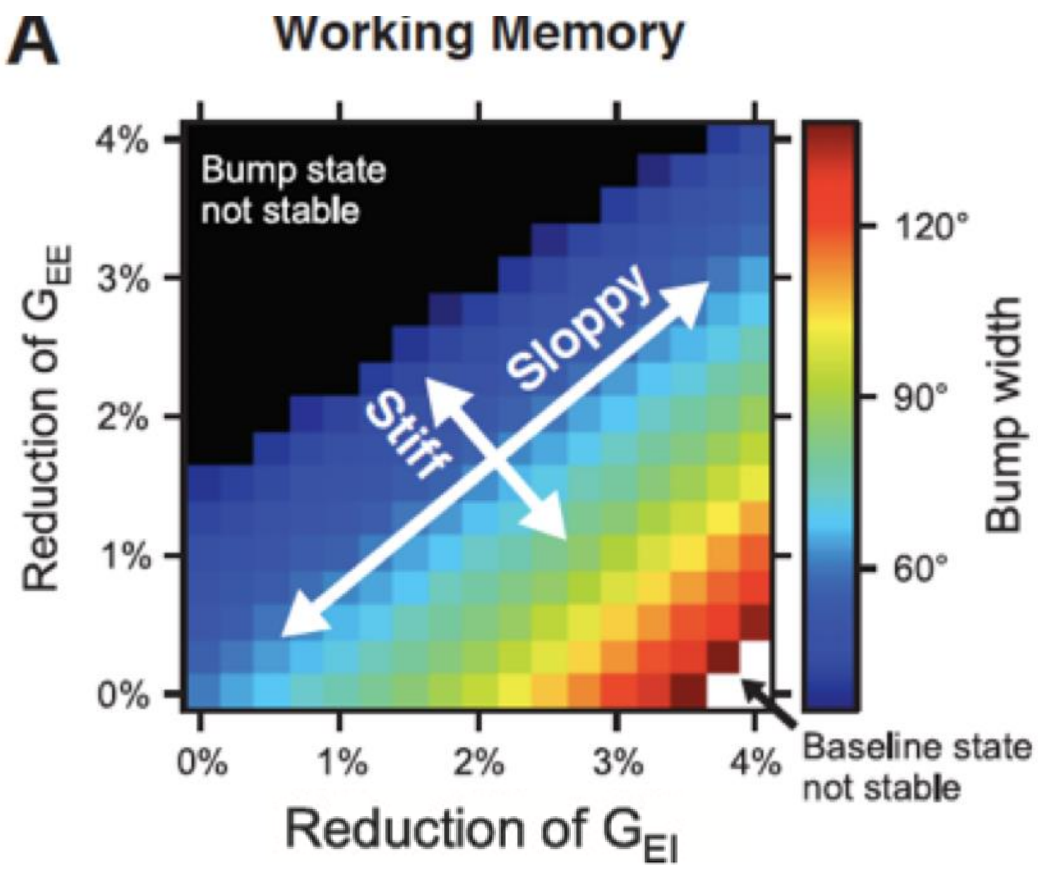
- **Desinhibition** via perturbation of NMDA receptors on Inhibitory cells.
- Broadens selectivity,
- Increases drift
- Increases vulnerability to distractors



[Murray et al 2012]

State diagram for the role of E/I balance in cognitive function

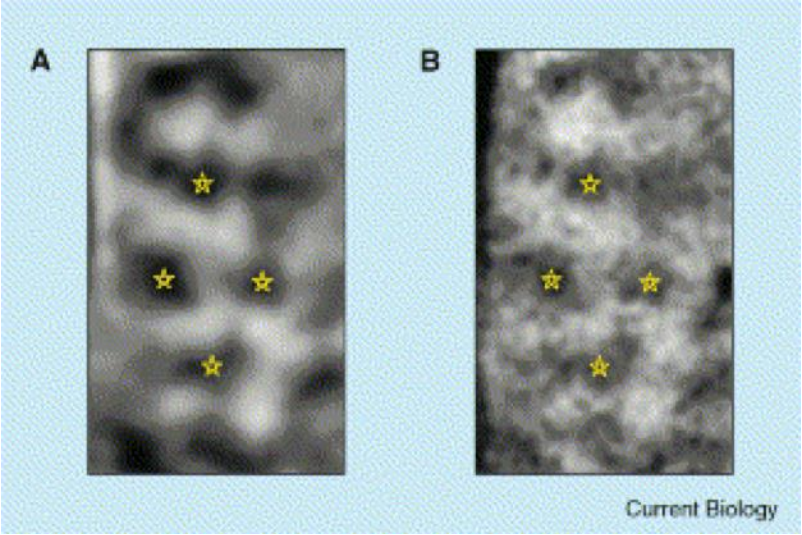
- Along some axes in parameters space, the model is relatively insensitive to perturbations (“sloppy”).
- E/I ratio is the key parameter for optimal function.



more generally, importance of E/I ratio in models of mental illness

A related problem: spontaneous activity

- Where does it come from?
- How is it maintained? How does it 'move'?
- Are these 'attractor states'?
- Is it structured?
- Why is it there? (any functional advantages?)
- Is it noise?
- Is it the brain trying to 'predict' the input?



evoked (horizontal orientation)   spontaneous (one frame)

Arieli et al 1997; Tsodyks et al, 1999; Fiser et al, Nature, 2004

Conclusions

- Attractor Networks as (main) model of working memory / sustained activity

- Effort to provide biologically plausible spiking models, comparable to recordings in Prefrontal cortex.

- Excitatory reverberation and maintenance of sustained activity is found to depend on NMDA receptors

- currently, interesting link with disease as well as ageing
-- working memory impairments as instability of attractor states e.g. due to deficits in NMDA, changes in E/I balance.

- Spontaneous activity as a similar problem.

2022

nature reviews neuroscience

https://doi.org/10.1038/s41583-022-00642-0

Review article

Check for updates

Attractor and integrator networks in the brain

Mikail Khona[1,2,3,4] & Ila R. Fiete[1,2,3]

Abstract

In this Review, we describe the singular success of attractor neural network models in describing how the brain maintains persistent activity states for working memory, corrects errors and integrates noisy cues. We consider the mechanisms by which simple and forgetful units can organize to collectively generate dynamics on the long timescales required for such computations. We discuss the myriad potential uses of attractor dynamics for computation in the brain, and showcase notable examples of brain systems in which inherently low-dimensional continuous-attractor dynamics have been concretely and rigorously identified. Thus, it is now possible to conclusively state that the brain constructs and uses such systems for computation. Finally, we highlight recent theoretical advances in understanding how the fundamental trade-offs between robustness and capacity and between structure and flexibility can be overcome by reusing and recombining the same set of modular attractors for multiple functions, so they together produce representations that are structurally constrained and robust but exhibit high capacity and are flexible.

Sections