Dynamical Mechanisms for Coordinating Long-term Working Memory Based on the Precision of Spike-timing in Cortical Neurons

Terrence J. Sejnowski[1,2]

1 Computational Neurobiology Laboratory, Salk Institute for Biological Sciences, La Jolla, 92037 CA, USA

2 Department of Neurobiology, University of California, San Diego, La Jolla, 92093 CA, USA

Abstract

In the last century, most sensorimotor studies of cortical neurons relied on average firing rates. Rate coding is efficient for fast sensorimotor processing that occurs within a few seconds. Much less is known about long-term working memory with a time scale of hours (Ericsson and Kintsch, 1995). The discovery of millisecond-precision spike initiation in cortical neurons was unexpected (Mainen and Sejnowski, 1995). Even more striking was the precision of spiking *in vivo*, in response to rapidly fluctuating sensory inputs, suggesting that neural circuits could preserve and manipulate sensory information through spike timing. High temporal resolution enables a broader range of neural codes. It could also support spike-timing-dependent plasticity (STDP), which is triggered by the relative timing of spikes between presynaptic and postsynaptic neurons in the millisecond range. What spike-timing mechanisms could regulate STDP *in vivo*? Cortical traveling waves have been observed across many frequency bands with high temporal precision. Traveling waves have wave fronts that could link spike timing to STDP. As a wave front passes through a cortical column, excitatory synapses on the dendrites of both pyramidal and basket cells are stimulated synchronously. Inhibitory basket cells form a calyx on pyramidal cell bodies, and inhibitory rebound following a strong transient hyperpolarization can trigger a backpropagating action potential, which arrives shortly after the excitatory inputs on pyramidal dendrites. STDP activated in this way could persist for hours, creating a second-tier network. This temporary network could support long-term working memory, a cognitive network riding above the long-term sensorimotor network. On their own, traveling waves and STDP have not yet yielded new insights into cortical function. Together, they could be responsible for how we think (Sejnowski, 2025).

Introduction

We are at a crossroads in systems neuroscience. In the 20th century, progress was made in characterizing the response properties of single neurons using sharp microelectrodes during behavioral tasks. The cortical responses of sensory neurons correlated with sensory inputs, those of motor neurons with movements, and those of neurons in association areas with higher cognitive functions. The tasks typically lasted a few seconds and required extensive training. However, in the wild, behavior is self-generated and is coordinated over much longer time spans. Recordings from freely moving rodents, for example, revealed that hippocampal neurons responded selectively to places in the environment, which would have been missed in head-fixed experiments. Self-generated cognition that occurs without any sensory inputs or motor outputs over minutes and hours, as in remembering and planning, is much more difficult to study than tasks that require working memory over seconds.

During the 1980s, I focused on network models of vision (Ballard, Hinton, and Sejnowski, 1983), grounded in rich psychophysical research and neural recordings from the visual cortex. I was inspired by pioneering network models, such as the Marr and Poggio (1976) model of stereopsis. Vision research in that era focused on images and object recognition. I knew that the visual system integrated information across eye movements and was curious how it was done. For example, as you read this article, your eyes make fast, saccadic movements across the page, taking in small groups of words in your fovea three times per second. Each saccade is a snapshot that must be integrated with previous words to build a conceptual understanding of what is being conveyed.

Psychologists call this long-term working memory (Ericsson and Kintsch, 1995). After reading this article, your brain will think about it in the context of experiences and thoughts previously stored in long-term memory. After listening and watching a lecture for an hour, you can retain enough details that you heard and the slides you saw to ask a relevant question. During a concert, recurring themes are expected and variations detected. Long-term working memory has a time scale of hours, much longer than short-term working memory, which lasts seconds and minutes, such as remembering a phone number by rehearsal.

We can now record distributed neural activity from many thousands of neurons simultaneously throughout the cortex. We need a new conceptual framework for how cognition arises from global activity on long time scales. Self-generated long-term working memory could be supported by the dynamical neural mechanisms proposed here.

The precision of spike timing in cortical neurons

While transitioning from theoretical physics to neuroscience, I was fortunate to work as a postdoctoral fellow with Stephen Kuffler in the Department of Neurobiology at Harvard Medical School. I gained skills in intracellular recordings from neurons and synaptic physiology, which introduced me to a pioneering tradition in electrophysiology going back to Alan Hodgkin, Andrew Huxley, and Bernard Katz in the 1950s. My first job was in the Thomas C. Jenkins Biophysics Department at Johns Hopkins University in the early 1980s, where I set up an experimental lab and started simulating Hodgkin-Huxley models of neurons using the NEURON computer program.

When I moved from Johns Hopkins to San Diego in 1989, I established the Computational Neurobiology Laboratory at the Salk Institute and the Institute for Neural Computation at the University of California at San Diego. My lab attracted some of the best and brightest graduate students and postdoctoral fellows. Zachary Mainen was a new graduate student in my lab from Brown University, coming from Tom Brown's lab at Yale. We had daily afternoon teas where new ideas and advances were discussed. Tony Zador, who was in Chuck Stevens' lab next door and previously Zach's graduate supervisor at Yale, would often join the discussions. Neural coding was a hot topic, and one day we had a lively debate over spikes. Tony Bell, a postdoctoral fellow, believed every spike mattered. This went against the seemingly random spike times in cortical recordings *in vivo*, consistent with the consensus view that cortical processing was noisy and a rate code carried most of the information in cortical networks. If spike times matter, what are they coding and how are they decoded?

Spike timing is determined by the highly fluctuating membrane potentials observed in intracellular recordings *in vivo*, which are driven by barrages of excitatory and inhibitory synaptic inputs from ongoing cortical activity. How would a cortical neuron respond if a controlled, randomly fluctuating current were to be injected into the cell body? One of the advantages of having a wet lab is that you do not have to wait for someone else to do your

experiment. It was not long before Mainen had the answer (Mainen and Sejnowski, 1995). When he repeatedly injected the same fluctuating current waveform, called frozen noise, the precision of action potential timing was in the millisecond range (Fig. 1). This revealed that the spike initiation mechanism in pyramidal neurons was surprisingly robust and raised an interesting question: Why?

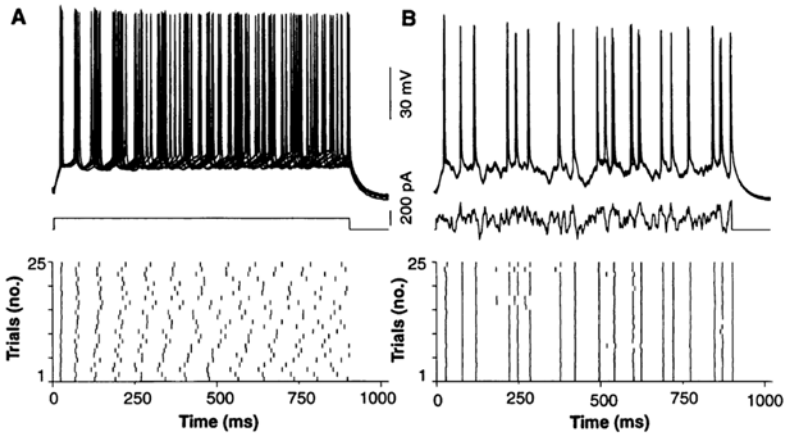

Figure 1. Precise spike timing in response to frozen noise. **A**) Constant current injection into a cortical pyramidal neuron results in a drifting spike raster. **B**) Precisely timed spike responses to repeated injection of the same fluctuating current (Adapted from Mainen and Sejnowski, 1995).

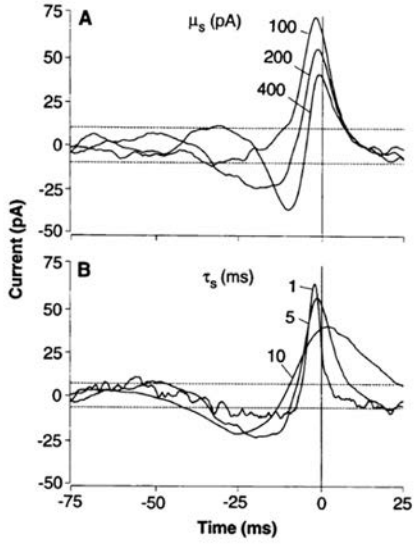

Figure 2. Spike-triggered average reveals a postinhibitory rebound mechanism underlying the precision of spike timing in Fig. 1. The mean and time constant of the filtered noise was varied (Adapted from Mainen and Sejnowski, 1995).

Our precision paper has been cited over 2,500 times. The citation rate on most papers falls exponentially, but 30 years later, this paper continues to receive around 100 citations per year. The difference between constant-current injection and frozen noise in Fig. 1 is what most remember from the paper. We also asked what feature in the current waveform triggered the spike. We expected a fast depolarizing fluctuation, but spike-triggered averaging revealed a consistent hyperpolarizing dip preceding the spike (Fig. 2). A strong inhibitory input controls spike timing by post-inhibitory rebound. In Hodgkin-Huxley models, sodium channels are partially inactivated near threshold, and hyperpolarization deinactivates them, making a neuron more excitable and ensuring precise spike timing.

There are many types of inhibitory neurons in the cortex. Which could provide strong inhibitory input to control spike timing? The soma of a pyramidal neuron receives a calyx of feedback inhibition from parvalbumin-positive (PV) basket cells. Synaptic inputs from basket cells are functionally equivalent to injecting current into the soma. Basket cells also receive strong excitatory inputs from pyramidal neurons. This feedback circuit is also responsible for precisely timed bursts of spikes during 40-80 Hz gamma oscillations (Jadi and Sejnowski, 2014).

In vivo spike timing precision

Current injected directly into the soma of neurons produces precise spike timing in slices. *In vivo* conditions may be different because current is injected throughout the dendritic tree. Thalamic neurons preserve the millisecond timing of a flickering visual stimulus (Reinagel and Reid, 2002) (Fig. 3). Is sensory timing information preserved upstream? A good place to look is area MT, which receives direct input from area V1 and is several synapses deeper in the visual cortex. First identified by John Allman and John Kaas in the medial temporal area of the owl monkey's visual cortex (Allman and Kaas, 1971), MT neurons selectively respond to the direction of motion of visual stimuli. Several researchers have focused on neurons in MT, including Tom Albright, Tony Movshon, and Bill Newsome.

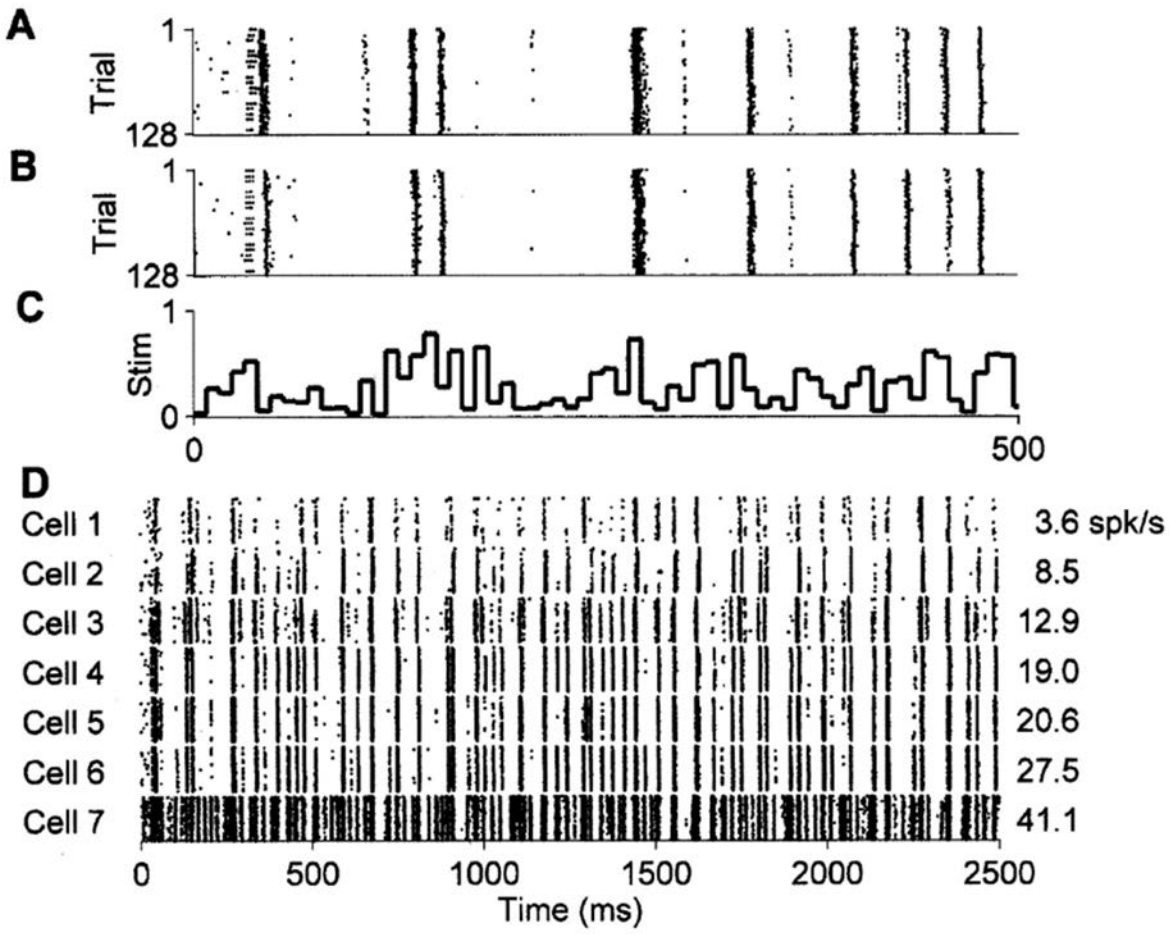

Figure 3. Response of thalamic relay neurons in the cat lateral geniculate nucleus to a fluctuating visual stimulus. **A, B**) Spike rasters. **C**) Time course of the intensity of the visual stimulus. **D**) Spike-time rasters for seven thalamic relay neurons with averge firing rates varying over an order of magnitude. (Adapted from Reinagel and Reid, 2002).

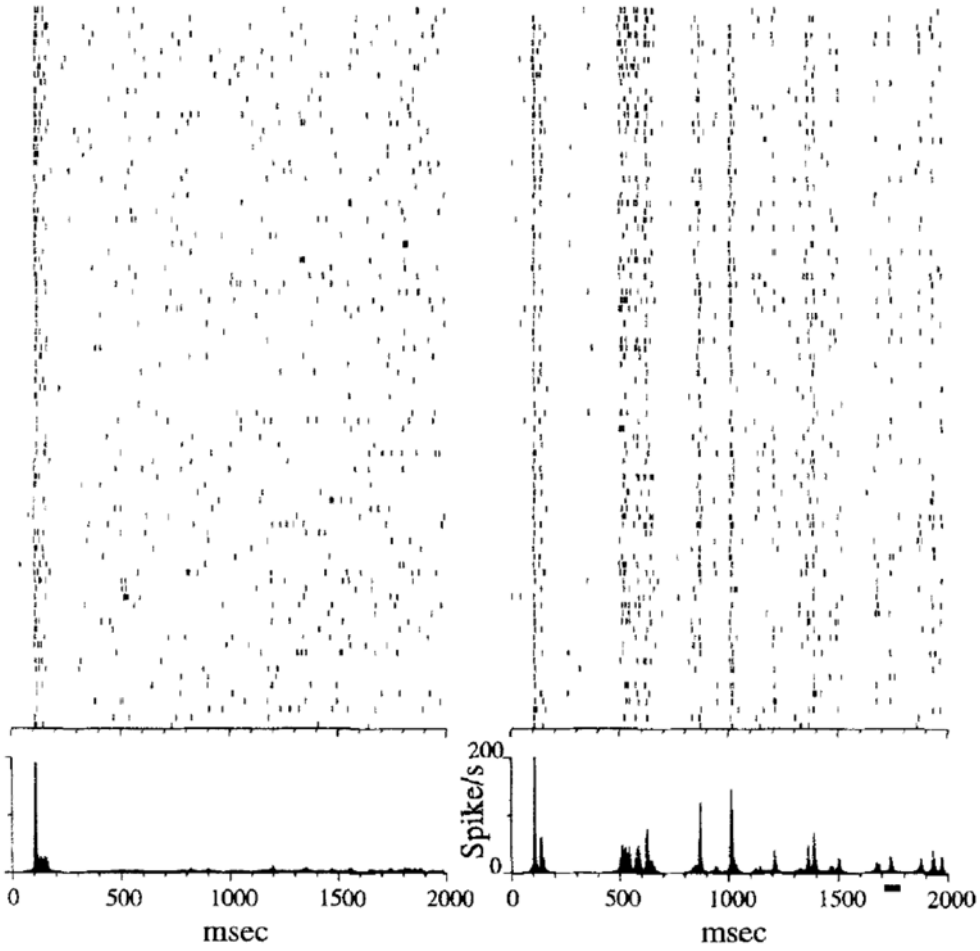

Figure 4. Responses of neurons in Monkey area MT in response to drafting random dots. **Left**: Display program initiated with a random seed on each trial. **Right**: Display program initiated with the same random seed on each trial (Adapted from Bair and Koch, 1996).

Newsome trained monkeys to detect the direction of random dots on a screen, with a fraction of them briefly moving in one direction (Bair and Koch, 1996). He varied the fraction and found the detection threshold and the rising psychometric function. Wyeth Bair and Christof Koch reanalyzed Newsome's recordings. They found that when trials with the same input motion correlation were displayed in a raster format, spike peaks appeared with 3 ms timing precision, similar to those found *in vitro* (Fig. 4, right). The random-dot display program used by the Newsome lab, obtained from Movshon, did not reset the random seed, so the same random dots appeared at the same locations and times in every trial for trials with the same motion

correlation. This was a fortuitous oversight since otherwise the precision of spike timing would not have been apparent (Fig. 4 left).

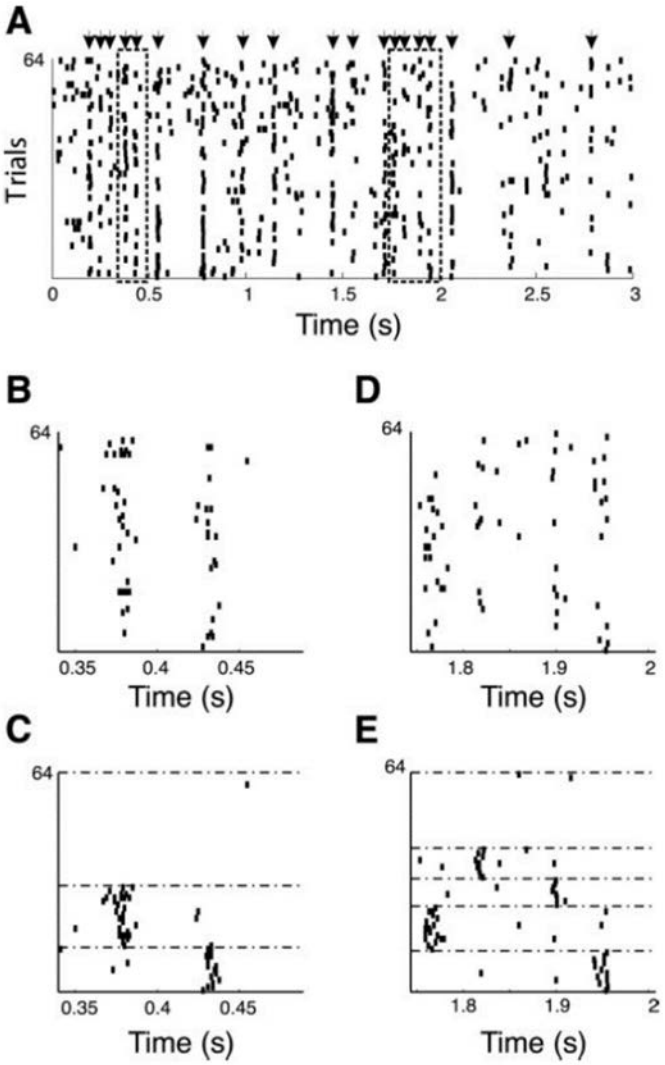

Figure 5. Multiple spike patterns in response to the same fluctuating visual stimulus in monkey area MT. **A**) spike responses to from 64 trials to the same fluctuating visual stimulus. **B**, **D**) Blow-up of two segments of the raster in panel A. **C,E**) Clustering of the trails in panels B and D, revealed two and four spike time patterns, respectively. (Adapted from Fellous, Tiesinga, Thomas, Sejnowski, 2004).

Shortly thereafter, researchers in Albright's lab confirmed that the spike timing of neurons in monkey area MT in response to fluctuations in moving visual stimuli was preserved with 3-

millisecond precision (Buracas, Zador, DeWeese, Albright, 1998). In a reanalysis of these data, clustering algorithms revealed that the same fluctuating stimulus produced not a single precisely timed spike sequence, but several precise spike patterns (Fellous, Tiesinga, Thomas, Sejnowski, 2004) (Fig. 5). A few trials in Fig. 1 show offsets, suggesting something similar may happen in cortical circuits, leading to multiple, interleaved dynamical spike patterns to the same fluctuating stimulus. The discovery that the cortex could respond nearly deterministically to sensory inputs with precisely timed spikes led to research into how diverse patterns of precisely timed spikes emerged from Hodgkin-Huxley models (Tiesinga, Fellous, Sejnowski, 2008).

Multiple cortical neural codes

The concept of a receptive field was an organizing principle for interpreting the properties of single neurons in the 20th century (Sejnowski, 1976a). The receptive field of a cortical neuron is the specific region of sensory space in which an appropriate stimulus elicits a change in its firing rate. Because of variation in the number and timing of spikes, the response is typically averaged over trials. Evidently, these neurons use firing rate to code sensory information. What was left for the cortex to process from sensory inputs that needed millisecond precision? Could it be that both types of coding are present and used for different purposes (Singer, 2021)?

David McCormick recorded intracellular membrane potentials from neurons in cat visual cortex in response to repeated video clips (Haider et al., 2010). They masked the visual stimulus within a neuron's receptive field and compared the spike trains to responses when the mask was widened (Fig. 6). When the stimulus was restricted to the classical receptive field, the neuron responded vigorously, with spike timing patterns that varied trial-to-trial. However, when the mask was widened, the neuron hyperpolarized, lowering the firing rate and revealing precisely timed spikes. Large inhibitory inputs appeared in the nonclassical stimulus condition (Fig. 6, lower panel). Compare these inhibitory events with the hyperpolarizing events preceding spikes in Fig. 2): The hyperpolarization was responsible for the precisely timed spikes. These results show that, in addition to modulating the average firing rate, inputs from the nonclassical receptive field shift neural circuits toward precise spike timing.

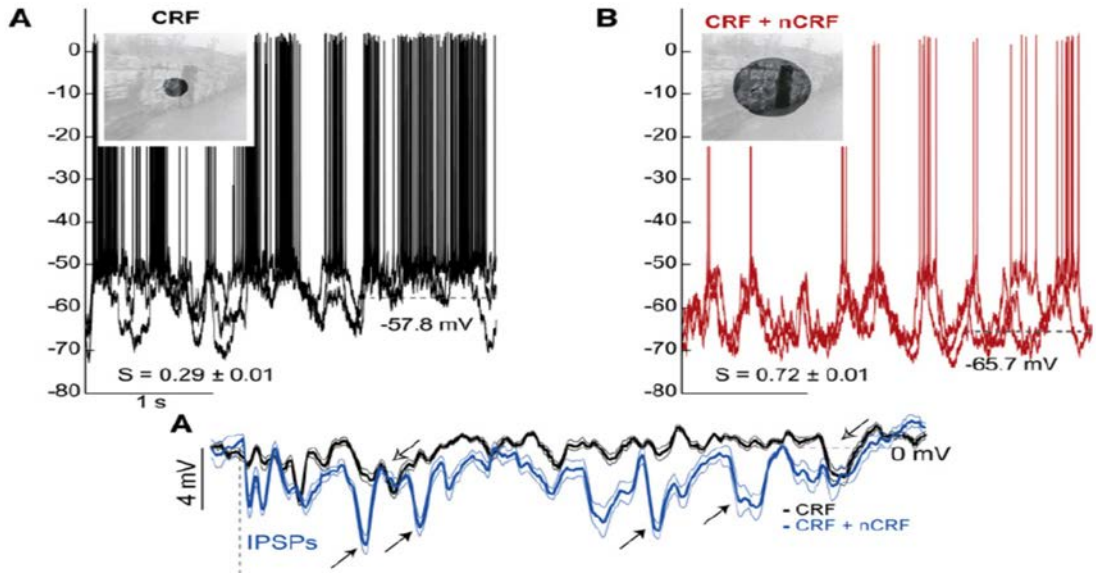

Figure 6. Intracellular responses of a cortical pyramidal neuron to a repeated video clip: **A**) when restricted to the classical receptive field (CRF) and **B**) when widened into the nonclassical receptive field (nCRF). **Below**: Intracellular membrane potential in response to the CRF (black) and CRF+nCRF (blue). The modulatory influence of visual inputs from outside of the classical receptive field was first observed in area MT by John Allman and has subsequently been shown to occur throughout the cortex (Allman, Miezin, McGuinness, 1985). (Adapted from Haider et al. 2010).

These results have interesting implications. First, under normal conditions, the visual cortex responds to full-field visual inputs and is in the nonclassical state. Conclusions about neural codes derived solely from recordings restricted to the classical receptive field should be reconsidered; second, there is a mixture of single spike responses and more rate-like responses in the nonclassical state. This blend suggests that both are present and may convey different kinds of information. Rate coding and spike timing codes could coexist within the same population of neurons, serving different purposes at different time scales; A third, less obvious implication is that inputs from the nonclassical surround, in addition to mixing spatial information, also mix temporal information. The long-range horizontal axons in the cortex that link neurons in cortical columns travel many millimeters. The axons are thin, unmyelinated, and their spikes travel slowly at 0.1 m/sec, arriving with time delays of 10-100 ms (Johnson and Frostig, 2016).

A new spacetime code based on temporal context

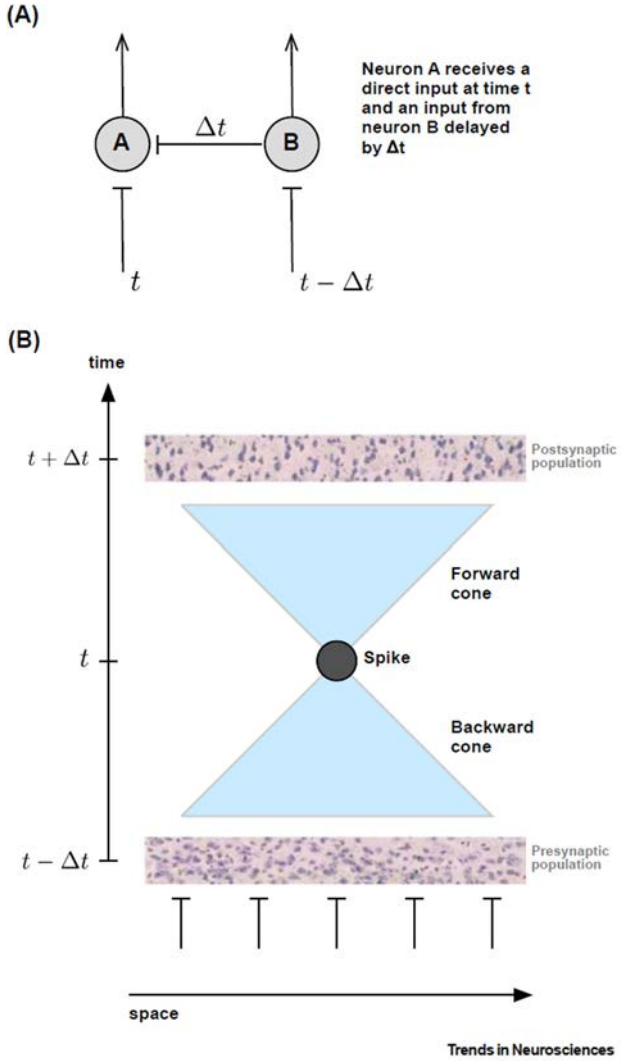

Figure 7. Spacetime population coding. (**A**) Reichardt model for detecting the direction of motion of a visual stimulus. The visual stimulus moves from location A to location B during a time delay Δ*t*. Neuron A responds only when it receives a direct visual input at time *t* and an indirect visual input with a time delay Δ*t* from neuron B. (**B**) The simple circuit in panel A can be expanded to include a wide range of converging inputs from neighboring neurons with a range of time delays. The timing of the spike at time *t* mixes information within the lower blue spacetime cone and, in turn, influences neurons in the upper blue cone. (Adapted from Muller, Churchland, and Sejnowski, 2024).

Nonclassical receptive fields provide both spatial context and temporal context. For example, a moving stimulus can be detected by sequential inputs in neighboring locations. This is the kernel of the Reichardt detector and Adelson-Bergen motion energy model (Fig. 7A) (Reichardt, 1957; Adelson and Bergen, 1985).

Cortical neurons receive delayed signals from many nearby neurons; the farther away, the longer the delay (Fig. 7B). Thus, a population of neurons could represent not just what is happening at a single moment, but also over a window of space and time (Muller, Churchland, and Sejnowski, 2024).

Spike-timing codes are based on neural dynamics on timescales appropriate for temporal context. If so, then it should be possible to reconstruct videos. This could be accomplished by training a deep learning network on spike-time recordings from large-scale neural populations to predict past movie frames.

Traveling waves mix temporal coding with spatial coding

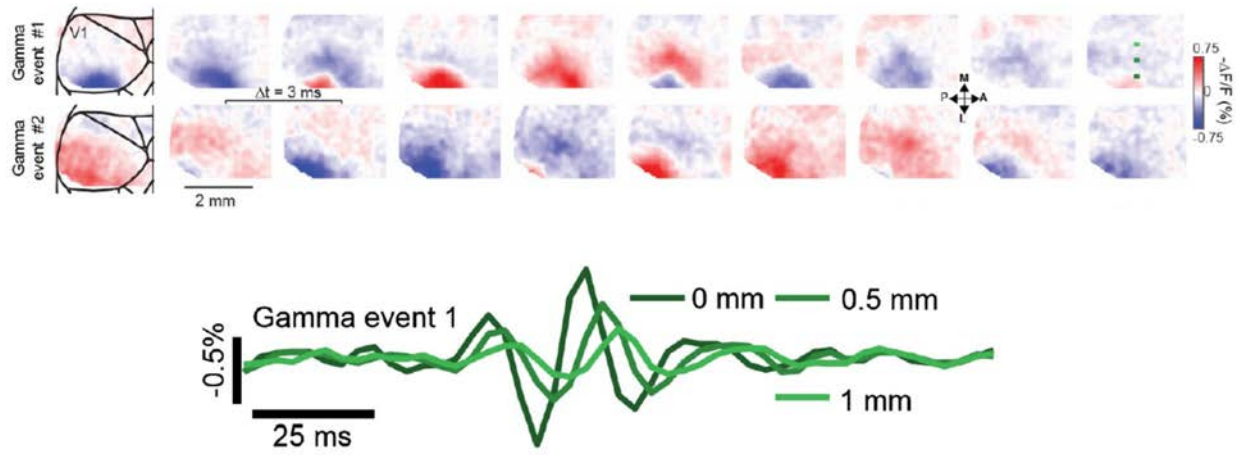

Figure 8. Membrane potentials of neurons in mouse visual cortex measured with a voltage-sensitive dye. Top: two sequences of 3-ms frames illustrating waves traveling from the bottom to the top of the frame. **Bottom**: Time course of voltage measurements at three locations in the top image showing the waveform and phase differences. Traveling waves are wave packets. (Adapted from Haziza et al., 2024).

Cortical oscillations are ubiquitous in recordings of local field potentials, which average extracellular signals from neighboring neurons. These oscillations have generally been

interpreted as evidence for synchronous spiking. Recordings from electrode arrays and optical recordings have instead revealed that local oscillations in all frequency bands extend spatially as traveling waves (Muller, Chavane, Reynolds, Sejnowski, 2018). Fig. 8 shows an optical recording of a traveling wave in the gamma band in mouse visual cortex. The dynamical patterns of spikes in traveling waves observed in the cortex allow information in relative spike timing to be propagated across the cortex by long-range horizontal axons (Gilbert and Wiesel, 1989). Spiking is sparse in cortical traveling waves, allowing many different subsets of interconnected neurons to represent many different spatiotemporal features. The circuit mechanisms for the origin of sparse mesoscale waves in the cortex have been identified and mathematically modeled (Davis et al., 2021).

Traveling waves are dynamical and allow the stable processing of signals undergoing motion and convey more information than synchronous oscillations (Keller, Muller, Sejnowski, Welling, 2025b). Because the same process is repeated at every layer in cortical hierarchies, the time window should increase from 100 ms in V1 to 10 seconds in the inferotemporal cortex (Hasson et al., 2008). The decrease in white matter in the prefrontal cortex also contributes to widening the spacetime window (Wang, 2020). Spacetime representations are counterintuitive, but recall that a three-dimensional object floating in space can be reconstructed from the speckles in a two-dimensional hologram that our visual system cannot interpret (Fig. 9). The object can be reconstructed from a portion of the hologram at reduced resolution (Pribram, 2013). In a holographic spacetime representation, one of the dimensions is time. There is no reason for evolution to make neural encoding easy for us to understand. We face the same challenge interpreting activity patterns in large language models.

Fast but temporary synaptic plasticity

Absence epilepsy, common in children, lasts for 10-20 seconds, during which there is a staring spell and a pause in motor behaviors (Destexhe and Sejnowski, 2023). In the end, the conversation and thinking continue as before the pause. Remarkably, absence epilepsy is characterized by massive, low-frequency thalamic bursting that sweeps across the cortex. What maintains traces of previous electrical activity? The continuity of thoughts could be sustained by fast synaptic weight changes lasting minutes to hours, preserved across the electrical storm. These temporary weight changes could serve as the substrate for long-term working memory.

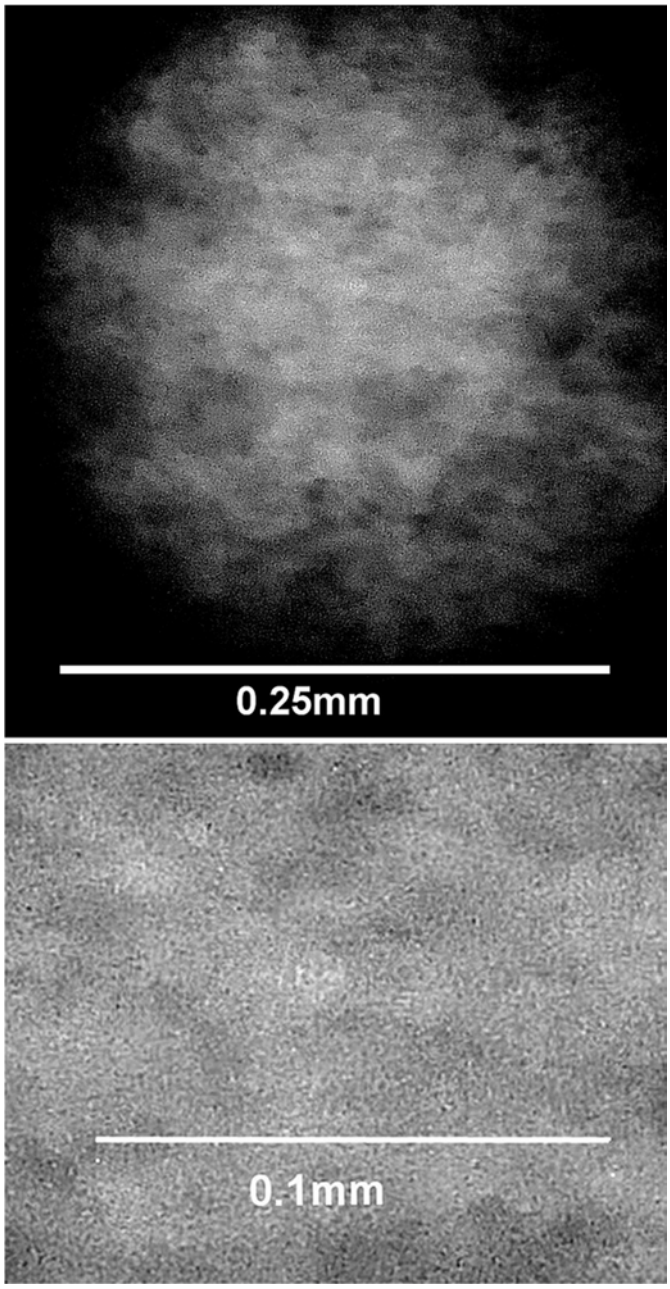

Figure 9. This is a small part of a transmission hologram photographic plate viewed through a microscope at two levels of magnification. The hologram captured images of a toy van and a car. It is no more possible to discern the subject of the hologram from this pattern than it is to identify what music has been recorded by looking at the surface of a DVD. Holographic information is encrypted in the speckle pattern. The three-dimensional objects can be recovered by shining a laser onto the hologram. (https://en.wikipedia.org/wiki/Holography).

A candidate for rapid weight change is spike-timing-dependent plasticity (STDP) (Markram, 1997; Feldman, 2012) (Fig. 10). STDP occurs when an excitatory input occurs within 10 milliseconds of a backward-propagating spike in the postsynaptic dendrite. The strength of the synapse is potentiated when presynaptic input precedes the postsynaptic spike (pre before post) and is depressed when the postsynaptic spike precedes the presynaptic input (post before pre). STDP has been demonstrated in slice preparations but remains elusive *in vivo*, partly because the mechanisms that regulate cortical spike timing *in vivo* are unknown.

Could traveling waves trigger STDP? As a traveling wave passes a pyramidal neuron in a neighboring column, excitatory synapses are activated on both the pyramidal neuron and neighboring PV interneurons. Strong transient activation of PV neurons could trigger inhibitory rebound and a backward-propagating spike in the pyramidal dendrites, arriving a few milliseconds after the feedforward excitatory inputs (Fig. 6). This dual pathway to the pyramidal neurons could coordinate the elusive 10-millisecond time window for pairing presynaptic and postsynaptic spikes during STDP.

Traveling waves are typically packets of spikes at 10-40 Hz, which also match the bursts used to elicit STDP *in vitro* (Markram et al., 1997). The simultaneous activation of local basket cells by the wave front could also trigger the ING mechanism for synchronizing reciprocally connected inhibitory neurons, thereby amplifying the transient hyperpolarizing pulse in the somas of nearby pyramidal neurons (Fig. 10) (Van Vreeswijk, Abbott, Ermentrout, 1994; Tiesinga and Sejnowski, 2009). Not all the basket cells, however, may synchronize. In a model of short-term working memory, activity in one group of basket cells could switch the network's output by inhibiting another group (Kim and Sejnowski, 2021). A chimera of locally synchronized basket cells could control both activity and plasticity.

Only a small fraction of cortical neurons spike as a traveling wave passes. In contrast, the traveling wave during a seizure is dense and is followed by postictal depression, during which the cortex is silent for 5-30 seconds. However, if even 1% of cortical neurons in a cortical column of 100,000 neurons spike during a sparse traveling wave, only 1,000 would be activated. In a cortical column, around 10% of pyramidal neurons are reciprocally connected, so around 100 neurons could undergo STDP together, strengthening their mutual connectivity and forming

a recurrent microcircuit, which Hebb called an assembly (Yoshimura, Dantzker and Callaway, 2005; Perin, Berger, and Markram, 2011).

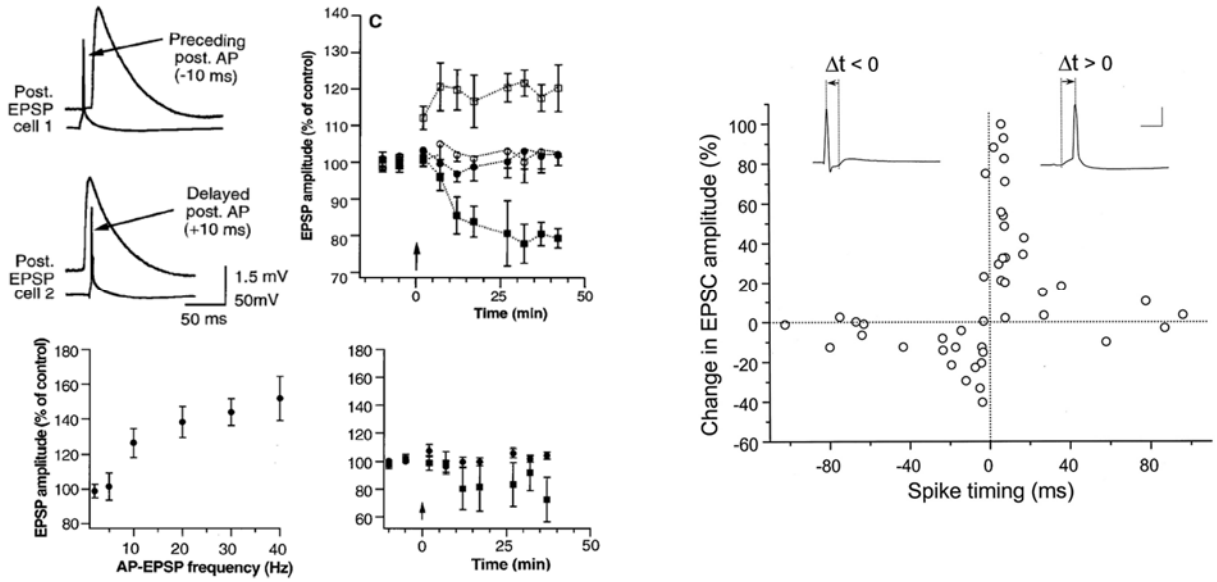

Figure 10. Spike-timing dependent plasticity. **Left**: Pairing presynaptic inputs with postsynaptic spikes with a 10 ms delay in rodent cortical slices produces fast weight increases (pre before post) and decreases (post before pre) at synapses. Bursts of 4 pairings were repeated 10 times, spaced by 1 second, at frequencies ranging from 4 Hz to 40 Hz. (Adapted from Markram et al., 1997). **Right**: The interval between the presynaptic input and the postsynaptic spike pairing was varied in cultured hippocampal neurons. (Adapted from Bi and Po, 1998)

Protocols that induce STDP produced plasticity that lasted for an hour *in vitro*, which were called long-term potentiation and long-term depression. However, another possibility is that STDP *in vivo* induces temporary plasticity for only a few hours, the timescale of long-term working memory. If so, then which cortical synapses are most likely to support STDP? Excitatory synapses on small spines are weaker than those with large spines, and entering calcium increases the calcium concentration in their small volume faster, making them more susceptible to STDP. They are also more labile and turn over with learning (Yang et al., 2014; Berry and Nedivi, 2017). Because the distribution of spine head sizes in the cortex is log-normal, the small, labile spines constitute the majority of excitatory synapses on pyramidal neurons (Lowenstein et al., 2022). Thus, the majority of cortical excitatory synapses could be recruited for cognitive processing over a rolling hours-long time window. The number of small synapses decreases

with age, paralleling the lower capacity of long-term working memory in older adults (Holtmaat et al., 2005).

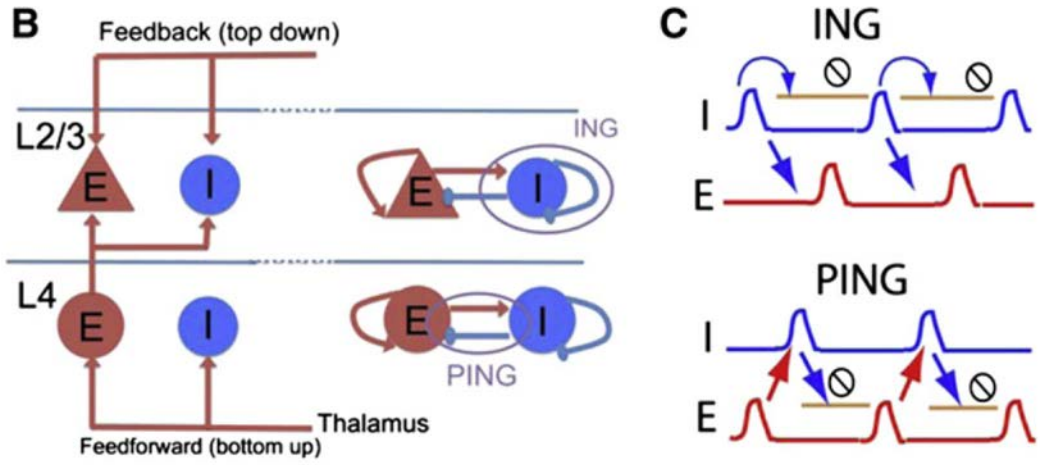

Figure 11. Neural mechanisms for synchronizing spiking in populations of neurons **B**) (Left) Simplified representation of the cortical laminar structure. (Right) Synchrony can be achieved by feedback between excitatory and inhibitory populations (PING) or reciprocal interactions between inhibitory populations (ING). The gray loops in the ING and PING insets encircle the essential interactions that lead to synchrony. Note that the inhibitory neurons are reciprocally connected. **C**) For the ING mechanism, I cells are sufficiently excited to spike synchronously without excitatory network activity. The light-brown lines with a stop sign indicate the period during which the network is inactive due to the high inhibitory conductance. (Adapted from Tiesinga and Sejnowski, 2009).

Origins of Traveling Waves

Global traveling waves occur during sleep in humans (Muller et al., 2016). Sleep spindles are global 10-14 Hz circular traveling waves lasting 1-2 seconds that coordinate the selective consolidation of experiences into long-term memories (Destexhe and Sejnowski, 2023). Experiences during the day are encoded in the hippocampus and replayed at night, triggering sleep spindles during slow-wave sleep. Hippocampal replay to the cortex originates spontaneously as 100 Hz sharp wave-ripples in CA3 at six time normal speed (Buzsáki, 2019). Cortical traveling waves are triggered by area CA3 in the hippocampus and spontaneously within highly recurrent networks in layer 2/3 of the neocortex. The architecture of these highly recurrent networks enables efficient learning of input sequences and their computational processing during replay (Keller, Muller, Sejnowski, Welling, 2024a; Chen, Zhang, H.,

Cameron, Sejnowski, 2024). Replay from the hippocampus also occurs during awake resting states and during repetitive activity such as running and gardening.

Traveling waves spontaneously originating in the cortex may be recalling old memories and planning future actions. Sensory inputs and cognitive tasks can trigger traveling waves. Circular traveling waves occur in the monkey prefrontal cortex during specific phases of working memory tasks (Bhattacharya et al., 2022; Batabyal et al., 2025) and in the human cortex (Koller, Schirner, and Ritter, 2024). Spontaneously circular traveling waves in rodent somatosensory cortex are coordinated with subcortical spiking patterns in the thalamus, striatum, and midbrain (Ye et al., 2025). Given their ubiquity and apparent involvement in so many aspects of cortical function, these traveling waves deserve our attention.

Phase Precession

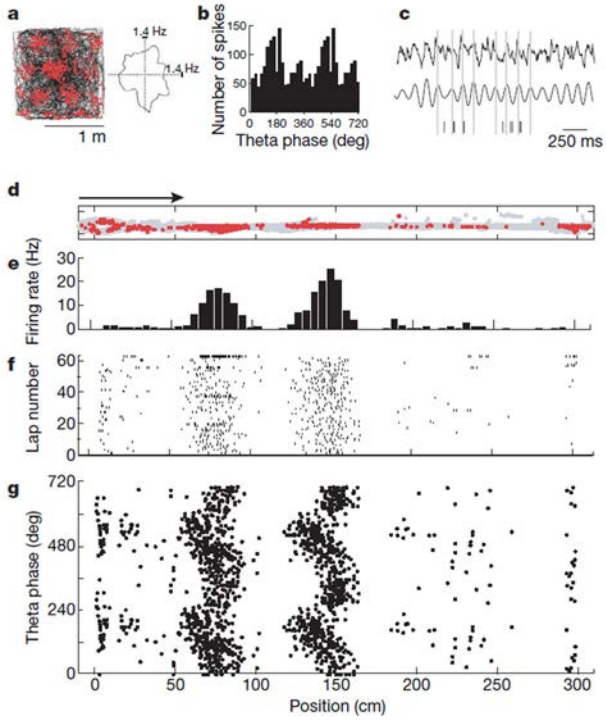

Figure 12. Phase precession in a layer 2 pyramidal neuron in the entorhinal cortex. **a)** Left: firing field of a grid cell during running in an open field (black, trajectory; red, individual spikes). Right: firing rate as a function of running direction (bin size 6°). **b)** Distribution of firing rate within the theta cycle for the cell during running on a linear track (bins of 20°, two theta cycles, peak of local theta rhythm is 0°). **c)** Local entorhinal EEG with spike times for the layer 2 cell in a and b during 2.1 sec of track running. Vertical ticks are individual spikes. Top: unfiltered EEG trace. Bottom: filtered at 6–11 Hz. Vertical grey lines indicate 0°. **d–g**: Rate distribution and phase relationship for spikes of the cell in a–c during running from left to right. **d)** Trajectory (gray) with locations of individual spikes (red). The arrow indicates the running direction. **e)** Linearized spatial firing rate map (bins of 5 cm). **f**, Raster plot indicating spike positions (vertical ticks) on the track. **g)** Theta phase as a function of position (two theta cycles). In d and e, all spikes are shown except at the turning points. Note the gradual advance of firing phase as the rat passes through each field. (Adapted from Hafting et al., 2008)

Little is known about the relative timing of spikes in a traveling wave as it passes through the cortex. Do all the spikes in a wave arrive synchronously, or are they offset? Only a few

milliseconds separate potentiation from depression in STDP (Fig. 10). How is this knife-edge achieved *in vivo*?

Traveling waves in the theta band (4-8 Hz) occur in the hippocampus and entorhinal cortex of rodents as they explore the environment (Lubenov and Siapas, 2009; Patel et al., 2012; Hafting et al., 2008). Neurons in the hippocampus have spatial place fields that release bursts of spikes at progressive phases of the theta wave as the rodent traverses the place field, a phenomenon called phase precession. Phase precession also occurs in the grid cells of the entorhinal cortex (Fig. 12). As the wave moves across the cortex, membrane potentials in a local population of neurons are sequentially swept across threshold in order of their initial depolarization levels. Phase precession has also been observed in recordings from the hippocampus, entorhinal cortex, parahippocampal gyrus, anterior cingulate cortex, orbitofrontal cortex, and amygdala of neurosurgical patients undergoing clinical treatment for drug-resistant epilepsy (Qasim, Fried, Jacobs, 2021).

When the time-ordered spikes arrive downstream, the earliest will activate basket cells, eliciting a backpropagating action potential that potentiates the early presynaptic inputs on dendrites (pre before post) but depresses subsequent synaptic inputs (post before pre). Repetitive pairing in the same order will ensure that a sparse subset of synapses is strengthened; however, because of phase precession, earlier spikes drop out on subsequent cycles, allowing synaptic inputs from later spikes to become potentiated. One advantage of this scheme is that it maintains a balance between potentiated and depressed synapses. The computational capabilities of time-ordered sequences of spikes, called polychrony, have been explored by Izhikevich (2006, 2025).

Decoding spike-time codes

Neural Coding in the PFC. The PFC is involved in high-level decision-making and planning, encoding decision variables and time intervals using a combination of rate and precise temporal codes. The specific timing of spikes can carry more information and greater stability for complex stimuli and behaviors than rate codes alone. The PFC sends excitatory glutamatergic projections to the striatum of the BG.

Timing-to-Rate Conversion. Striatal neurons exhibit rich, variable population dynamics that encode timing information and temporal context for subsequent actions (Jin, Fujii, Graybiel, 2009). The convergence of precisely timed spikes from many PFC neurons onto many fewer striatal neurons summates their excitatory postsynaptic inputs, effectively converting the temporal pattern into a stronger, suprathreshold signal that influences firing rates in downstream targets. The precise timing of PFC spikes, after processing through the BG circuit, potentially involving complex feedback loops and post-inhibitory rebound in the thalamus, influences the average firing rates of thalamic neurons.

Thalamocortical Neurons. The conversion of a decision represented by spike timing in the second tier of the prefrontal cortex (PFC) into a firing-rate code in the motor cortex is a complex process that involves several subcortical structures. The cortex projects topographically to the ganglia (BG) and loops back to the cortex through the thalamus. This key circuit transforms temporal dynamics into motor commands. The BG effectively acts as a filter and integrator, translating timing-based signals into amplified firing-rate "decision" and "command" signals that the motor system reliably uses to execute a specific sequence of movements.

Thalamic relay neurons convey rate-encoded information to the motor cortex and premotor cortices. This feedforward input to the cortex is integrated with cortical feedback from layer 6 (Berry, 2025). The motor cortex primarily uses a rate code to represent motor commands, such as movement direction, force, and timing. The incoming rate-encoded signal from the thalamus directly modulates the firing rate of motor cortex neurons, which then initiate and refine the persistent movement or action.

Temporal context. Transformers in large language models are feedforward deep learning networks trained to predict the next word in a sentence. The temporal context of a word in a sentence helps us understand its meaning. In "Bob, who is hopeful, likes Alice," "Bob" is positively associated with "hopeful," while "Alice" is negatively associated. These associations are represented in transformers as “self attention" (Fig. 13). "Attention" here refers to the strength of association between words, not to the way the term “attention” is used in psychology and neuroscience. Topographically organized loops between the cortex and the BG, shown in Fig. 13, could provide temporal context for cortical representations.

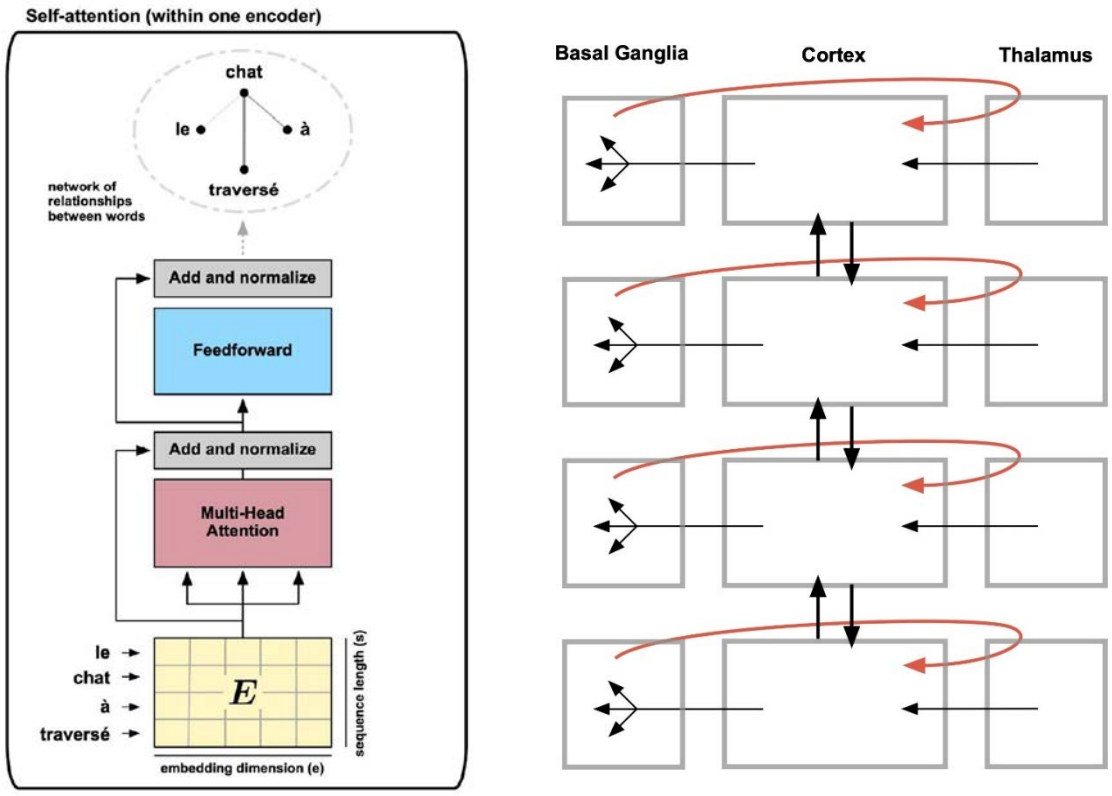

Figure 13. **Left**: One module in a feedforward transformer with input on the bottom and output from the top. **Right**: Hierarchy of cortical areas receiving inputs from the thalamus and a feedback loop from the BG. (Adapted from Muller, Churchland, and Sejnowski, 2024).

Sequence learning. The basal ganglia can learn sequences of actions to achieve goals and rewards. Dopamine neurons in the substantia nigra pars compacta (SNc) compute reward prediction error for actions and spoken words, and in the ventral tegmental area (VTA) for thoughts and plans. The prediction error is then used to modify synaptic strengths via the temporal difference learning algorithm (Montague, Dayan, and Sejnowski, 1996). TD learning constructs a value function used in model-free reinforcement learning, enabling fast decisions based on past experience. The value function is distributed across neurons in the ventral striatum and the orbital frontal cortex. These same structures could learn associations within sequences of words in the posterior cortex and thoughts in the prefrontal cortex, which, in transformers, are implemented with self-attention.

Discussion

It has been 30 years since Mainen and Sejnowski (1995) established the millisecond precision of spike initiation in cortical pyramidal neurons *in vitro*. Since then, evidence has accumulated for precise spike timing *in vivo*, but a function for this level of precision has yet to be established. Cortical traveling waves have been reported in hundreds of papers (Muller, Chavane, Reynolds, Sejnowski, 2018). Without a known function, it has been difficult to generate much interest in cortical waves. The traditional approach to assigning function to spikes is by their impact on behavior. Because spikes on the second tier are sparse, they do not directly move muscles or interfere with the sensorimotor rate code. Traveling waves are more ephemeral and are not anchored to the Broadman map of cortical areas. Thinking is also ephemeral and crosses borders, but no one would deny its importance. Large-scale spiking models will be needed to explore the integration of the neural mechanisms underlying long-term working memory.

Smoking gun? We have evidence that traveling waves can affect perception when a visual stimulus is near threshold. As spontaneous waves pass through area MT of a marmoset, the wave's phase modulates the responses of neurons to low-contrast stimuli and is highly correlated with behavioral reports that the weak signal was perceived (Davis, Muller, Martinez-Trujillo, Sejnowski, Reynolds, 2020). This reveals a link to behavior, but with weak strength on the second tier. Waves do not change percepts above threshold or directly cause an action. The coordination of traveling waves is through shared subthreshold membrane potentials, organized by massive but weak synaptic signaling. This is a form of *E pluribus unum* in populations of neurons.

Predictions. Two testable functions for cortical waves are suggested here: First, transmitting information across the cortex to create spacetime population codes, a type of holographic representation; and second, triggering STDP at synapses to temporarily change synaptic strengths. Together, waves and STDP could create long-term working memories that last for hours. Protocols that induce STDP by pairing pre- and postsynaptic bursts of spikes at 10-40 Hz mesh well with the characteristics of traveling waves (Markram et al., 1997). Interfering with spike timing, traveling waves, and STDP should affect the cortical processing that underlies long-term working memory.

Noise. Does every spike matter? Reaction times to sensory stimuli have a broad distribution, suggesting noise in the sensorimotor chain. The release and recycling of neurotransmitters at

cortical synapses is expensive. The release is probabilistic, with the probability of release proportional to the strength of the synapse (Schikorski and Stevens, 1997). The largest synapses are the most reliable. Paradoxically, the majority of synapses are small and have the lowest release probabilities. We can appeal to the Central Limit Theorem, which states that for large sample sizes, the distribution of sample means from any distribution will be approximately Gaussian. This mitigates, but does not eliminate, synaptic noise, making possible the observed spike-timing precision in vivo (Figs. 3-5). The precision of spike timing on the second tier is compatible with the presence of noise on the first sensorimotor tier that does not depend on spike timing and is consistent with the mixture of precisely timed spikes and rate-coded spikes in Fig. 6. Brain noise remains an open question.

Thinking. Brains are self-generative in a way that does not exist in current large language models (LLMs). When your dialog with an LLM ends, its internal activity stops. We continue to think and plan without any external prompting. LLMs have a feedforward hierarchy of layers. In contrast, cortical hierarchies have massive feedback between layers and recurrent connectivity within layers. These feedback connections support attention, decision making, and learning, which are regulated by powerful neuromodulatory systems.

Attention: You are more likely to remember something the next day if you pay attention to it. Attention releases the neuromodulator acetylcholine, shifting the state of local cortical circuits and triggering bursts of gamma waves (Fig. 8) grouped within lower-frequency theta waves (Howe et al., 2017). Cortical states marked by attention in the hippocampus can then be transferred to long-term memory via global traveling waves during sleep spindles (Muller et al., 2016). Traveling waves control what we remember.

Psychosis. Abnormal gamma waves are linked to thought disorders. Schizophrenia patients have reduced amplitude gamma waves, which is associated with symptoms that include delusions and disorganized thinking. Similar symptoms appear after serial use of a party drug, Special K. After several days of raving under the influence of Special K, ravers present with symptoms indistinguishable from a schizophrenic break. Fortunately, the symptoms go away after a few days. Special K is ketamine, an NMDA channel blocker and anesthetic at high doses. At low doses, ketamine induces hallucinations and dissociative out-of-body experiences.

This ketamine model of human psychosis has been studied in mice, which has revealed the underlying neural mechanisms (Behrens and Sejnowski, 2009). Ketamine downregulates the enzyme that synthesizes the inhibitory neurotransmitter GABA in basket cells. Reduced inhibition unbalances cortical circuits, making them hyperactive, and also reduces gamma waves, as in schizophrenia patients (Jadi, Behrens, Sejnowski, 2016). Ketamine is effective at relieving clinical depression, perhaps by boosting reduced cortical activity to normal levels. Additional experiments are needed to confirm a causal role for reduced gamma waves in psychosis.

Conclusion

The debate between rate coding and spike timing is a false dichotomy. There is room for multiple codes to serve multiple functions. For sensorimotor tasks, rate coding is well established. The evidence presented here for high-precision spike timing underpinning long-term working memory is provisional and needs direct experimental tests of predictions. Methods that selectively manipulate spike timing are needed.

There are many open questions: How do neuromodulators shape temporary STDP and the long-term dynamics in the second-tier network? How does STDP interact with synaptic tagging and capture and active currents in dendrites (O'Donnell and Sejnowski, 2014)? How are working memories selected for integration into long-term memories? Could the second-tier temporary network form a global workspace for cognition (Baars, 1988)?

Many philosophical questions about the "mind" could have scientific answers if we knew more about the neural mechanisms underlying cognition. Descartes (1637) put it simply: "Cogito, ergo sum." A corollary of this dictum is: "Once we understand thinking, we will know ourselves." We should be open to more surprises as we explore the global dynamics of spiking neurons and their underlying subthreshold dynamics over a wide range of time scales.

Acknowledgements

This paper is based on a talk given on November 14, 2025, at a gathering of former lab members celebrating the 35th anniversary of the Computational Neurobiology Laboratory (CNL) at the

Salk Institute. The focus was on research arising from CNL. An extended, more complete history of spike timing is planned. My research is supported by an NIH Director's Pioneer Award (DP1NS149613) and by a grant from the Office of Naval Research (N00014-23-1-2069).

Appendix

My Ph.D. thesis at Princeton was on "A Stochastic Model of Nonlinearly Interacting Neurons," which collected several papers that I had previously published based on traditional rate-coded recurrent neural network models with a sigmoid activation function (Sejnowski, 1976a). In one of these papers, I proved that the covariance equation is linear when the membrane potentials are Gaussian distributed (Sejnowski, 1976b). The interaction matrix governing the covariance is modulated by the activity levels of the network's nonlinearly interacting neurons, and only the sparse network of neurons near threshold propagates correlated activity (Sejnowski, 1981).

Theorem (Sejnowski, 1976b): A network of neurons is stochastically interacting

$$\tau \frac{d}{dt} \phi_a + \phi_a = \eta_a + \sum_b K_{ab} L_b(\phi_b)$$

where ϕ_a are the membrane potentials, K_{ab} is the interaction matrix, $L_b(\phi_b)$ is a sigmoid activation function, and η_a are the inputs. If the membrane potentials are Gaussian, then the covariance equation is linear with the interaction matrix:

$$K'_{ab}(t) = K_{ab} P'_b(\hat{\phi}_b(t))$$

where P'_b is $E[L_b(\phi_b)]$, the derivative of the expectation of the sigmoid $L_b(\phi_b)$ with respect to the mean membrane potential. Intuitively, P'_b is nearly zero when the membrane potential is well below or well above threshold, so that covariance only depends on the interactions between those neurons near threshold. The covariance interaction matrix rapidly changes with the mean membrane potentials. Traveling waves are sparse, and only a sparse skeleton of connections between them controls the dynamics of their covariance during a wave (Sejnowski, 1981). I also worked on a model of synaptic plasticity to store the covariance in long-term memory

(Sejnowski, 1977). Covariance is a second-order measure of coordinated activity and is all that is needed for Gaussian statistics.

This model is stochastic but not spike-based. In the linear-Nonlinear-Poisson (LNP) cascade model, a rate-modulated Poisson generator converts a continuous signal into random spikes. Neither of these models can fully capture the higher-order statistics of spikes in neural circuits. Spike synchrony is a higher-order property, and traveling wave dynamics in a spatial array of neurons could support more complex computational systems (Keller, Muller, Sejnowski, Welling, 2024b).

My research has come full circle, back to where it began many decades ago. The missing link was spike timing, and the turning point was Mainen and Sejnowski (1995).

References

Adelson, E. H., Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*(2), 284-299.

Allman, J. M., & Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (Aotus trivirgatus). *Brain research*, *31*(1), 85-105.

Allman, J., Miezin, F., McGuinness, E. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local–global comparisons in visual neurons. *Annual review of neuroscience* (1985).

Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

Bair, W., Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation*, *8*(6), 1185-1202.

Ballard, D. H. Hinton, G. E. Sejnowski (1983). T. J. Parallel Visual Computation, Nature, 306, 21-26,

Batabyal, T., Brincat, S. L., Donoghue, J. A., Lundqvist, M., Mahnke, M. K., Miller, E. K. (2025). State–Space Trajectories and Traveling Waves Following Distraction. *Journal of Cognitive Neuroscience*, 1-21.

Behrens, M. M., Sejnowski, T. J. (2009). Does schizophrenia arise from oxidative dysregulation of parvalbumin-interneurons in the developing cortex? *Neuropharmacology*, 57(3), 193-200.

Berry, K. P., and Nedivi, E. (2017). Spine dynamics: are they all the same? *Neuron*, 96(1), 43-55.

Berry. M. J., A Theory of Thalamocortical Loops in Evidence Accumulation and Decision-Making (2025). bioRxiv. https://www.biorxiv.org/content/10.64898/2025.12.23.696263v1

Bhattacharya, S., Brincat, S. L., Lundqvist, M., Miller, E. K. (2022). Traveling waves in the prefrontal cortex during working memory. *PLoS Computational Biology*, 18(1), e1009827.

Bi, G. Q., Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24), 10464-10472.

Buracas, G. T., Zador, A. M., DeWeese, M. R., & Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20(5), 959-969.

Buzsáki, György. (2019). *The brain from inside out*. Oxford University Press.

Chen, Y., Zhang, H., Cameron, M., Sejnowski, T. (2024). Predictive sequence learning in the hippocampal formation. *Neuron*, 112(15), 2645-2658.

Davis, Z. W., Muller, L., Martinez-Trujillo, J., Sejnowski, T., Reynolds, J. H. (2020). Spontaneous travelling cortical waves gate perception in behaving primates. *Nature*, 587(7834), 432-436.

Davis, Z. W., Benigno, G. B., Fletterman, C., Desbordes, T., Steward, C., Sejnowski, T. J., Reynolds, J. H., Muller, L. (2021). Spontaneous traveling waves naturally emerge from horizontal fiber time delays and travel through locally asynchronous-irregular states. *Nature Communications*, 12(1), 6057.

Descartes, R. (1637). *Discours de la Méthode pour bien conduire sa raison, et chercher la vérité dans les sciences*.

Destexhe, A., Sejnowski, T. (2023). *Thalamocortical Assemblies: Sleep spindles, slow waves and epileptic discharges*. Oxford University Press.

Mainen, Z. F., and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268, 1503-1506.

Ericsson K. A. and Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211.

Feldman, D. E. (2012). The spike-timing dependence of plasticity. *Neuron*, 75(4), 556-571.

Fellous, J. M., Tiesinga, P. H., Thomas, P. J., Sejnowski, T. J. (2004). Discovering spike patterns in neuronal responses. *Journal of Neuroscience*, 24, 2989-3001.

Gilbert, C. D., Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, *9*(7), 2432-2442.

Haider, B., Krause, M. R., Duque, A., Yu, Y., Touryan, J., Mazer, J. A., & McCormick, D. A. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron*, 65(1), 107-121.

Holtmaat AJ, Trachtenberg JT, Wilbrecht L, Shepherd GM, Zhang X, Knott GW, Svoboda K. (2005) Transient and persistent dendritic spines in the neocortex in vivo. *Neuron*. 45(2):279-91

Hafting, T., Fyhn, M., Bonnevie, T., Moser, M. B., Moser, E. I. (2008). Hippocampus-independent phase precession in entorhinal grid cells. *Nature*, 453(7199), 1248-1252.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, *28*(10), 2539-2550.

Haziza, S., Chrapkiewicz, R., Zhang, Y., Kruzhilin, V., Li J., Li J, Delamare, G., Swanson, R., Buzsáki, G., Kannan, M., Vasan, G., Lin, M. Z., Zeng, H., Daigle, T. L., Schnitzer, M. J. (2025).

Imaging high-frequency voltage dynamics in multiple neuron classes of behaving mammals. *Cell*, 188(16), 4401-4423.

Howe, W. M., Gritton, H. J., Lusk, N. A., Roberts, E. A., Hetrick, V. L., Berke, J. D., & Sarter, M. (2017). Acetylcholine release in prefrontal cortex promotes gamma oscillations and theta–gamma coupling during cue detection. *Journal of Neuroscience*, 37(12), 3215-3230.

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Computation*, 18(2), 245-282.

Izhikevich, E. M. (2025). Spiking Manifesto. arXiv:2512.11843

Jadi, M. P., Behrens, M. M., & Sejnowski, T. J. (2016). Abnormal gamma oscillations in N-methyl-D-aspartate receptor hypofunction models of schizophrenia. *Biological Psychiatry*, 79(9), 716-726.

Jadi, M. P. and Sejnowski, T. J. (2014). Cortical Oscillations Arise from Contextual Interactions that Regulate Sparse Coding, *Proceedings of the National Academy of Sciences, USA*, 111, 6780-6785.

Jin, D. Z., Fujii, N., and Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences, USA, 106*(45), 19156-19161.

Johnson, B.A., and Frostig, R. D. (2016). Long, intrinsic horizontal axons radiating through and beyond rat barrel cortex have spatial distributions similar to horizontal spreads of activity evoked by whisker stimulation. *Brain Structure and Function* 221: 3617-3639.

Keller, T. A., Muller, L., Sejnowski, T., Welling, M. (2024a). Traveling Waves Encode the Recent Past and Enhance Sequence Learning, *International Conference on Learning Representations* (*ICLR*).

Keller, T. A., Muller, L., Sejnowski, T., Welling, M. (2024b). A spacetime perspective on dynamical computation in neural information processing systems. *arXiv preprint arXiv:2409.13669*.

Kim, R., and Sejnowski, T. J. (2021). Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *Nature Neuroscience, 24*(1), 129-139.

Koller, D. P., Schirner, M., and Ritter, P. (2024). Human connectome topology directs cortical traveling waves and shapes frequency gradients. *Nature Communications*, *15*(1), 3570.

Loewenstein, Yonatan, Annerose Kuras, and Simon Rumpel. (2011) Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo." *J. Neuroscience.* 31.26: 9481-9488.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*(5216), 1503-1506.

Markram H, Lübke J, Frotscher M, Sakmann B. (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*. 275(5297):213-5.

Marr, D., Poggio, T. (1976). Cooperative Computation of Stereo Disparity: A cooperative algorithm is derived for extracting disparity information from stereo image pairs. *Science*, 194(4262), 283-287.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936-1947.

Muller, L., Churchland, P. S., Sejnowski, T. J. (2024). Transformers and cortical waves: encoders for pulling in context across time. *Trends in Neurosciences* 47(10):788-802.

Muller, L., Chavane, F., Reynolds, J., Sejnowski, T. J. (2018). Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience*, 19(5), 255-268.

Muller, L., Piantoni, G., Koller, D., Cash, S. S., Halgren, E., & Sejnowski, T. J. (2016). Rotating waves during human sleep spindles organize global patterns of activity that repeat precisely through the night. *Elife*, 5, e17267.

O'Donnell, C., & Sejnowski, T. J. (2014). Selective memory generalization by spatial patterning of protein synthesis. *Neuron*, 82(2), 398-412.

Patel, J., Fujisawa, S., Berényi, A., Royer, S. and Buzsáki, G. (2012). Traveling theta waves along the entire septotemporal axis of the hippocampus. *Neuron*, 75(3), 410-417.

Perin, R., Berger, T. K. and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences USA, 108*(13), 5419-5424.

Pribram, K. H. (2013). Brain and perception: Holonomy and structure in figural processing. Psychology Press.

Qasim, S. E., Fried, I., & Jacobs, J. (2021). Phase precession in the human hippocampus and entorhinal cortex. *Cell*, 184(12), 3242-3255.

Reichardt, W. (1957). Autokorrelations-Auswertung als Funktionsprinzip des Zentralnervensystems: (bei der optischen Bewegungswahrnehmung eines Insektes). *Zeitschrift für Naturforschung B*, 12(7), 448-457.

Reinagel, P. and Reid, R. C. (2002). Precise firing events are conserved across neurons. *Journal of Neuroscience*, 22(16), 6837-6841.

Schikorski T., Stevens C. F. (1997). Quantitative ultrastructural analysis of hippocampal excitatory synapses. *Journal of Neuroscience* 17:5858–5867.

Sejnowski, T. J., On global properties of neuronal interaction, *Biological Cybernetics* 22, 85-95 (1976a).

Sejnowski, T. J. On the Stochastic Dynamics of Neuronal Interaction (1976b). *Biological Cybernetics*, 22, 203-211.

Sejnowski, T. J., (1977a). Statistical constraints on synaptic plasticity, *Journal of Theoretical Biology* 69, 385-389

Sejnowski, T. J. (1977b). Storing Covariance with Nonlinearly Interacting Neurons, *Journal of Mathematical Biology*, 4, 203-211.

Sejnowski, T. J. (1981). Skeleton Filters in the Brain Hinton, G. E. Anderson, J. A. (Ed.), In: Parallel Models of Associative Memory, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 189-212,

Sejnowski, T. J., (2025) Thinking about thinking: AI offers theoretical insights into human memory, The Transmitter, https://www.thetransmitter.org/human-neurotechnology/thinking-about-thinking-ai-offerstheoretical-insights-into-human-memory/

Singer, W. (2021). Recurrent dynamics in the cerebral cortex: Integration of sensory evidence with stored knowledge. *Proceedings of the National Academy of Sciences*, *118*(33), e2101043118.

Lubenov, E. V., Siapas, A. G. (2009). Hippocampal theta oscillations are travelling waves. *Nature*, *459*(7246), 534-539.

Tiesinga, P. Fellous, J.M. Sejnowski, T. J. Regulation of spike timing in visual cortical circuits, *Nature Reviews Neuroscience*, 9: 97-109, 2008

Van Vreeswijk, C., Abbott, L. F., Bard Ermentrout, G. (1994). When inhibition not excitation synchronizes neural firing. *Journal of Computational Neuroscience*, *1*(4), 313-321.

Wang, X. J. (2020). Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nature Reviews Neuroscience*, *21*(3), 169-178.

Yang, G., Lai, C. S. W., Cichon, J., Ma, L., Li, W., & Gan, W. B. (2014). Sleep promotes branch-specific formation of dendritic spines after learning. *Science*, *344*(6188), 1173-1178.

Ye, Z., Bull, M. S., Li, A., Birman, D., Daigle, T. L., Tasic, B., Zeng, H., Steinmetz, N. A. (2023). Brain-wide topographic coordination of traveling spiral waves. *bioRxiv*, 2023-12.

Yoshimura, Y., Dantzker, J. L and Callaway, E. M. (2005). Excitatory cortical neurons form fine-scale functional networks. *Nature*, *433*(7028), 868-873.