An Indexing Theory for Working Memory based on Fast Hebbian Plasticity

Florian Fiebig, Pawel Herman, Anders Lansner
1 Computational Science and Technology, Royal Institute of Technology, Stockholm, Sweden
2 Department of Mathematics, Stockholm University, Sweden

Objectives

- Extend a previous spiking neural network model of prefrontal cortex (PFC) with fast Hebbian plasticity to also include **interactions between short-term (STM) and long-term memory stores (LTM)**.
- Implement a functional multi-item WM leveraging PFC Indices, transiently formed assemblies that can bind and maintain multi-modal LTM representations with dynamic on-line **learning, updating, and associative recall**.
- Identify key anatomical and electrophysiological parameters that realize known properties of WM activity (e.g. intermittent gamma bursting) at comparable **corticocortical latencies**[2] between macaque PFC and inferior temporal cortex (ITC).

Introduction

We present the extension of a spiking neural network model of PFC with fast Hebbian plasticity[1] to also include interactions between short-term and long-term memory stores. We investigated how PFC could bind and maintain multi-modal long-term memory representations in parietotemporal cortical areas by simulating three cortical patches based on macaque data. The resulting biophysically constrained multiarea network model allows for a quantitative analysis of interactions between STM and LTM and suggests a novel approach to the 'binding problem'.

The Binding Problem

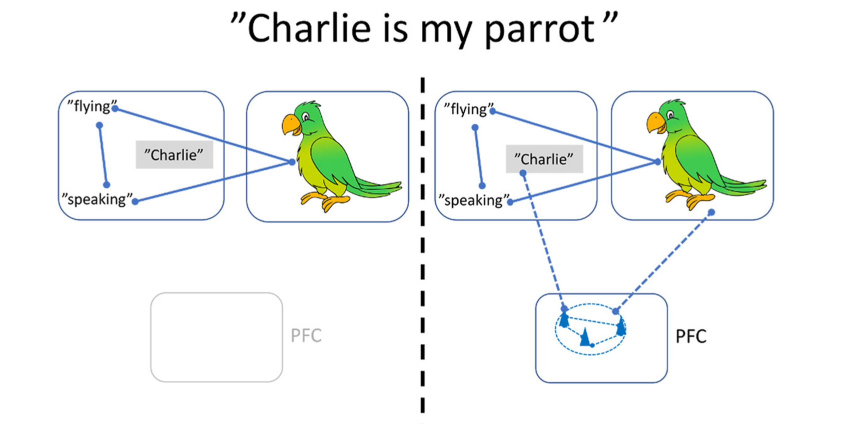

Figure 1: **Variable value binding via index in PFC.** Initially, the multimodal representation of 'parrot' exists in LTM comprising symbolic and subsymbolic components side by side with 'Charlie' as a representation of a proper name. It is hypothesized here that when someone states that 'Charlie is my parrot', the name 'Charlie' is temporarily and reciprocally bound to the parrot representation via PFC, mediated by fast Hebbian plasticity. Pattern completion effect now allows 'Charlie' to trigger the entire assembly and, analogously, makes 'flying' or the sight of a given parrot trigger 'Charlie'. If important enough or repeated a couple of times this association could consolidate in LTM.

Cortical Model Architecture and Connectivity

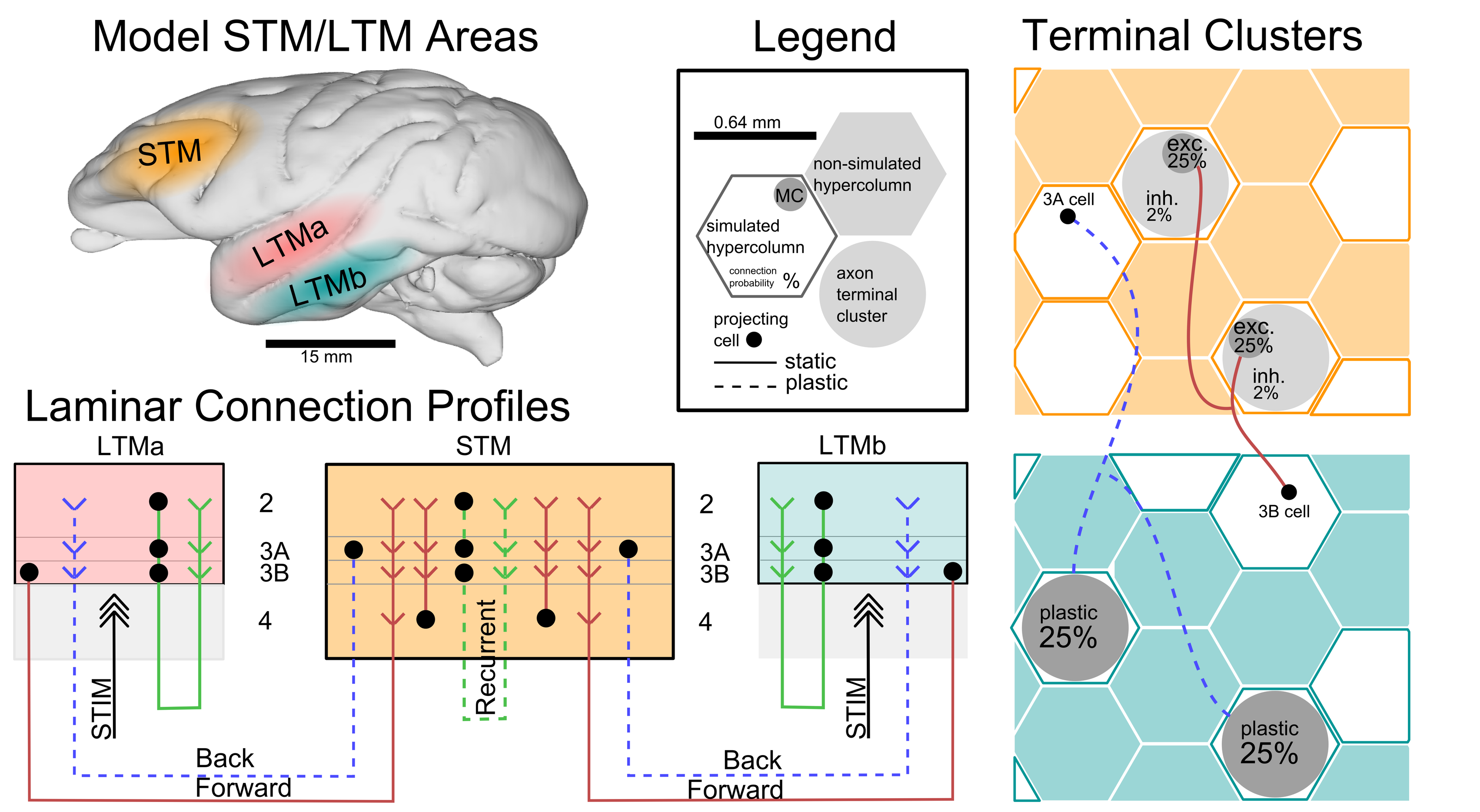

Figure 2: The model organizes cells into grids of nested hypercolumns (HC) and minicolumns (MC). Each network spans several hundred mm^2, and the simulated columns constitute a spatially distributed subsample. Pyramidal cells in the simulated supragranular layers form connections both within and across columns of their respective networks. Corticocortical connections are sparse (<1% connection probability) and implemented with terminal clusters and specific laminar connection profiles realizing a direct supragranular forward-projection, as well as a supragranular back-projection. Layer 2/3 recurrent connections in STM and corticocortical backprojections feature fast Hebbian plasticity.

Rapid LTM Indexing and WM Updating

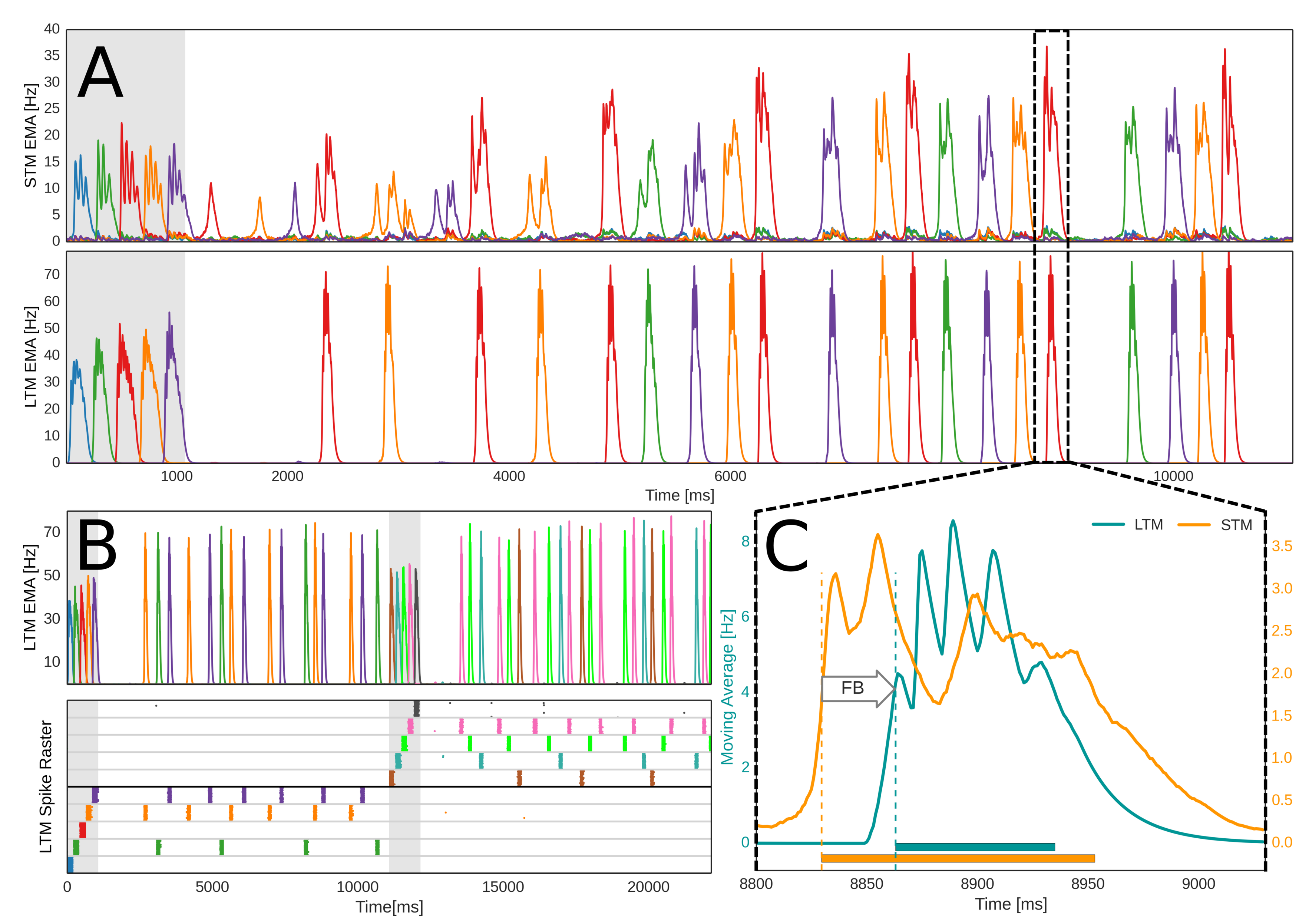

Figure 3: **Indexing and reactivation of LTM. A:** Exponential moving averages of pattern-specific subpopulations in STM (top) and LTM (bottom). 0-1 s (shaded grey), five LTM memories are cued via brief external stimuli. Plasticity is temporarily boosted during this time and activated attractors thus indexed in STM. Thereafter (1-11 s), strong noise drive to STM causes random activations and consolidation of pattern-specific subpopulations in STM, while backprojections reactivate associated LTM memories. **B: Rapid Updating of WM**, replacing a previously maintained set of memories. **C: STM-to-LTM Feedback dynamics during a reactivation event.** Horizontal bars at the bottom indicate activation half-width. Onset is denoted by vertical dashed lines.

Multi-modal LTM Binding, Maintenance, and Recall

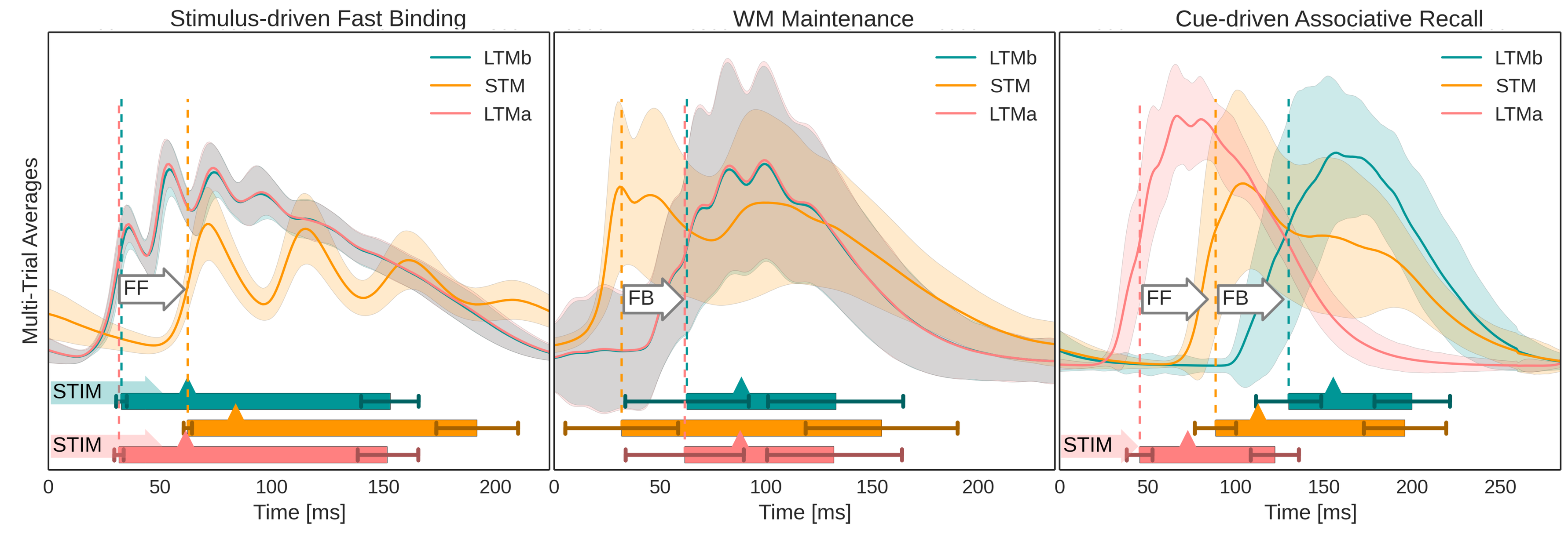

Figure 4: **Multi-trial peri-stimulus activity traces from the three cortical patches.**. Pattern Activity denoted by a rectangle underneath the traces. Error bars and shaded areas indicate one standard deviation (100 Trials, 495 Traces). Shaded arrows denote stimuli. White arrows annotate feedforward (FF) and feedback (FB) delay.

Cortical Model Activity

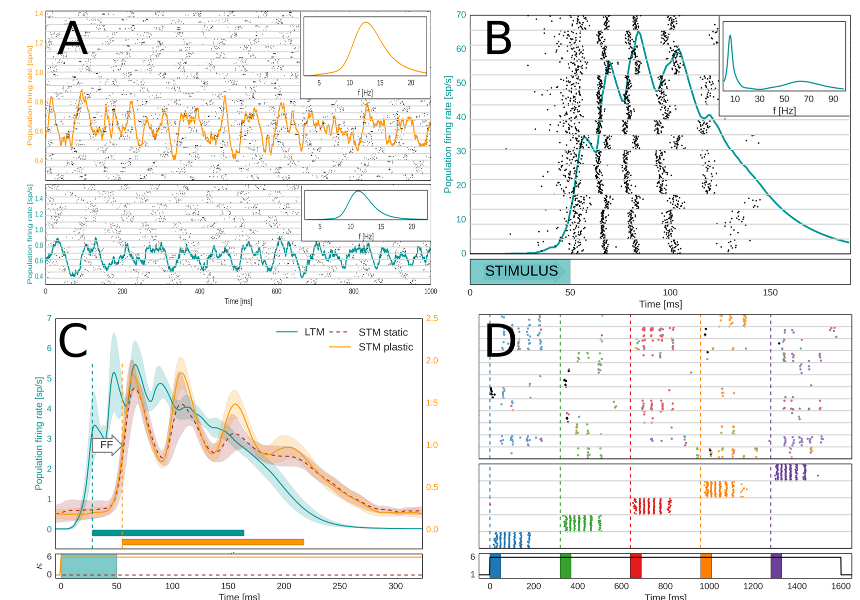

Figure 5: **A:** Alpha/beta oscillations characterize the ground state. **B:** LTM attractor activations express as theta-nested gamma bursts. **C:** LTM-to-STM forward dynamics following a 50 ms targeted LTM stimulus **D:** STM Activation (top) by five different LTM-attractors (bottom).Spikes are colored according to each cells dominant pattern-selectivity.

Conclusion

Our simulation instantiates a novel theory for WM and results demonstrate how simultaneous, brief multi-modal memory cues could build a temporary joint memory representation as an 'index' in the PFC by means of fast Hebbian synaptic plasticity. The latter can then reactivate spontaneously and thereby also reactivate the associated long-term representations. Cueing one LTM item rapidly pattern-completes the associated un-cued item via PFC. The STM network in PFC updates flexibly as new stimuli arrive thereby gradually over-writing older representations and bringing on-line the new LTM representations. In a wider context, this WM model suggests a novel explanation for 'variable binding', a long-standing and fundamental phenomenon in cognitive neuroscience, which is still poorly understood in terms of detailed neural mechanisms.

References

[1] Florian Fiebig and Anders Lansner. A Spiking Working Memory Model Based on Hebbian Short-Term Potentiation. *The Journal of Neuroscience*, 37(1):83–96, 2017.

[2] Hyoe Tomita, Machike Ohbayashi, Kiyoshi Nakahara, Isao Hasegawa, and Yasushi Miyashita. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, 401(6754):699–703, oct 1999.

[3] F. Fiebig, P. Herman, and A. Lansner. An Indexing Theory for Working Memory based on Fast Hebbian Plasticity. *eNeuro, March 2020,* doi.org/10.1523/ENEURO.0374-19.2020.

Acknowledgements

This work was supported by the EuroSPIN Erasmus Mundus doctoral program, SeRC (Swedish e-science Research Center), and StratNeuro (Strategic Area Neuroscience at Karolinska Institutet, UmeÃě University and KTH). Simulations were performed using computing resources by the Swedish National Infrastructure for Computing at PDC Centre for High Performance Computing.