UNIVERSITÉ DE PARIS

ED Cerveau-Congition-Comportement (158)

Laboratoire Cerebral Dynamics, Plasticity and Learning (UMR 8002)

Exploring the dynamics of Prefrontal cortex and its implication in working memory and learning

Alejandro SUÁREZ PÉREZ

Thèse de doctorat en Neurosciences

Dirigée par **Prof. David HANSEL**

Présentée et soutenue publiquement le 31 de janvier de 2022

Devant un jury composé de:

David HANSEL, DR, Université de Paris, directeur de thèse
Thomas BORAUD, DR, Université de Bordeaux, rapporteur
David GOLOMB, professeur des universités, Université Ben-Gurion, rapporteur
Gianluigi MONGILLO, CR, Sorbonne Université, examinateur
Carl VAN VREESWIJK, CR, Université de Paris, examinateur
Carole LEVENES, CR, Université de Paris, examinatrice

Abstract

Working memory or, maintaining a mental representation of past immediate information to drive coherent actions, is critical feature in the realisation of complex cognitive functions such as planning, decision making, learning, thinking, etc. Experimental evidence point that neural correlates of working memory involves persistent selective activity in the Prefrontal Cortex (PFC). Main accepted mechanisms of persistent activity rely on strong recurrent excitation in PFC whereas selectivity or spatial tuning arises from columnar PFC architecture together with lateral inhibition. However, the mechanisms underlying network bistability remain elusive. On the other hand, mental illnesses such Schizophrenia present PFC disfunction and particularly Working Memory (WM) deficits. Although main theories of the Schizophrenia etiology underline the role of PFC dysconnectivity with subcortical areas, the mechanisms underlying WM deficits in Schizophrenia remain elusive. In this thesis we argue the idea that understanding the neurobiology of the PFC and WM will provide an important knowledge in order to unravel the etiology of mental illnesses with salient cognitive symptoms such Schizophrenia. With the aim to shed some light into these questions, in this thesis, we present original mechanisms for network bistability based on the nonlinearities in the neuronal input-output transfer functions. With this model, we study different mechanisms of state transitions. Additionally, we study how alterations of the biological parameters seen in schizophrenia can shape the spatiotemporal dynamics. Finally, we evaluate a promising treatment for schizophrenia based on cerebellar stimulation.

Keywords: Visuospatial working memory; persistent activity; direction selectivity; neural nonlinerities; neuronal transfer functions; prefrontal cortex; cerebellum; cerebellar stimulation; schizophrenia.

Résumé

La mémoire de travail (MT), ou le maintien d'une représentation mentale des informations du passé immédiat pour conduire des actions cohérentes est une caractéristique critique dans la réalisation de fonctions cognitives complexes telles que la planification, la prise de décision, l'apprentissage, la pensé, etc. L'évidence expérimentale indique que le corrélat neuronal de la MT comprend une activité persistant et sélective dans le cortex préfrontal (CPF). Les mécanismes les plus acceptés par rapport à l'activité persistante s'appuient sur une forte excitation récurrent dans le CPF tandis que l'activité sélective ou hétérogène en l'espace neuronal découle de l'architecture en colonne du CPF associée à une forte inhibition latérale. Néanmoins, les mécanismes sous-jacents à la multistabilité du réseau restent incertains. D'autre part, les maladies mentales telles que la schizophrénie présentent une marquée dysfonction du CPF et particulièrement déficits en la MT. Bien que les principales théories concernant l'étiologie de la schizophrénie soulignent le rôle de la disconnectivité entre le CPF et areas subcorticales, les mécanismes concernant les déficits de MT liées à la squizophrénie demeurent vagues. Dans cette thèse nous soutenons l'idée que la compréhension de la neurobiologie du CPF et les mécanismes de la MT révêléront des connaissances importantes afin d'éclaircir l'étiologie des maladies mentales avec des symptômes cognitifs saillants tels que la schizophrénie. Dans le but d'apporter certains éclaircissements sur ces questions, dans cette thèse, nous présentons des modèles qui nous permettent d'étudier et de proposer des mécanismes originaux de mutistabilité de réseau fondés sur les non linéarités de la fonction de transfert neuronale. En utilisant ces modèles, nous étudions comment des altérations des paramètres biologiques, comme c'est le cas dans la schizophrénie, pourraient transformer la dynamique spatio-temporelle du réseau de neurones. Finalement, nous évaluons un traitement prometteur pour la schizophrénie basée sur la stimulation électrique du cervelet.

Mots clefs: Mémoire de travail visuo-spatiale; activité persistante; sélectivité de direction; non linéarités neuronales; fonction de transfert neuronale; cortex préfrontal; cervelet; stimulation cérébellaire schizophrénie.

Resumen

La memoria de trabajo (MT), es decir, el mantenimiento de una representación mental de la información del pasado inmediato para llevar a cabo acciones coherentes, es una característica crítica en la realización de funciones cognitivas complejas como la planificación, la toma de decisiones, el aprendizaje, el pensamiento, etc. Las evidencias experimentales apuntan a que los correlatos neuronales de la memoria de trabajo implican una actividad selectiva persistente en el cortex prefrontal (CPF). Los principales mecanismos mayormente aceptados sobre la actividad persistente se basan en una fuerte excitación recurrente en el CPF, mientras que la selectividad o la sintonización espacial surge de la arquitectura columnar del CPF junto a una fuerte inhibición lateral. Sin embargo, los mecanismos que subyacen a la biestabilidad de la red siguen siendo un tanto esquivos. Por otra parte, las enfermedades mentales tales como la esquizofrenia, presentan una marcada disfunción del CPF y, en particular, déficits de la MT. Aunque las principales teorías sobre la etiología de la esquizofrenia subrayan el papel de la desconexión del CPF con las áreas subcorticales, los mecanismos subyacentes a los déficits de la MT en la esquizofrenia no son del todo claros. En esta tesis defendemos la idea de que la comprensión de la neurobiología del CPF así como de la MT nos proporcionará un conocimiento de vital importancia para desentrañar la etiología de los síntomas cognitivos en enfermedades mentales como la esquizofrenia. Con el objetivo de arrojar algo de luz sobre estas cuestiones, en esta tesis presentamos diversos mecanismos originales sobre la multiesabilidad de red basados en las no linealidades de las funciones de transferencia neuronales. Con este modelo, estudiamos diferentes mecanismos de transiciones entre los estados de actividad espontanea y estados de memoria. Además, investigamos cómo las alteraciones en los parámetros biológicos observadas en la esquizofrenia pueden dar forma a la dinámica espacio-temporal de la red neuronal. Finalmente, y basándonos en la hipótesis de que los síntomas cognitivos en la esquizofrenia podrían originarse debido a una desconectividad entre el CPF y areas subcorticales tales como el cerebelo, evaluamos un tratamiento prometedor para la esquizofrenia basado en la estimulación cerebelar.

Palabras clave: Memoria de trabajo visuoespacial; actividad persistente; selectividad direccional; no linearidades neuronales; función de transferencia neuronal; cortex prefrontal; cerebelo; estimulación cerebelar; esquizofrenia.

Resum

La memòria de treball (MT), és a dir, el manteniment d'una representació mental de la informació del passat immediat per a dur a terme accions coherents, és una característica crítica en la realització de funcions cognitives complexes com la planificació, la presa de decisions, l'aprenentatge, el pensament, etc. Les evidències experimentals apunten al fet que els correlats neuronals de la memòria de treball impliquen una activitat selectiva persistent en el cortex prefrontal (CPF). Els principals mecanismes majorment acceptats sobre l'activitat persistent es basen en una forta excitació recurrent en el CPF, mentre que la selectivitat o la sintonització espacial sorgeix de l'arquitectura columnar del CPF juntament amb una forta inhibició lateral. No obstant això, els mecanismes que subjeuen a la biestabilitat de la xarxa continuen sent esquius. D'altra banda, les malalties mentals com ara l'esquizofrènia, presenten una marcada disfunció del CPF i, en particular, dèficits de la MT. Encara que les principals teories sobre l'etiologia de l'esquizofrènia subratllen el paper de la desconnexió del CPF amb les àrees subcorticals, els mecanismes subjacents als dèficits de la MT en l'esquizofrènia no són del tot clars. En aquesta tesi defensem la idea que la comprensió de la neurobiologia del CPF així com de la MT ens proporcionarà un coneixement de vital importància per a desentranyar l'etiologia dels símptomes cognitius en malalties mentals com l'esquizofrènia. Amb l'objectiu de esclarir aquestes qüestions, en aquesta tesi presentem diversos mecanismes originals sobre la multiesabilitat de xarxa basats en les no linealitats de les funcions de transferència neuronals. Amb aquest model, estudiem diferents mecanismes de transicions entre els estats d'activitat espontanea i estats de memòria. A més, investiguem com les alteracions en els paràmetres biològics observades en l'esquizofrènia poden donar forma a la dinàmica espai-temporal de la xarxa neuronal. Finalment, i basant-nos en la hipòtesi que els símptomes cognitius en l'esquizofrènia podrien originar-se a causa d'una desconectivitat entre el CPF i areas subcorticals com ara el cerebel, avaluem un tractament prometedor per a l'esquizofrènia basat en l'estimulació cerebelar.

Paraules clau: Memòria de treball visuoespacial; activitat persistent; selectivitat direccional; no linearitats neuronals; funció de transferència neuronal; cortex prefrontal; cerebel; estimulació cerebelar; esquizofrènia.

Résumé Substantiel

Dans cette thèse nous explorons la dynamique neuronal du cortex préfrontal (CPF) impliqué dans la mémoire de travail (MT) et ainsi les mécanismes sous-jacents a niveau neuronal et de réseau. D'autre coté, nous étudions comment une activité aberrante dans le CPF peut générer des états pathologiques similaires à ceux qui sont présentes dans la schizophrénie. Nous formulons ensuite l'hypothèse que cette activité aberrante pourrait être liée à une diminution des afférences et nous justifions cette hypothèse en montrant un cas clinique de stimulation cérébelleuse qui semble améliorer la mémoire de travail chez le patient avec schizophrénie parmi autres symptômes cognitives.

La thèse est organisée en trois parties : un prologue consistant en une introduction générale et un aperçu du travail (chapter 1) suivi d'une revue de la littérature concernant le rôle du CPF dans la MT et son implication dans la schizophrénie ainsi que dans le cervelet (chapter 2) ; une partie concernant la recherche originale qui est subdivisée en chapter 3, chapter 4 et chapter 5. , chapter 4 et chapter 5, où nous présentons les résultats de trois projets séparément, chacun avec une introduction, des résultats, une discussion et une conclusion spécifiques ; et finalement, un épilogue consistant en une discussion générale et des conclusions dans chapter 6.

Nous passons maintenant à présenter plus en détail les trois parties concernant la recherche originale :

Dans le chapter 3, nous présentons un cadre théorique pour étudier les mécanismes de réseau de l'activité sélective persistante comme celle observée dans les CPF pendant les tâches MT. Nous étudions spécifiquement comment l'interaction entre les non-linéarités dans la fonction de transfert (FT) des neurones individuels et la dynamique du réseau récurrent donne lieu à une multi-stabilité entre l'activité continue et un continuum d'états dans lesquels l'activité est localisée. Nous montrons que le mécanisme qui sous-tend l'initiation et la fin de l'activité sélective persistante dépend des non-linéarités

des neurones individuels. Dans la première section, nous présentons un modèle de réseau de neurones connectés tous à tous avec une connectivité structurée en forme de "chapeau mexicain" par laquelle les neurones ayant des directions préférentielles (DP) similaires sont plus fortement connectés. Les non-linéarités de la FT neuronale permettent d'avoir une persistance et une sélectivité au sein d'un même réseau. Les deux non-linéarités étudiées sont l'accélération près du seuil de décharge et la concavité près de la partie saturante. Ces non-linéarités se retrouvent dans les FT de neurones typiquement mesurés dans des expériences *in vitro* (Arsiero et al., 2007; Lafon et al., 2017; Schiff and Reyes, 2012). Afin de simplifier les calculs analytiques, nous avons choisi que les FT soient des fonctions définies par segments linéaires. Dans l'accélération ou la non-linéarité expansive, la pente des FT augmente alors que dans la concavité ou la non-linéarité compressive, la pente diminue. Nous montrons que la nature de la non-linéarité détermine la configuration des états stables. Dans les deux cas, pour un paramètre de connectivité donné, nous trouvons une région d'état homogène de multistabilité et un état de bosse ("bump state"). Nous explorons ensuite les mécanismes permettant de mettre fin à l'état de bosse. En raison de la configuration des états stables, dans le modèle avec une non-linéarité expansive, la transition de l'état de bosse à l'état homogène peut être médiée par un input inhibitoire globale. D'autre part, dans le modèle avec une non-linéarité compressive, la transition de l'état d'activité persistante à l'état d'activité continue peut être induite par un input excitateur globale. Ce mécanisme est intéressant car la plupart des enregistrements électrophysiologiques des neurones CPF pendant la tâche MT montrent une augmentation globale du taux de décharge neuronal autour de la période de réponse. Le seul inconvénient de ce mécanisme est que les neurones au centre de l'état de bosse ont un taux de tir proche de la saturation, ce qui n'est pas cohérent avec les données expérimentales montrant que les neurones qui déchargent à leurs DP sont très éloignés des taux de décharge typiques de la saturation. Pour résoudre ce problème, nous proposons un mécanisme dans lequel l'interaction entre les populations excitatrices et inhibitrices avec des FT d'accélération différents donne lieu à une saturation effective. L'idée principale de ce mécanisme est qu'à un input donné,

la pente du FT inhibiteur est plus grande que la pente du FT excitateur, ce qui donne lieu à un scénario qualitativement similaire à celui de la non-linéarité compressive. Nous montrons différentes configurations des FTs excitateurs et inhibiteurs qui conduisent à une saturation effective et nous prouvons que les mêmes mécanismes concernant la fin de l'état de bosse par excitation globale sont toujours valables. Afin d'étudier comment les populations excitatrices et inhibitrices interagissent dans le mécanisme de saturation effective et d'étudier en détail la dynamique qui a lieu dans les transitions entre les états stables et la formation de modèles spatio-temporels d'activité, nous implémentons le mécanisme de saturation effective dans un modèle de réseau de neurones qui présentent décharges (spiking network model). Le modèle consiste en une population excitatrice et inhibitrice de neurones "leaky integrate-and-fire" dont la connectivité est identique à celle de nos modèles de taux de décharge ("firing rate neurons"). Les non-linéarités des FTs de chaque neurone sont implémentées en ajustant les propriétés intrinsèques de la membrane neuronal ainsi que le bruit de fond. Nous démontrons que le mécanisme de saturation effective ainsi que la terminaison de l'état de bosse par une excitation globale s'appliquent toujours. Nous étudions en détail la dynamique pendant les transitions entre les états d'activité persistante et spontanée. Nous constatons que pour des durées d'input de l'ordre des constantes de temps synaptiques et membranaires, un motif spatio-temporel oscillatoire apparaît. Il est intéressant de noter qu'en raison de ces oscillations, seules certaines valeurs d'intensité d'input permettent de mettre fin à l'état persistant. Dans l'ensemble, ce travail fournit un mécanisme dans lequel les FTs excitateurs et inhibiteurs interagissent, donnant lieu à une multistabilité dans laquelle la terminaison de l'état sélectif persistant peut être médiée par une excitation globale.

Ensuite, dans le chapter 4, nous exploitons le précédent modèle de réseau de neurones qui présentent décharges décrivant la dynamique du CPF au cours d'une tâche MT pour étudier comment les dysfonctionnements du réseau pourraient conduire à une dynamique aberrante qui pourrait être à l'origine des déficiences cognitives typiquement rencontrées dans les maladies mentales. Le CPF semble être l'une des régions les plus touchées dans les maladies mentales telles que la schizophrénie, entre autres. Les

troubles de la pensée, y compris les hallucinations, et les symptômes cognitifs tels que l'altération de l'MT et de l'apprentissage décrits chez les sujets atteints de schizophrénie sont fortement associés à un dysfonctionnement de le CPF. Le CPF, en synchronisation avec les aires sensorimotrices et les aires sous-corticales, est chargée d'orchestrer le cycle perception-action en reliant les informations sensorielles passées et le comportement prospectif. Ainsi, des altérations de la fonction CPF pourraient potentiellement conduire à une perception anormale de la réalité conduisant à des hallucinations, ainsi qu'à des altérations de l'MT. L'une des principales théories sur l'étiologie de la schizophrénie tourne autour d'une hypofonction des récepteurs NMDA (NMDAR)(Gao and Snyder, 2013). Les principales preuves à l'appui de cette théorie proviennent d'études dans lesquelles des antagonistes des NMDAR (comme la kétamine) pouvaient reproduire les symptômes des patients atteints de schizophrénie chez des patients sains (Krystal et al., 1999). En outre, les médicaments améliorant la fonction NMDAR atténuent les symptômes négatifs et cognitifs. Par ailleurs, le blocage persistant de la fonction des NMDAR chez l'animal développe des caractéristiques pathologiques de la schizophrénie telles qu'une altération des tâches MT (Aura and Riekkinen Jr, 1999; Driesen et al., 2013; Coyle, 2012). De plus, des études de neuro-imagerie et d'anatomie de cerveaux schizophrènes ont décrit une diminution du volume du CPF ainsi que des corps cellulaires plus petits et une perte d'épines dendritiques dans les neurones pyramidaux de la couche III du DLPFC (Hoftman et al., 2017). On pense que la couche III du DLPFC est l'endroit où résident les microcircuits du CPF. Ces circuits sont composés de réseaux de cellules pyramidales excitatrices fortement connectées de manière récurrent et s'excitant mutuellement via des récepteurs AMPA et NMDA sur les épines dendritiques. Ce circuit entraîne une activité persistante qui est façonnée par l'inhibition latérale des cellules GABAergiques, créant ainsi les propriétés de sélectivité spatial observées dans les expériences de mémoire de travail (Arnsten, 2011). L'hypothèse de l'hypofonctionnement des NMDAR suggère que le neurotransmetteur glutamate n'active plus les récepteurs de NMDA qui sont responsables des courants excitateurs postsynaptiques (CEPS) lents. Au lieu de cela, les récepteurs AMPA qui ont une

constante de temps de décroissance plus courte (environ 2 à 5 ms), seraient la principale contribution aux CEPS (Rubio et al., 2012). Afin d'explorer comment les différentes contributions des récepteurs NMDA pourraient modifier la dynamique du CPF, nous avons réalisé une étude exploratoire analysant le rôle des constantes de temps synaptiques sur la dynamique spatio-temporelle et sur la stabilité des états stables. Nous avons constaté que, pour des valeurs physiologiques de la constante de temps de décroissance synaptique inhibitrice (environ 5-10 ms pour le récepteur $GABA_A$), la constante de temps de décroissance synaptique excitatrice doit être plus lente afin d'éviter la synchronie et les motifs d'activité oscillatoire. Comme déjà remarqué par Wang (1999), ces résultats suggèrent que la dynamique lente du NMDA joue un rôle important dans la stabilisation de la décharge neuronal irrégulier présent à la fois dans l'état homogène et dans l'état de bosse. Consécutivement, dans une deuxième partie, nous explorons comment la taille du réseau du modèle affecte la stabilité des états stables du système. Nous avons constaté qu'une réduction du nombre de neurones pouvait conduire à des transitions spontanées donnant lieu à l'apparition de bosses d'activité en l'absence de stimulations externes. Ces résultats pourraient suggérer un mécanisme pour l'origine des hallucinations dans les circuits CPF.

Finalement, dans le chapter 5 nous présentons un article publié en 2020 dans la revue Frontiers in psychiatry où nous décrivons le cas d'un patient de 50 ans atteint de schizophrénie qui a subi un protocole de stimulation cérébelleuse non invasive. Des données issues d'évaluations cliniques et psychométriques incluant la mémoire de travail, la mémoire verbale à long terme, les fonctions exécutives et d'attention ont été recueillies avant et après la stimulation, ainsi que des données issues d'un protocole de conditionnement du clignement des yeux dépendant du cervelet.

Bien que nous n'ayons pas rapporté de changements dans les symptômes positifs ou négatifs de la schizophrénie avant et après la stimulation, nous avons constaté une amélioration globale des mesures psychométriques après la stimulation. Nous avons également constaté une amélioration des performances en matière d'attention

sélectionnée, de mémoire à court et long terme, de mémoire de travail et d'inhibition de la réponse, domaines cognitifs connus pour être altérés chez les patients atteints de schizophrénie. On a constaté une nette amélioration du test de conditionnement de clignement d'œil (eye-blink conditionning ou EBC) après la stimulation. En l'absence de données sur des sujets sains dans nos conditions, nous ne pouvons exclure un éventuel effet de retest ("saving") dans cette amélioration (Hoffland et al., 2012). Ceci est cependant peu probable étant donné que dans nos données : 1) aucune eyeblink conditioning (EBC) nette n'a pu être observée à la fin de la session de pré-stimulation, et 2) le conditionnement est apparu progressivement au cours de la session EBC de post-stimulation, en partant d'une absence de réponse prédictive (et ne montrant donc aucune sauvegarde). En outre, et contrairement aux sujets témoins, il a été démontré que les patients atteints de schizophrénie n'améliorent pas leurs performances au cours de sessions EBC consécutives (Bolbecker et al., 2009). Ces éléments de preuve soutiennent donc l'idée que l'amélioration de l'EBC après la stimulation était due à la stimulation elle-même et non à une quelconque rétention de la première session d'EBC. Sur la base de l'abondante littérature, tant chez l'homme que chez l'animal (Thompson and Steinmetz, 2009; Mauk et al., 2014; Timmann et al., 2010; Freeman and Steinmetz, 2011), nos données indiquent que la fonction cérébelleuse du patient était altérée de manière basale, comme cela a été décrit précédemment dans la schizophrénie (Kent et al., 2015). Notre évaluation est cohérente avec plusieurs études qui ont rapporté un effet de la stimulation cérébelleuse non invasive sur l'apprentissage associatif mesuré par EBC. (Zuchowski et al., 2014; Monaco et al., 2014). Plus important encore, elle met en évidence la stimulation du cervelet comme un outil puissant pour améliorer significativement la fonction cérébelleuse dans la schizophrénie. L'EBC s'est révélé être une méthode pertinente pour étudier le dysfonctionnement cérébelleux dans les troubles neuropsychiatriques. Démêler les aspects motivationnels des déficits cognitifs peut s'avérer difficile chez les patients atteints de schizophrénie. Cet aspect est pourtant important car les déficits de motivation communément présents chez les patients schizophrènes peuvent biaiser les tests neuropsychologiques classiques. EBC ne

nécessite pas la participation active du sujet. Dans notre paradigme, le patient regardait un film muet pendant l'expérience ; chez les nouveau-nés, l'EBC peut même être réalisée pendant le sommeil (Fifer et al., 2010; Tarullo et al., 2016). Ainsi, il est peu probable que les résultats des EBC soient liés à des déficits motivationnels chez les patients atteints de schizophrénie.

(Brady Jr et al., 2019) a rapporté dans une population de patients atteints de schizophrénie, une amélioration des symptômes négatifs après une stimulation magnétique transcraneal liée à la connectivité dorso-latérale du CPF au cervelet. Cependant, les auteurs n'ont pas étudié l'effet de la stimulation cérébelleuse sur les symptômes cognitifs.

Chez les sujets sains, il existe des preuves que la stimulation cérébelleuse non invasive peut moduler la mémoire de travail, le contrôle moteur, l'apprentissage et le traitement des émotions (Ferrucci et al., 2015). Ces résultats sont en accord avec nos résultats où la stimulation cérébelleuse non invasive a eu un effet sur la mémoire verbale, la fonction exécutive et la fonction d'attention.

Nos résultats soutiennent plusieurs points forts. À notre connaissance, ce cas est le premier à rapporter l'effet de la stimulation électrique du cervelet sur la cognition (y compris l'apprentissage associatif mesuré avec le EBC) dans la schizophrénie.

En conclusion, nos résultats suggèrent que la stimulation du cervelet peut avoir un impact sur les troubles cognitifs chez les patients atteints de schizophrénie. Nous suggérons que le EBC, connu comme une méthode pertinente pour étudier le dysfonctionnement cérébelleux dans les troubles neuropsychiatriques, pourrait être utilisé pour évaluer l'impact de la stimulation sur le cervelet chez les patients atteints de schizophrénie.

Science is what you know. Philosophy is what you don't know

— Bertrand Russell

Contents

List of Figures

List of Tables

Acronyms

AHRS Auditory Hallucination Rating Scale.

ALM anterolateral motor cortex.

AMPAR AMPA receptor.

BOA basin of attraction.

CR conditioned response.

CS conditioned stimulus.

DA delayed alternation.

DA dopamine.

dlPFC dorsolateral prefrontal cortex.

DN dentate nuclei.

DR delayed response.

EBC eyeblink conditioning.

EPSCs excitatory postsynaptic currents.

fMRI functional magnetic resonance imaging.

FPs fixed points.

FR firing rate.

GABA γ-aminobutyric acid.

IPSCs inhibitory postsynaptic currents.

IR infrared.

MD mediodorsal nucleus of the thalamus.

NCP non clinical psychosis.

NMDAR N-methyl-d-aspartate receptor.

OB oscillatory bump.

ODR oculomotor delayed response.

OU oscillatory uniform.

PANSS Positive And Negative Symptom Scale.

PCP phenylciclidine.

PD preferred direction.

PFC Prefrontal Cortex.

PN pontine nuclei.

PPT pedunculo-pontine nuclei.

PSCs postsynaptic currents.

rTMS repetitive transcranial magnetic stimulation.

RTN reticulo-tegmental nucleus.

SA sequential activation.

SW standing wave.

t2t time to transition.

tDCS transcranial direct current stimulation.

TF neuron transfer function.

TMS transcranial magnetic stimulation.

TW travelling wave.

UB uniform bump.

UR unconditioned response.

US unconditioned stimulus.

V1 primary visual cortex.

VTA ventral tegmental area.

WM Working Memory.

Part I

PROLOGUE

1 Introduction

If we wanted to simplify at maximum the function of the brain we could picture it as a black box that receives some inputs, does some transformations of the inputs and gives an output. In order to fully understand this black box we should consider all the inputs and know all the transformation steps from inputs to outputs. This becomes a very complex task if the number of inputs is very large and even more if the transformations depend on the internal state of the black box which in addition can change because of the external inputs. The first problem can be relatively overcome by reducing the number of inputs in a controlled experiment. Following this idea, brain research has made important advances in understanding sensory processing and movement execution, but less is known as we go deeper in the input processing chain towards action and try to understand high order cognitive functions such as reasoning, memory, decision-making, emotions, etc. Paradoxically, these functions are the reason why we start to ask complex questions and have a need to find the answers. The ambition to comprehend reality and existence is probably one of the most characteristic features of the human being. Besides many other features that are common to other species, humans have developed an intricate infrastructure to address these ontological questions. Language, writing, mathematics and other tools have served as a framework to share knowledge, ideas and reasonings transcending the spatiotemporal distance among other humans allowing a sort of collective processing and the creation of novel knowledge with the final purpose of satisfying our need of understanding. Not by chance, the genesis of this framework is a direct consequence of brain evolution. The development of neocortical areas has led to complex behavior consisting in the realization of a considerable number of

planned actions in order to reach a certain goal. A central role of the temporal coordination of these actions is attributed to the PFC. Metaphorically, we humans have developed a framework to understand complex things that plays a similar role as the PFC does by processing novel information and integrating it by contrasting with past experience to come up with a new behavior (or thought) that is intended to achieve a certain goal.

In fact, the evolution of the brain reveals a so-called hierarchy between subcortical and cortical structures. While simple movements and innate actions are mostly controlled by subcortical structures, conscious and complex behavior is attributed to cortical areas which developed later in the evolutionary timeline, being the premotor and prefrontal cortices the latest evolved areas, thus, at the top of this hierarchy. A simple example to understand the difference between simple and complex behavior and the brain structures that are involved could be this: it's a stormy day and you have hypersensibility to loud noises such as thunder, the storm is quite close so whenever you see a bolt you cover your ears to alleviate the rumble, in this case in which the bolt and the thunder occur close in time your cerebellum allow you to associate both events and cover your ears at the right time. But now, the storm is moving away so there is a larger time gap between the bolt and the thunder, in this case the cerebellum alone is not sufficient to pair both events because of the temporal delay, here is where the coordination with higher cortical areas comes into place by processing and coordinating in time a coherent response (Weiss and Disterhoft, 1996). In order to orchestrate an action that is separated in time a few seconds from the initial stimulus we have to retain in our brain the stimulus, this capacity is known as WM and this process is largely attributed to the PFC. Many mechanisms of WM have been proposed but the neural mechanisms underlying the formation, the maintenance and extinction of this memory are still a matter of debate. Besides, it is well known that in mental illnesses such Schizophrenia, cognitive function including WM is impaired, however it is poorly understood how the PFC in coordination with subcortical areas is implicated in cognitive dysfunction and what are the network mechanisms underlying the main symptoms in Schizophrenia.

With the aim to shed some light on the network mechanisms that support WM we devise a biophysical theoretical framework that allows us to study the role of the

neural nonlinearities in the formation and extinction of memories. In particular, we present novel network mechanisms for memory extinction based on biological evidences. Thereafter, in a separated work, we study how changes in the PFC network leads to aberrant working memory dynamics that could explain some symptoms in mental illness such as schizophrenia. Finally, in the last work, we present a novel treatment for schizophrenia based on cerebellar stimulation.

The thesis is organised in three parts: a prologue consisting in an overview (chapter 1) and a literature review regarding the role pf the PFC in WM and its implication in Schizophrenia together with the cerebellum (chapter 2); an original research which is subdivided in chapter 3 , chapter 4 and chapter 5, where we present the results of three projects separately, each one with a specific introduction, results, discussion and conclusion; and a epilogue consisting in the general discussion and conclusions in chapter 6.

2 Literature Review

2.1 Anatomy and connections of the Prefrontal Cortex

The prefrontal cortex is located in the anterior region of the mammalian neocortex (Figure 2.1). In mammals, it is anatomically defined as the part of the cerebral cortex that receives projections from the mediodorsal nucleus of the thalamus (MD) (Fuster, 2015). The size of the PFC increases along the evolution of the species and its relative size is directly proportional to the cognition complexity of the animal. In humans the relative size reaches a maximum configuring almost one third of the entire neocortex (Fuster, 2015). In newborns the prefrontal area is not fully developed and it experiences a proliferation of dendrites and formation of new synapses during the firsts years of life (Koenderink et al.,

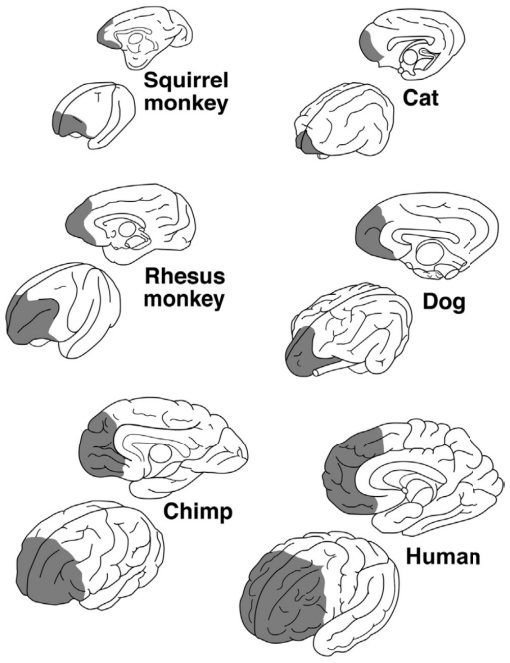

Figure 2.1. The prefrontal cortex in six different mammalian species (shaded in gray). From Fuster (2015)

1994; Bourgeois et al., 1994). Furthermore it is the last area to undergo myelination reaching its maximum in adulthood (Conel, 1939; Brody et al., 1987).

The homologues of the PFC are not clear in small mammals regarding its cytoarchitecture. In dogs and cats granular areas (with large body density in layer IV) resemble agranular areas in primates, and rodents have agranular areas in the frontal region (Passingham and Wise, 2012). These discrepancies among species in the cytoarchitecture (cellular composition) and in the number of connections with other brain areas, makes it difficult to define a PFC homologue among small mammals(Markowitsch and Pritzel, 1977; Preuss, 1995; Passingham and Wise, 2012). Even so, experiments support that lesions in frontal areas of these small mammals results in similar functional and behavioral impairments as seen in monkeys (Larsen and Divac, 1978). In addition, later studies comparing structural and functional data have found homologues of certain areas of the primate PFC in dogs (Markow-Rajkowska and Kosmal, 1987) as well as rats (see review in Uylings et al. (2003)) and even in mice (Laubach et al., 2018). These evidences point out that the cytoarchitecture alone is not sufficient for determining the prefrontal function being the anatomical relationship with the MD to be a more universal feature among mammals (Fuster, 2015). In this line of thinking, maybe the one most relevant characteristic that determines largely the function of the PFC is the high degree of reciprocal connections with other brain areas. The principal subcortical afferents (inputs) come from the MD reaching all the areas conforming the PFC. Inputs from subcortical areas such as cerebellum, globus pallidus and hippocampus are relayed through the thalamus to the lateral prefrontal areas (Middleton and Strick, 1994). But not all subcortical inputs are mediated by the thalamus; the hippocampus, the pons, the hypothalamus, the amygdala and the cerebellum reach the PFC by direct afferent connections Figure 2.2. Although the nature of these inputs is not fully understood, some inferences can be drawn from the function of the structures that project to the PFC. Inputs from the hippocampus could be relevant for motor learning and memory; inputs from hypothalamus and amygdala are related to the internal state and the drives of the organism; and inputs from cerebellum are thought to be involved in the temporal organization of motor actions (Fuster, 2015). Besides the subcortical afferents, the PFC also receives important cortical inputs from both sensory areas indicating its role integrating different sensory modalities

(Figure 2.2).

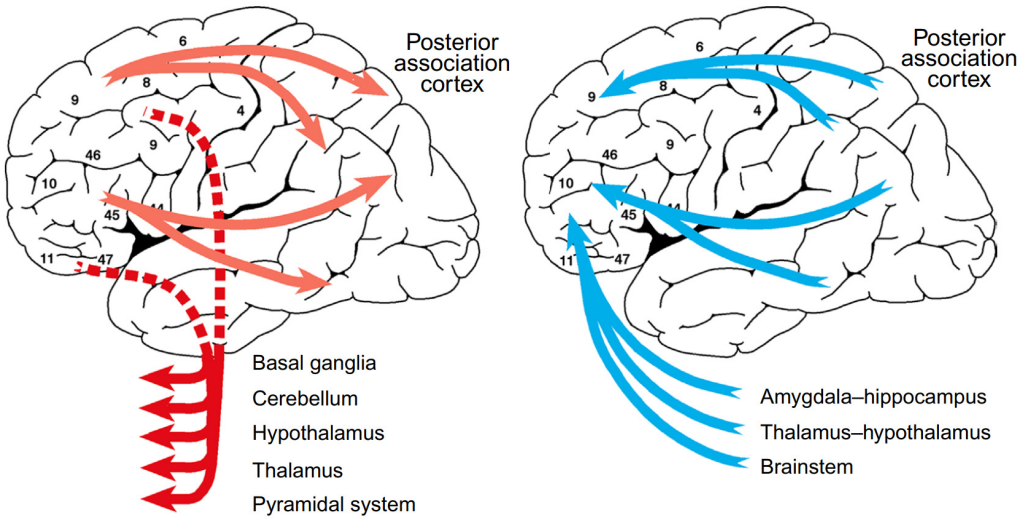

Figure 2.2. Diagram of the efferents (*left*) and afferents (*right*) of the prefrontal cortex relevant in the perception-action cycle. From Fuster (2015)

Primary sensory areas for the three main modalities (vision, audition and somatosensory) projects to an adjacent associative area and to a specific region of PFC that also send reciprocal connection. The next area in the processing line again sends projections to the next one and to the prefrontal Figure 2.3. Other corticocortical connections are the reciprocal connections between the prefrontal and the premotor cortices and this with the primary motor cortex. This frontal motor hierarchy in which information flows from prefrontal to primary motor cortex is symmetrical to the perceptual one in which information flows from primary to associative areas. The reciprocity of connections of PFC and its afferents is a late-motive feature with the exception of the basal ganglia and pontine nuclei which receive inputs from PFC but do not reciprocate.

2.2 Working Memory in the Prefrontal cortex

The first reported historical evidence of the anatomical substrates of WM came from Eduard Hitzig (1874) and later David Ferrier (1886). Their studies revealed

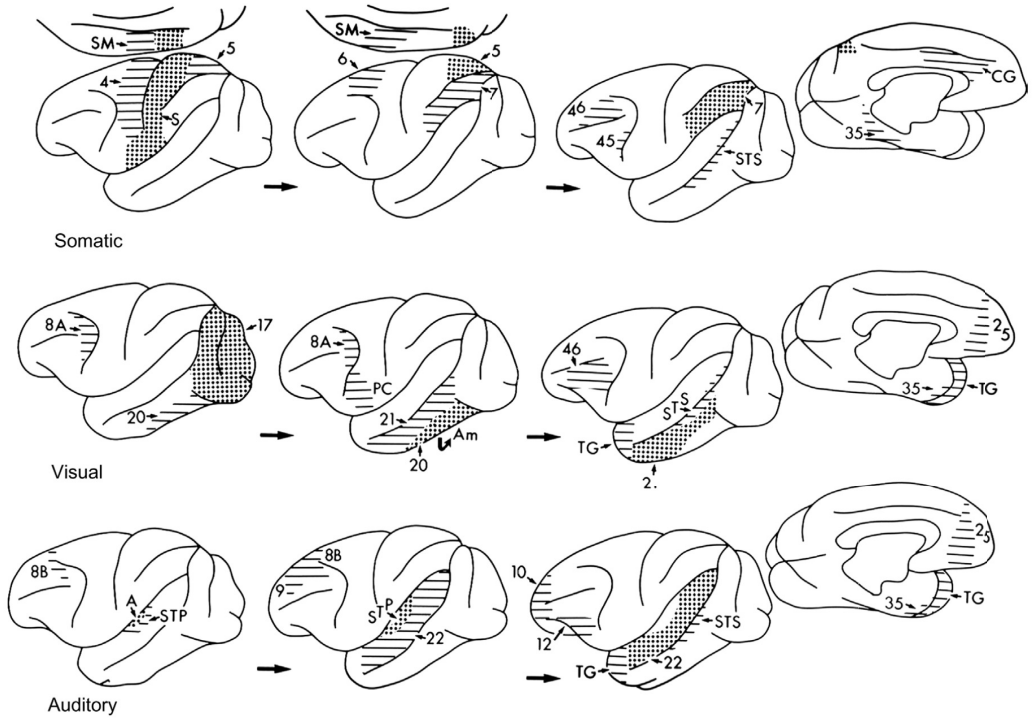

Figure 2.3. Stepwise projection of connections originating in the somatic, visual, and auditory areas of the monkey's cortex. Each step is shown by a dotted pattern, and the succeeding regions of termination by a pattern of horizontal lines. Note that most areas project to others nearby and to the prefrontal cortex. Abbreviations: A, auditory cortex; CG, cingulate gyrus; S, somatosensory cortex; SM, supplementary motor cortex; STP, superior temporal plane; STS, superior temporal sulcus; TG, temporal polar cortex. From Fuster (2015), adapted by Amaral (1987) from Jones and Powell (1970)

that the prefrontal region of the cortex is indispensable to conduct tasks in which WM is required. In its work "The functions of the Brain" published in 1886, Ferrier described how dogs after the ablation of the prefrontal brain region were not able to remember the location of food they have just seen. Later on, Caryle Jacobsen (1935; 1936) explored the implications of the PFC in a more systematized way by training monkeys to perform a delayed task in which the animal observes how a piece of food is located in one of the two cups and few seconds later the animal have to choose between the two cups. If the answer is correct the animal will be rewarded with the piece of food. This task was conducted with several delays and

at different stages: preoperatory (control), unilateral and bilateral lesions. While the unilateral lesion did not differ from control condition, the bilateral extirpation of the PFC dramatically impairs the delayed task involving WM. Furthermore, the lesions on other cortical areas (parietal association area, premotor and temporal lobe) did not impair the ability to perform the task. Subsequent works in monkeys have confirmed and confined the specific PFC areas relevant to WM tasks in which the spatial location of an item has to be remembered. The region is known as dorsolateral prefrontal cortex (dlPFC) which corresponds to anatomical areas 46, 9/46 and 9 (surroundings of principal sulcus in primates) (Figure 2.4) (Blum, 1952; Pandya et al., 2015; Levy and Goldman-Rakic, 1999) .

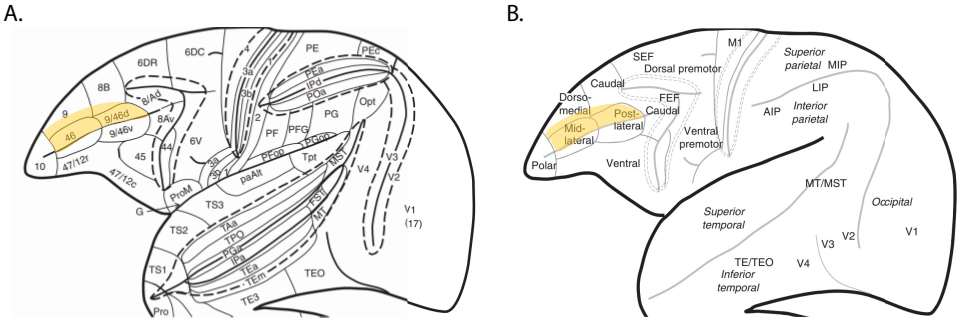

Figure 2.4. A. Architectonic map of the macaque monkey cortex by Petrides and Pandya (2007) **B.** Regional nomenclature used in Passingham and Wise (2012) for macaque monkeys. The dlPFC location is shaded in yellow, right above the *Principal Sulcus*. Adapted from Passingham and Wise (2012)

First electrophysiological recordings of activity in single neurons in dlPFC during WM tasks in monkeys were conducted by Kubota and Niki 1971 and Fuster and Alexander 1971. Kubota and Niki used a delayed alternation (DA) task in which the monkey has to alternately press two levers after a delay. Fuster and Alexander used a delayed response (DR) task in which the animal has to recall the location (right or left food wells) of the previously presented food after a time delay of a few seconds (Figure 2.5A). Those studies described a rich neural activity in which some neurons increased their firing when information was presented, others responded strongly to the motor onset and, the most relevant finding, others were active all along the delay period between the sensory cue and the response action (Figure 2.5B). This sustained activity triggered by the sensory cue can cover

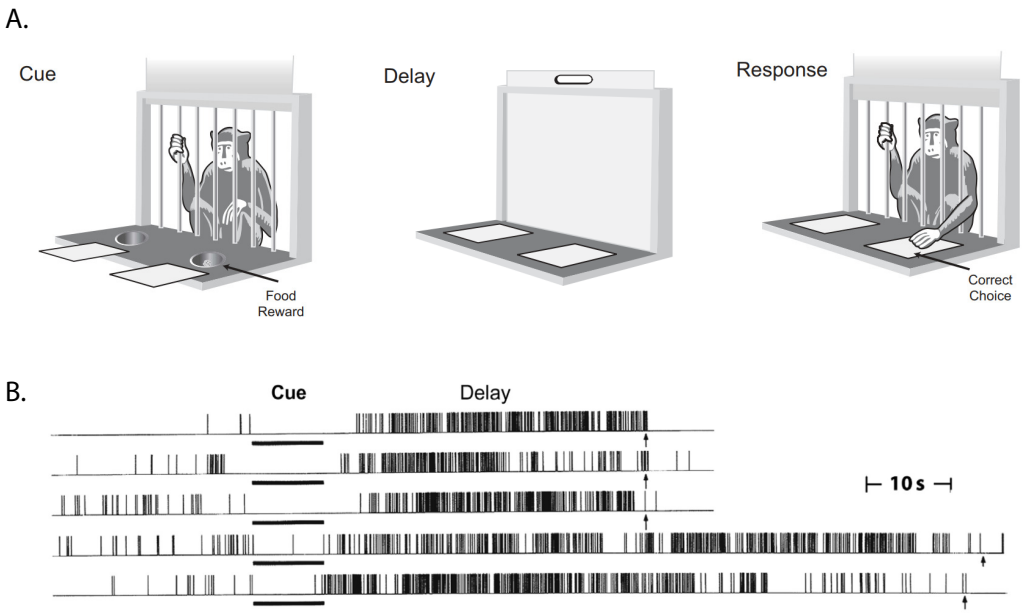

Figure 2.5. A. Schematic depictions the spatial delayed-response task. It is a test of spatial working memory and requires the monkey to remember the baited location of a food well over a delay. Adapted from Curtis and D'esposito (2004) **B.** Activity of a prefrontal unit during five delayed-response trials. In each trial, a horizontal bar marks the cue period and an arrow the end of the delay (i.e., the presentation of the choice stimuli). Note the activation of the cell during the delay: over 30s in the upper three trials, 60s in the lower two trials. Adapted from Fuster and Alexander (1971)

delays up to 1 min or longer and it recovers its previous spontaneous activity right after the response is completed. Kubota and Niki relate this activity to a delayed motor execution while Fuster and Alexander relate this sustained firing to neural correlates of short-term memory. Later on, Fuster showed that this persistent activity was only present in correct trials, but not in the error trials (Fuster, 1973). This led to the idea that the tonic persistent activity during the delay is linked to the maintenance of the spatial information that is critical to perform the task.

Few years later, Hiroaki Niki (1974a,b) trained monkeys to perform a delayed response (DR) task similar to that used by Fuster and Alexander 1971. With this experiment showed that some neurons respond differently depending on the

spatial location of the response, i.e., the neurons display directional selectivity. These findings were further confirmed later on by the seminal work of Shintaro Funahashi et al. (1989). In this case the monkey is trained to perform an oculomotor delayed response (ODR) task consisting in a screen which displays a visual cue that can appear in one of 8 possible spatial locations distributed evenly and separated 45°one form each other in a circumference. The animal has to retain the spatial location of the cue during a delay and when the fixation point in the center of the screen disappears it has to make a saccade towards the location where the cue was presented (Figure 2.6B). In these experiments they found that some neurons display a marked spatial location selectivity, in which the firing rate during the delay increases significantly for certain cue locations in the space (Figure 2.6C). This selectivity, similar to that found in the receptive fields of the visual system, but maintained in the absence of external stimulus was described as mnemonic receptive fields (Funahashi et al., 1989; Rainer et al., 1998a,b). Each selective neuron seems to always respond to the same visual location so the response curve versus cue location draws a bell-shaped tuning curve in which the angle of the maximum response is named the preferred direction (PD) (Figure 2.6D). Notice that this feature selectivity is also present in neurons in the visual cortex (Hubel and Wiesel, 1962, 1968). As a direct consequence, theoretical models of visuospatial working memory have inherited the orientation tuning theories from the visual system as we will see in the next section.

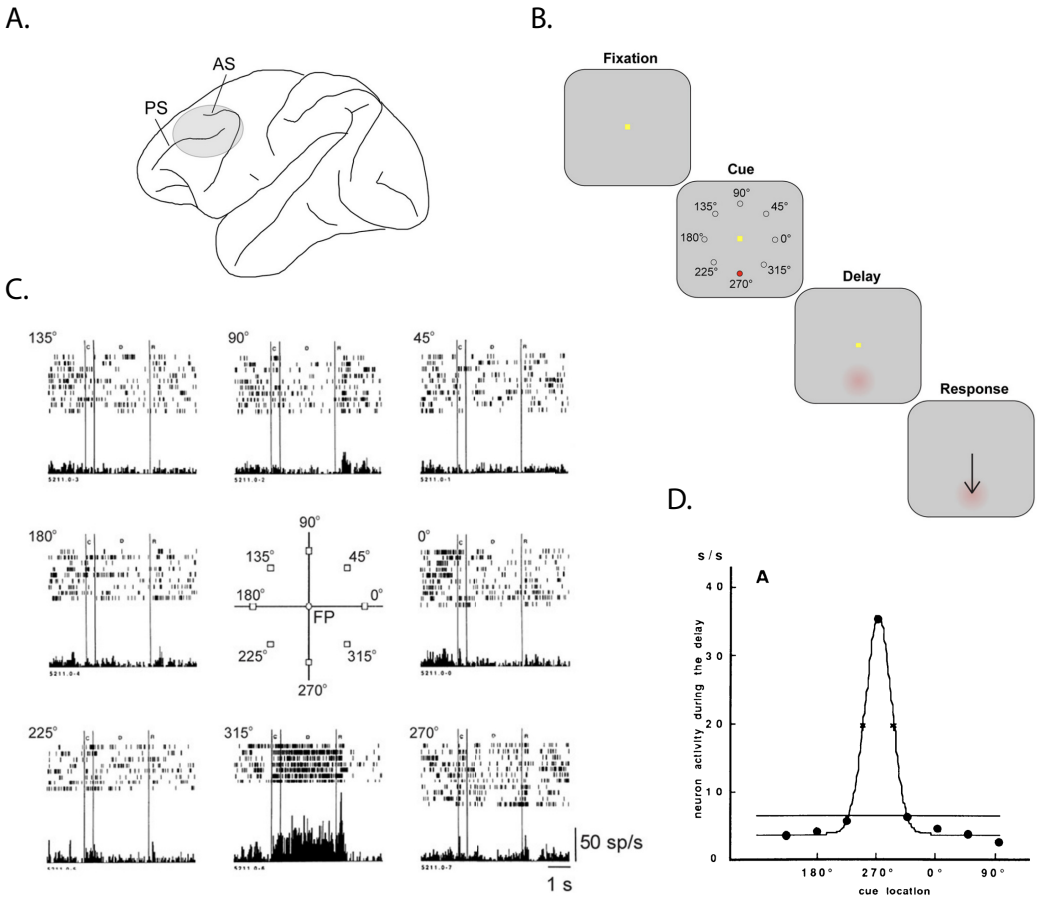

Figure 2.6. A. Recording area. Schematic diagram of the monkey brain with caudal area of the *Principal Sulcus* (PS) highlighted. AS, *Arcuate sulcus*. **B.** Sequence of events in the oculomotor delayed response (ODR) task. **C.** Directional delay period activity of a *Principal Sulcus* neuron during the ODR task. This neuron has strongly directional delay period activity, responding only when the cue had been presented at the bottom (270°) location. **D.** Directional tuning of delay period activity for neuron in **C**. Plot shows discharge rate during the delay period for 8 different cue locations, with a Gaussian fit to the data. Adapted from Constantinidis et al. (2018) and Funahashi et al. (1989)

2.3 Network mechanisms of Working Memory

Until now we have described the neural dynamics during a spatial working memory task and so far we can distinguish two main features: persistent activity and selectivity. The persistent activity is seen as a combination of the active internal representation of a sensory stimulus that is maintained when the stimulus is not present anymore and a representation of the future action (Takeda and Funahashi, 2002). On the other hand, the stimulus selectivity is the specific activation of a subgroup of neurons when the stimulus is presented in a specific location. Despite other features of this delay activity such as high spiking variability (inter and intratrial), heterogeneity of the neural activity, resistance to distractors, etc. we would like to address the main mechanisms that underlie persistence and selectivity.

2.3.1 Persistent Activity Mechanisms

The vast majority of spatial working memory models rely on mechanisms of persistent activity supported by the previous exposed literature about the increase of firing activity of some neurons during the delay period. (See however, Mongillo et al. (2008); Barak et al. (2010); Barak and Tsodyks (2014); Lundqvist et al. (2018) but also see Li et al. (2021)).

Although persistent activity is broadly accepted as a neural correlate of the working memory, the mechanisms underlying the generation and the extinction of this persistent activity remain poorly understood. In essence, models of persistent activity are based in dynamical systems with multiple stable states. The system displays at least two stable steady states or fixed points and transient changes due to an external input can cause a transition from a low activity state to a high activity state. Since this high activity state is a stable fixed point, the system will remain in this point unless an external input restores the system to the initial state. Although this is a very recurrent approach to model persistent activity several mechanisms that generates this multistability have been proposed. The main streams regarding the origin of this multistability are; (1) the intrinsic properties of the neurons (Camperi and Wang, 1998; Loewenstein and Sompolinsky, 2003), (2)

the network interactions (Amari, 1977; Hopfield, 1982) and (3) a combination of the previous two (Compte et al., 2000; Hansel and Mato, 2013). The main question that is common to all these approaches is what physiological mechanism gives rise to this multistability. Although, up to now, there is no evidence of intrinsic neuronal bistability in PFC neurons, bistability in form of plateau potential have been found in neurons from the entorhinal cortex (Egorov et al., 2002) and motoneurons (Kiehn, 1991). Despite this, the hypothesis of cellular bistability in PFC has been used in a model of WM presented by Camperi and Wang 1998 that we will describe later. On the other hand, most research has been conducted based on the network mechanism relying on strong recurrent excitation (Hebb, 1949). This mechanism is well supported by empirical data from describing high horizontal connection in the monkey's PFC in *in vitro* studies (González-Burgos et al., 2000) as well as in *in vivo* studies (Funahashi and Inoue, 2000). The basic idea of this mechanism is represented in the Hopfield model (Hopfield, 1982) for storing discrete memories in the synaptic connectivity matrix creating a discrete attractor network. In this model of associative memory, which neglects some of the basic biological properties, a finite number of activity patterns can be stored as memories by creating strong connectivity patterns within a cluster of neurons (also describing a mechanism for selectivity). These connectivity patterns could be previously established by Hebbian learning in which neurons that activate by a specific input are strongly connected. Once the connectivity is established the working memory mechanism can be described as follows: an external input reaches the network and activates a group of neurons, if a great majority of these neurons belong to a pre-established cluster i.e. are strongly connected, they will recover the memorised pattern maintaining the activity through time due to the strong recurrent connection.

The main problem residing in the persistence mechanism based in strong excitatory feedback is the blow up of the neural activity that exceeds the physiological firing rates ranging from 10-50 spikes/s. Solutions to control the firing rate were proposed by Xiao-Jing Wang (1999) by including slow and saturating synapses that stabilise persistent states at low firing rate. These types of synapses were attributed to N-methyl-d-aspartate receptor (NMDAR) channels. Although saturation of the NMDAR is debatable (see McAllister and Stevens (2000)), the blockade of NMDAR results in the abolition of the persistent

firing during the delay (Wang et al., 2013; van Vugt et al., 2020). However it is not clear whether this is a consequence of the suppression of slow synaptic currents or a reduction in the overall excitation. In fact, in van Vugt et al. (2020) the selective blockade of AMPA receptor (AMPAR), which have fast dynamics, was sufficient to cause a suppression of the delay activity. Inhibitory feedback was also a proposed solution to control the firing rate (Amit and Treves, 1989; Rubin and Sompolinsky, 1989) which also contributes to stabilise the spontaneous state with non-zero firing rate (Amit and Brunel, 1997). The principal problem of the introduction of inhibition is the appearance of oscillations due to the excitatory-inhibitory loop if the recurrent excitation is not slow enough (Van Vreeswijk et al., 1994; Hansel et al., 1995; Wang and Buzsáki, 1996). A possible solution to this problem will be to use a slower excitation (as in Wang (1999)) but, the asynchronous activity can also be restored provided that inhibitory to inhibitory interactions are strong enough (Hansel and Mato, 2001). A recurrent mechanism that appears in many works is the use of a non-linear synaptic or neuronal input-output transfer function in order to obtain bistability (Amit and Tsodyks, 1991; Amit and Brunel, 1995; Durstewitz et al., 1999; Wang, 1999; Hansel and Mato, 2001). In these works they describe the dynamics of networks of integrate and fire neurons with homogeneous connectivity and then, using mean-field analysis, they obtain the self-consistent equation governing the population mean firing rate (FR), ν. The steady states equation is of the form $\nu = \phi(\nu)$, where $\phi(x)$ is the TF. Using a saturating TF such a sigmoid the system could have up to three possible solutions for a given value of external input I applied ($\nu = \phi(\nu + I)$), two of them stable and one unstable. From the two stable states, one displays a low firing rate (down state) and the other high firing rate (up state). In this configuration, a transient input can induce a transition between these two bistable states (Figure 2.7).

2.3.2 Selectivity Mechanisms

As previously seen in the Hopfield model, the connectivity is prestablished by Hebbian learning in which neurons that coactivate are strongly connected. These connectivity patterns, in which neurons are organised in clusters, originate a finite

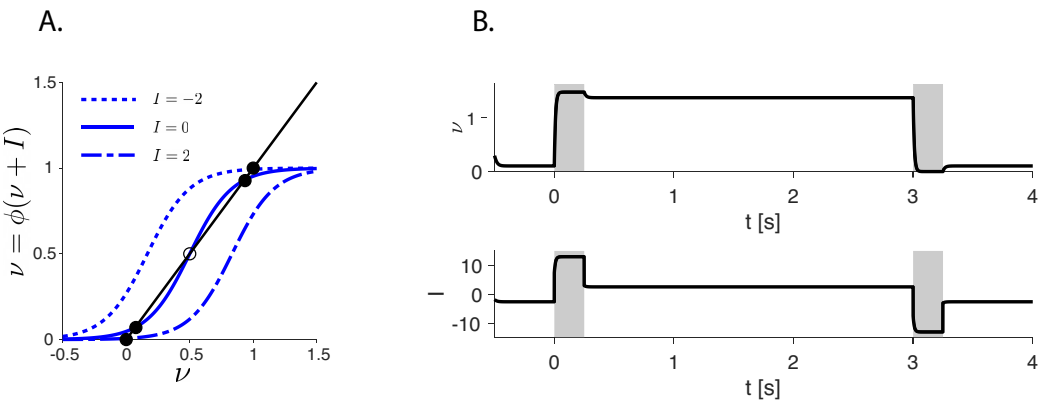

Figure 2.7. Bistability in a network of rate neurons A. Graphical solutions of the equation $\nu = \phi(\nu + I)$ for different input values with ϕ as a sigmoid. Solid black dots represents stable fixed points and empty black dots unstable fixed points. **B.** Example of switching between *down* and *up* stable states by a transient input. Top: average population firing rate. Bottom: External input to the network.

number of discrete attractor states (Figure 2.8 A). This selectivity mechanism has been widely exploited in associative memory models of spiking neurons in which the item to store is a discrete variable (Mongillo et al., 2003; Amit and Mongillo, 2003; Brunel, 2003). A possible drawback of this elegant mechanism is that the need of a pre-existing connectivity is not consistent with the capacity to maintain in the memory novel information. Furthermore, the number of memories is finite and don't allow to store continuous variables such as space (Funahashi et al., 1989), stimulus frequency (Romo et al., 1999) or visual motion (Zaksas and Pasternak, 2006). These limitations can be overcome by using continuous attractor models in which the network connectivity leads to a line of attractor states (Figure 2.8 B). The foundation of the attractor theory was developed by Wilson and Cowan (1973) and Shun-Ichi Amari (1977) in recurrent networks of firing rate neurons. The main idea behind continuous attractor models is the symmetry of the connectivity among neurons and the implementation of a connectivity function that decays as the spatial distance between two neurons increases (but also see Darshan and Rivkind (2021)).

In the specific case of the visuospatial WM the most relevant selectivity mechanism is inherited from the theory of orientation selectivity in the primary

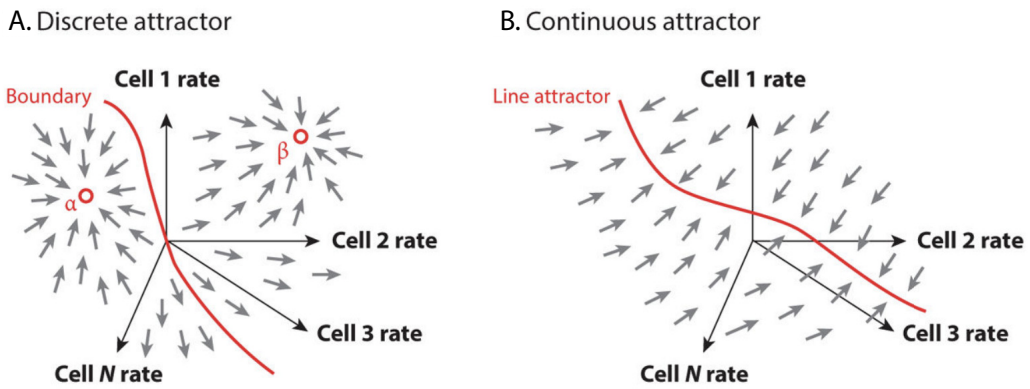

Figure 2.8. Network models of persistent representations. A. Discrete attractor models. A population of neurons is described by their firing rates (axes of the diagram). The network dynamics cause movement within this space: At each point, the small arrows indicate the direction in which the population activities move. Here, all direction arrows point toward either point α or point β, and so the network activity patterns will evolve toward one of these two activity patterns. Which of these patterns gets generated depends on whether the inputs push the network to the left of the marked boundary or the right. This boundary is known as a separatrix. **B.** Continuous attractor models are similar to the model in panel **A**, but now the direction arrows all point toward a continuous line. The network dynamics cause the activity patterns to evolve to points (patterns) on the marked line.

visual cortex (V1) from Rani Ben-Yishai et al. (1995). In this model, selectivity arises from the spatial connectivity in which neurons with similar preferred orientations are strongly connected. Neurons are spatially distributed on a ring and each neuron is labeled with a preferred orientation (PO, θ) which represents the orientation of the stimulus at which the response is maximal. In the limit of large network, the mean activity, $m(\theta, t)$, is described by the following dynamics:

$$\tau \frac{dm(\theta, t)}{dt} = -m(\theta, t) + g[h(\theta, t)] \tag{2.1}$$

where $g(x)$ is the TF that can be taken as a threshold linear function and $h(\theta, t)$ is the total recurrent input from the network described as follows:

$$h(\theta, t) = \frac{1}{\pi} \int_{-\pi/2}^{+\pi/2} J(\theta - \xi) m(\xi, t) d\xi \tag{2.2}$$

the interaction between the postsynaptic neuron θ and the presynaptic neuron ξ is described by a "Mexican hat" connectivity described as $J(\theta - \xi) = J_0 + J_1 \cos[2(\theta - \xi)]$ representing local excitation and distal inhibition (Figure 2.9A).

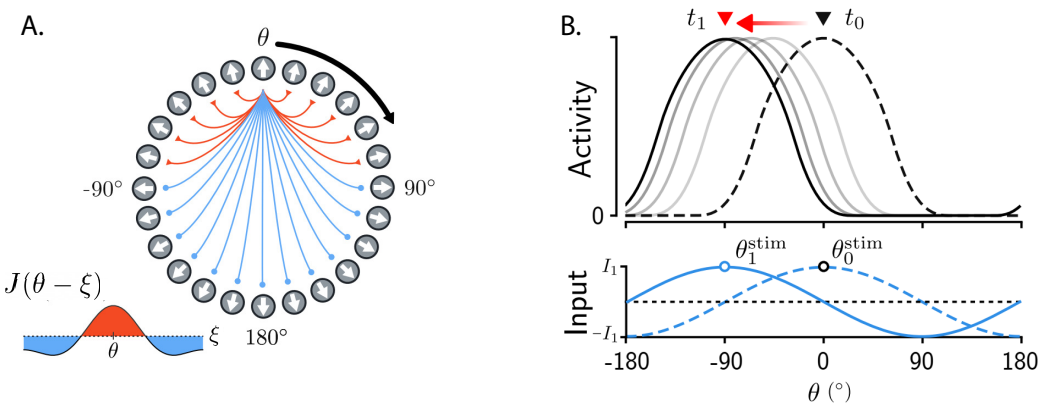

Figure 2.9. A. Ring network with Mexican-hat connectivity, i.e, strong local excitation (*red*) and broader inhibitory coupling (*blue*). **B.** Network activity for neurons arranged according to their position in the ring (**A**). Due to strong recurrent connectivity, a bump of activity (black dotted line) emerges in this network at a position determined by the external input (blue dotted line centered at θ_0). If new tuned input centered at -90°(θ_1^{stim}) is applied, the bump of activity drifts towards a new position. Adapted from Esnaola-Acebes et al. (2021)

Under a homogeneous excitatory input I, the fixed points (FPs) of the network are represented by the solutions of the following equation:

$$m(\theta) = g\left(I + \frac{1}{\pi}\int_{-\pi/2}^{+\pi/2}[J_0 + J_1\cos(2(\theta - \xi))]m(\xi)d\xi\right) \tag{2.3}$$

It represents the profile of activity of the network relative to the stimulus orientation. It also represents the tuning curve of a neuron centered at its PO. There is always a homogeneous solution $m(\theta) = m_0$ which is consistent with the fact that if the input is homogeneous the orientation tuning disappears. An inhomogeneous solution exists depending on the angular modulation of the connectivity J_1. This solution will be stable as far as the slope of $g(x)$ is larger than $2/J_1$. This inhomogeneous solution representing spontaneous

generation of orientation tuning will display a profile of activity with the shape of a "bump" centered at a certain PO (Figure 2.9 B). Due to the spatial symmetry, also exist shifted solutions of the form $m(\theta - \phi)$ describing a line of attractor states parametrised by the angle ϕ which indicates the peak of the bump activity profile. In fact, in a more realistic case, the location of the bump is not arbitrary but is determined by the orientation of the external input. Taking the external input I to be slightly tuned around ϕ (heterogeneous), neurons with PO close to ϕ will amplify the tuned input creating a bump of activity centered at ϕ. This behavior leads to the representation of the stimulus location by the network activity (Figure 2.9B). However in this network, when the stimulus is withdrawn the selective activity encoding for the stimulus is replaced by the homogeneous state, that means there is no persistence.

2.3.3 Bump attractor Models of Visuospatial Working Memory

In order to have persistence together with selectivity, the inhomogeneous solution (line of attractors) must coexist with an homogeneous solution in which all the neurons display similar firing rates. This was presented in the previously cited work of Amari (1977) studying the dynamics of pattern formation for networks dominated by lateral inhibition. The neural field dynamics of the single population model is described by the following equation:

$$\tau \frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int w(x-y) f[u(y,t)] dy + I \tag{2.4}$$

where u is considered as the total input to a neuron, $f[x]$ is the TF and $w(x)$ is the weighting function describing the connectivity between neurons. Notice the similarities of the dynamics equation with the one presented in Ben-Yishai et al. (1995). In this model, $w(x)$ also has a "Mexican hat" shape describing local excitation and lateral inhibition. The main difference relies on the TF which is taken to be the Heaviside function, this allows the network to have bistability between a bump state and a homogeneous state.

In a model presented by Camperi and Wang (1998), the network bistability is

achieved by using bistable neurons. This persistent mechanism relies on intrinsic neuronal bistability in which a single neuron displays a higher activity state that coexists with a lower activity state. These bistable neurons were embedded in a network with a "Mexican hat" connectivity in order to have selectivity in the form of a bump of activity. The rate dynamics of the network is described as follows:

$$\tau \frac{dr(\theta, t)}{dt} = -f[r(\theta, t)] + g[I(\theta, t)] \tag{2.5}$$

as seen in Ben-Yishai et al. model the TF $g[x]$ is taken to be threshold linear but the main difference is that the firing rate of each neuron is transformed by a cubic non-linear function $f[x]$. As stated at the beginning of this section, the main drawback of this model is that there is no evidence of intrinsic neural bistability in PFC neurons. However, in the same study they address a mechanism without neuronal bistability which relies on the shape of the TF. In this case, they approximate the TF to a sigmoidal shape using a piecewise linear function representing. The network displays bistability between a homogeneous state and a bump profile of activity which crucially depends on this specific shape of the TF. Unfortunately the mechanisms for bistability regarding the shape of the TF were not well characterised. Moreover, it is worth to point that in the previous presented models (Amari, 1977; Camperi and Wang, 1998) the mechanisms for bistability relies on the saturation shape of the TF, thus the neurons around the peak of the bump activity profile are firing at its saturation rates. The typical saturation firing rates of a pyramidal neuron in monkey's PFC is around 40-60 Hz (Arsiero et al., 2007; Zaitsev et al., 2012) and this is not consistent with the average firing rates around 15-20 Hz during the delay period of a pyramidal neuron responding to a stimulus at its preferred orientation (Funahashi et al., 1989; Takeda and Funahashi, 2002).

A couple of years later Albert Compte et al. (2000) presented a visuospatial WM model of spiking neurons with a very detailed synaptic interaction. With this model they reinforce the hypothesis that slow synaptic transmission mediated by NMDAR is crucial for the stability of the persistent activity following the previous work presented in Wang (1999) (without spatial selectivity). The model consists of two populations of excitatory and inhibitory neurons modelled as

integrate and fire neurons. They used an excitatory-to-excitatory (E-E) structured connectivity in which neurons with similar PO are strongly connected and the rest of interactions (E-I, I-E and I-I) are unstructured so the strength of connection is the same (Figure 2.10B). Due to the strong local excitatory connection and the strong global inhibition the network displays a bistability between a homogeneous state with low firing rate and a bump state (Amit and Brunel, 1997; Wang, 1999). The transition between the two states can be mediated by transient inputs: explicitly the visual cue, represented by a tuned input centered at a certain orientation, cause a transition from the homogeneous state to the bump state and a subsequent homogeneous excitatory input restores the homogeneous state (Figure 2.10A). Altogether the model reproduces very well the ODR task presented in Funahashi et al. (1989).

One year later Boris Gutkin et al. (2001) propose a two population network model of conductance-based single compartment neurons (Hudking-Huxley) (Figure 2.11). In this case all the interactions are structured and described by a Gaussian function. Importantly, the E-E and E-I interactions are described by sharper Gaussian (smaller variance) while the I-E and I-I are wider (larger variance). In this model as well as in the previous one, the persistence mechanism is based on strong recurrent excitation together with a more global inhibition which is balanced with the excitation. Interestingly, and conversely to the Compte et al. (2000) model, the persistent state is maintained by fast AMPA-type recurrent synapses thus, not requiring slow synaptic dynamics from NMDAR. They claim that this is due to the use of a biophysical model with ionic channels that present a time delay from the moment at which the firing threshold is crossed to the time that takes to develop the full action potential. This delay allows recurrently connected neurons that have fast synapses to display low firing rates. It is worth pointing out that in this model the spontaneous activity is silent, maybe due to the fact that the inhibitory feedback is not strong enough (Amit and Brunel, 1997). The results of this model highlight the role of fast AMPA synapses in WM mechanisms, this is consistent with the experimental results in van Vugt et al. (2020) in which the selective blockade of AMPAR leads to a decrease of the persistent firing related to the WM task.

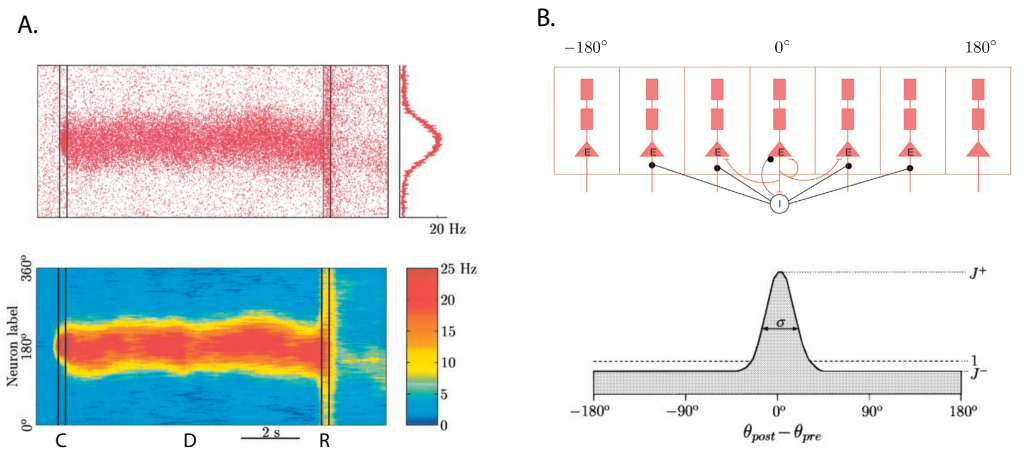

Figure 2.10. Working memory maintained by a spatially tuned network activity pattern ("bump attractor") in an oculomotor delayed-response task **A.** Top: Pyramidal neurons rastergram. The x axis represents time, while the y axis represents neuron label according to its preferred cue. A dot in the rastergram indicates a spike of a neuron whose preferred location is at y, at time x. Note the enhanced and localized neural activity that is triggered by the cue stimulus and persists during the delay period. The population firing profile, averaged over the delay period, is shown on the right. Bottom: Color-coded spatiotemporal activity pattern. C, cue period (250 ms, peak stimulus 200 pA); D, delay period (8.75 s); R, response period (250 ms, external current increase 500 pA). **B.** Top: Basic architecture of the model. Pyramidal cells (E) of similar preferred cue (spatial location, between -180°and 180°) are interconnected by local excitatory connections and send inhibitory feedback upon themselves through interneurons (I). Bottom: Structured connectivity of the model. The synaptic connection strength decreases with the difference in the preferred cues of two neurons, with strong interactions between neighboring neurons and weak interactions between more distant neurons. Adapted from Compte et al. (2000) and Wang et al. (2006)

2.3.4 Mechanisms for "Switching OFF" the Memories

Hitherto, we have reviewed some of the most relevant mechanisms for the generation of persistent activity and one can conclude that the strong recurrent excitation is the most accepted mechanism. However, the mechanisms underlying the extinction of this persistent state remain a poorly studied topic. In Camperi and Wang (1998) as well as in Wang (1999) and Hansel and Mato

(2001), they propose a switching off mechanism by a transient inhibitory input which brings the activity back to the homogeneous steady state. However this approach might be lacking of biologically relevance since some neurons, regardless of their preferred direction, experiment a global increase of the activity at the vicinity of the response period suggesting a global excitatory input (Funahashi et al., 1989, 1990, 1991). The switch off by excitation is well supported by physiological evidences that long range connectivity to the PFC is mostly excitatory (González-Burgos et al., 2000), in addition, anatomical studies have found excitatory projections from MD to supragranular layers on PFC that form excitatory synapses mostly with pyramidal neurons (Melchitzky et al., 1998). Hypotheses of the origin of the excitatory input revolve around the "go" signal that indicates the animal to respond, the motor command input from premotor areas, and the preparation for the imminent reward related to dopaminergic inputs from subcortical areas (Funahashi et al., 1991; Gutkin et al., 2001; Ott and Nieder, 2019). In the work of Gutkin et al. (2001), they directly address the question of switching off the bump state by a brief strong excitatory input. This mechanism relies on the synchronisation of the excitatory population. The simultaneous firing of all excitatory neurons sets them in a refractory period in which they do not respond to synaptic inputs. The condition for this to take place is to have fast synaptic dynamics leading to a decay of the synaptic activity in a period shorter than the refractory period of the neurons. In this scenario, the synchronisation will lead to an extinction of the spatially tuned synaptic inputs and thus the switch off to the homogeneous state. In addition, the duration of the input to mediate the extinction has to be of the order of a few milliseconds, this short duration seems physiologically unrealistic. In contrast, in Compte et al. (2000) model, the switch off was performed by a long global excitatory input. The mechanism, discussed in the article, suggests that the global excitation leads to an increase of the firing rate of the inhibitory neurons sufficiently large to suppress the excitatory firing and thus restoring the homogeneous state. A similar mechanism is addressed in Hansel and Mato (2013). Alternatively, the mechanism could be due to the saturating nature of the NMDAR present in the excitatory synapses. Unfortunately, the degree of complexity of the model makes it challenging to discern the true mechanism underlying the switch off.

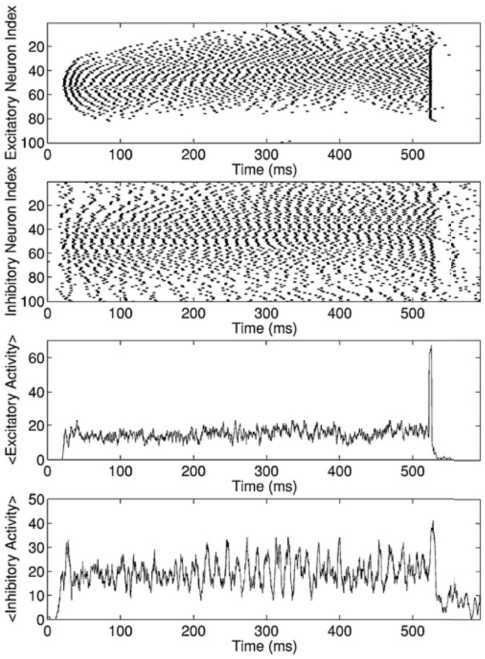

Figure 2.11. Space-time rastergrams of the excitatory and inhibitory populations. The initial stimulus is a transient focus of afferent excitation to the excitatory neurons only. The recurrent excitation supports the persistent activity, while the inhibition constrains it spatially. A transient excitatory current terminates the focus. Two bottom plots: average activity of the excitatory (upper) and inhibitory (lower) populations. From Gutkin et al. (2001).

2.4 The Role Of The Cerebellar-Prefrontal Loop In Schizophrenia

Schizophrenia is a chronic and complex mental disease with a strong hereditary component that causes a significant impact on the quality of life of around 20 million people worldwide (Spencer et al., 2018). The clinic manifestation comprises a set of symptoms, that may or may not coexist, which can be categorised in three different groups: positive symptoms such as hallucinations, delusions and abnormal motor behavior; negative symptoms such as depression, lack of motivation, social withdrawal and alogia; and cognitive symptoms such as impairment of attention, working memory, learning and executive function

(Goldman-Rakic et al., 1994; Hoff et al., 2005; Albus et al., 2006). The origin of Schizophrenia is attributed to multiple genes with small effects interacting with exogenous and environmental factors (Frankle et al., 2003). However, the large heterogeneity of symptoms attributed to Schizophrenia has led to a challenging dilemma regarding the etiology and pathophysiology of the disease (Patel et al., 2014). Thus, a better understanding of the brain network substrates involved in Schizophrenia symptoms will help to develop useful treatments.

2.4.1 Prefrontal Cortex and Schizophrenia

More than a century ago Emil Kraepelin (1919) noticed that the behavioral abnormalities in patients with Schizophrenia (named "dementia praecox" at that time) were similar to those in patients with lesions in the frontal lobe. This causal relationship between Schizophrenia and PFC has gained prominence due to a better understanding of the role of the PFC in cognitive processes together with a vast amount of evidence pointing out prefrontal abnormalities in patients with Schizophrenia. Studies on cerebral blood flow as well as measurements of local cerebral metabolism showed that patients with Schizophrenia have lower blood flow in the prefrontal regions indicating a decreased level of activity (Ingvar and Franzén, 1974; Weinberger et al., 1986; Buchsbaum and Hazlett, 1997). Although this prefrontal hypofunction is not clear in the resting state, it is reliably found during cognitive tasks engaging the PFC. For example, in a functional magnetic resonance imaging (fMRI) study in which subjects with Schizophrenia performed a sequential "n-back" working memory task, they found a strong correlation between dlPFC dysfunction and poor task performance as the memory load increases (Perlstein et al., 2001). This is consistent with the works of Sohee Park and colleagues showing that patients with Schizophrenia present an impairment in spatial working memory tasks similar to the ODR task used in monkeys (Park et al., 1992, 1995, 1999). Additionally, neuroimaging data indicate a significant volume reduction and gray matter loss in the PFC (Goldstein et al., 1999; Gur et al., 2000; Antonova et al., 2004; Cannon, 2015). Moreover, morphological studies found smaller cell bodies and dendritic spine loss in pyramidal neurons from layer III of the dlPFC

in Schizophrenia (Glantz and Lewis, 2000; Hoftman et al., 2017). This could potentially affect cognitive function, specifically WM, by a reduction in the recurrent excitation, that is critical for persistent activity as seen above in section 2.3.1. These findings could explain the reduction in the excitatory activity seen as a PFC hypofunction in the functional imaging experiments. As a consequence of the deficit in the pyramidal cell activity, several studies suggest that there is a homeostatic reduction of the feedback inhibition mediated by GABAergic interneurons (for a review see Lewis et al. (2012). Several studies found decreased levels of synaptophysin, a presynaptic terminal protein, in PFC of patients with Schizophrenia (Glantz and Lewis, 1997). The reason is not known but few evidences suggests that is due to a reduction of the afferent projections from the MD (Lewis and Lieberman, 2000). Furthermore, there are three main neurotransmitter systems causally related to Schizophrenic symptoms: glutamate, dopamine (DA) and γ-aminobutyric acid (GABA). The three of them seem to be altered in the PFC of Schizophrenic patients. On one hand, NMDA/Glutamate antagonists such as ketamine and phenylciclidine (PCP) induce psychotic symptoms and cognitive dysfunction in healthy patients (Krystal et al., 1999) and exacerbate negative and cognitive symptoms in Schizophrenic patients (Lahti et al., 1995). On the other hand, the chronic administration of these agonists produced a NMDA hypofunction that results in an increase of mesolimbic DA release which is related to positive symptoms (Kegeles et al., 2000), and a decrease in DA release in the PFC which is related with negative symptoms (Jentsch and Roth, 1999; Patel et al., 2014) (Figure 2.12). A reduction in DA in PFC could explain the cognitive symptoms since the DA is implicated in cognitive control, for instance gating the sensory input, maintaining and manipulating WM contents and relying motor commands (Ott and Nieder, 2019).

Although many evidences converge in the idea that most of the Schizophrenic symptoms arise from abnormalities in the PFC, one should reject the idea that Schizophrenia is merely a disease of the PFC. Is important to have in mind that the PFC did not perform any physiological process entirely by itself, but instead it is an association cortex which is characterised by a large number of reciprocal connections with sensor and motor areas as well as thalamus and other subcortical regions. Moreover, the complexity and the polymorphism of the disease suggest

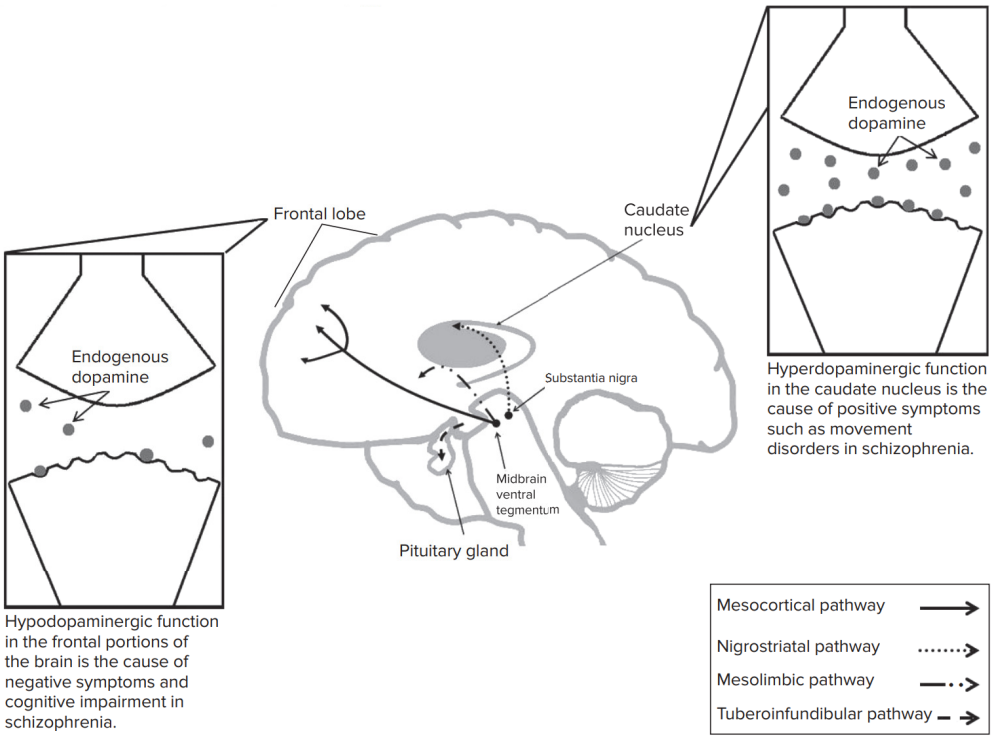

Figure 2.12. Pathophysiology of Schizophrenia. From Patel et al. (2014) adapted from Kandel et al. (2000) and Stahl (2000)

an interaction between multiple brain areas as we will see in the next subsection.

2.4.2 Cerebellar-Prefrontal Loop in Schizophrenia

For many years the scientific community considered Schizophrenia as a disease affecting the cognition areas in the cortex, for instance the PFC. However, an increasing number of evidences have unveil that the cerebellum largely influences the PFC playing an important role in higher cognitive functions and therefore in Schizophrenia. The work from Nancy Andreasen and colleagues has set the foundations of the "cognitive dysmetria" theory explaining the Schizophrenic symptoms (Andreasen et al., 1996, 1998, 1999). This theory postulates that a dysfunction or disruption of the prefrontal-thalamic-cerebellar circuit will lead to a desynchronisation between those areas resulting in deficits in the processing

and coordination of the information. Both anatomical and functional evidences strongly supports this theory. Abnormalities or lesions in the cerebellum result in cognitive impairment in associative learning, planning and working memory, indicating that the cerebellum is part of the brain circuits involved in these functions (Rogers et al., 2011). Similar to the PFC, in Schizophrenia the cerebellum shows a reduction in volume, abnormalities in the blood flow and, importantly, a reduction of the Purkinje cells (Antonova et al., 2004; Martin and Albers, 1995; Andreasen and Pierson, 2008; Mittleman et al., 2008). Worth mentioning that Purkinje cells provide input to the dentate nuclei (DN) in the deep nuclei of the cerebellum and have the important role of deciding which information is returned to the cortex by inhibiting the DN through GABA neurotransmitter. Regarding anatomical evidences of the role of the cerebellum in Schizophrenia, the studies from Strick and colleagues using retrograde and anterograde tracers found cerebellar input from DN to parietal and frontal cortex, notably in areas 9 and 46, via the MD (Middleton and Strick, 2001; Kelly and Strick, 2003) (Figure 2.13B). This circuit is closed by the input from the PFC to cerebellum passing through the pontine nuclei (PN). In another circuit the cerebellum sends inputs to the ventral tegmental area (VTA) (via the reticulo-tegmental nucleus (RTN) and the pedunculo-pontine nuclei (PPT)) that project directly to the PFC (Schwarz and Schmitz, 1997) (Figure 2.13A). The VTA is composed of dopaminergic neurons that are excited by glutamatergic inputs coming from the cerebellum release DA into the PFC (Carta et al., 2019). This circuit corresponds to a part of the mesocortical dopaminergic pathway. These neuroanatomical evidences of the connectivity between PFC and cerebellum have been also corroborated in electrophysiology studies describing how stimulation in the cerebellum modify the dynamics of the PFC (Watson et al., 2009, 2014; Mittleman et al., 2008) and widely proven its implication in WM (Sobczak-Edmans et al., 2019; Manoach, 2003; Ilg et al., 2013).

On the other hand, fMRI studies have found abnormalities in the functional connectivity between PFC and cerebellum in schizophrenic patients (Meyer-Lindenberg and Bullmore, 2010). More recently, also a fMRI study from Brady and colleagues (Brady Jr et al., 2019) revealed that a lower connectivity between the dlPFC and the MD strongly correlates with higher negative Schizophrenic symptoms. Moreover, in the same study, to prove the hypothesis that disconnection between cerebellum and PFC is the causal mechanism of the

A. B.

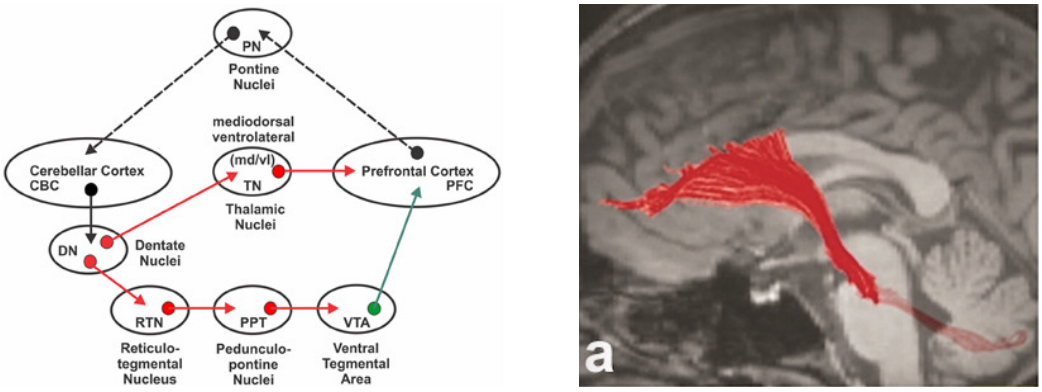

Figure 2.13. Prefrontal-Cerebellar neural circuits A. Neural circuitry involved in cerebellar modulation of medial PFC dopamine. With the exception of inhibitory cerebellar to dentate nucleus projections, red arrows indicate glutamatergic pathways. The green arrow indicates the mesocortical dopaminergic pathway. Dotted black arrow Indicates feedback loop. From Rogers et al. (2013). **B.** Fronto-ponto-cerebellar tractography reconstructed on a 3D T1-weighted image. From Kamali et al. (2010).

negative symptoms they performed repetitive transcranial magnetic stimulation (rTMS) targeting the cerebellum which produced and enhancement of the functional connectivity between the two areas and relevantly the increase was correlated with the reduction of the negative symptoms. The mechanisms underlying this amelioration due to cerebellar stimulation are not entirely known but the evidences suggest that glutamatergic and dopaminergic synaptic transmission could play an important role.

Part II

ORIGINAL RESEARCH

3 The Role of Neuronal Nonlinearities in Working Memory

In this chapter we present a theoretical framework to study network mechanisms of persistent selective activity like the one observed in PFC during WM tasks., We specifically investigate how the interplay between the nonlinearities in the input-output transfer function (TF) of individual neurons and the recurrent network dynamics gives rise to multi stability between ongoing activity and a continuum of states in which the activity is localized. We show on the mechanism underlying the initiation and termination of the persistent selective activity depends the single neuron nonlinearities. In the first section, we present a model network of rate neurons connected all-to-all with a "Mexican's hat" structured connectivity by which neurons with similar PDs are strongly connected. The nonlinearities in the neuron TF allows to have persistence and selectivity within a single network. The two studied nonlinearities are the acceleration close to the firing threshold and concavity near the saturating part. These nonlinearities can be found in typical input-output neuron TF measured in *in vitro* experiments (Arsiero et al., 2007; Lafon et al., 2017; Schiff and Reyes, 2012). In order to simplify the analytical calculations we chose the TF to be piecewise linear functions. In the acceleration or expansive nonlinearity, the slope of the TF increases while in the concavity or compressive nonlinearity, the slope decreases. We show that the nature of the nonlinearity shapes the

configuration of the steady states. In the both cases, for a given connectivity parameters, we found a region of multistability homogeneous state and a bump state. Then we explore the mechanisms for terminating the bump state. Due to the configuration of the steady states, in the model with expansive nonlinearity the transition from the bump state to the homogeneous can be mediated by a global inhibitory input. On the other hand, in the compressive nonlinearity, the transition from persistent to ongoing activity states can be induced by a global excitatory input. This mechanism is appealing since most electrophysiological recordings of PFC neurons during WM task show a global increase in the firing rate around of the response period. The only drawback of this mechanisms is that the neuron in the center of the bump state are firing close to their saturation rates which is not consistent with the experimental data showing that the neurons firing at their PDs are far away from the typical saturating firing rates. To overcome this issue, we propose a mechanism in which the interaction between excitatory and inhibitory populations with different accelerating TFs give rise to an effective saturation. The main idea behind this mechanism is that at a given input the slope of the inhibitory TF is larger than the slope of the excitatory TF resulting in a qualitatively similar scenario as in the compressive nonlinearity. We show different configurations of the excitatory and inhibitory TFs that leads to an effective saturation and we prove that the same mechanisms regarding the termination of the bump state through global excitation still holds. In order to study how the excitatory and inhibitory populations interact in the effective saturation mechanism and to investigate in detail the dynamics taking place in the transitions between steady states and the formation of spatiotemporal patterns of activity we implement the effective saturation mechanism into a spiking network model. The model consists in an excitatory and inhibitory populations of leaky integrate-and-fire neurons whose connectivity is as in our rate models. The nonlinearities of the single neuron TFs are implemented by adjusting the intrinsic membrane properties together with background noise. We demonstrate that the effective saturation mechanism as well as the bump termination by a global excitation still applies. We study in detail the dynamics during the transitions between the persistent and ongoing activity states We find that for input durations of the order of the synaptic and membrane time constants an oscillatory spatiotemporal pattern appears.

Interestingly, due to these oscillations, only certain values of input intensity will successfully terminate the persistent state. Taken together, this work provides a mechanism in which excitatory and inhibitory TFs interact giving rise to multistability in which termination of the persistent selective state can be mediated by a global excitation.

3.1 INTRODUCTION

Working memory (WM), the ability to temporarily hold, integrate, and process information to produce goal-directed behavior, is crucial to higher cognitive functions such as planning, reasoning, decision-making, and language comprehension Baddeley (1986); Fuster (2015). The persistent activity recorded in neocortex during WM tasks is thought to be the main neuronal correlate of WM Fuster and Alexander (1971); Miyashita and Chang (1988); Goldman-Rakic (1995). For example, in an oculomotor-delayed response (ODR) task in which a monkey has to remember the location of a stimulus for several seconds to make a saccade in its direction, a significant fraction of the neurons in the prefrontal cortex (PFC) modify their activity persistently and selectively to the cue direction during the delay period Funahashi et al. (1989, 1990, 1991); Constantinidis et al. (2001); Takeda and Funahashi (2007). The classical view is that this reflects a multistability in the dynamics of the PFC circuit because sensory inputs are the same in the precue and in the delay periods but neuronal activity is different Hebb (1949); Hopfield (1984); Amit and Brunel (1995, 1997); Wang (2001).

In monkeys performing an ODR task, neurons in the dlPFC show elevated activity during the delay period which depends on the direction of the presented cue (e.g. Constantinidis et al. (2001); Constantinidis and Wang (2004); Funahashi et al. (1989, 1990, 1991)). As the direction of the cue is varied the activity of the neuron changes. The response of neuron in the dlPFC is therefore characterized by its tuning curve. Neurons are characterized by its preferred direction i.e. the direction for which the response is maximum. In the classical view, the “line of attractors” hypothesis posits that this direction tuning is an emergent property of the recurrent dynamics in the dlPFC. According to this hypothesis the location

of the cue is encoded in persistent states, each state characterized by an activity profile in the feature space which is "bumpy". Since this space has the geometry of a ring, the location of the bump can be parametrized by an angle which match the direction of the cue. Thus, the set of attractors is continuous which is invariant by rotation. Recent experiment results support this "line of persistent attractors" hypothesis (see Knierim and Zhang (2012); Hulse and Jayaraman (2020) for a review). The results reported in Wimmer et al. (2014) show that the behavioral error patterns in ODR tasks correlate with shifts in the tuning curves of individual dlPFC neurons on these error trials, implying a drift of an activity bump during the delay period. Persistent activity bumps encoding for a direction were observed in *drosophila* flies using calcium imaging Seelig and Jayaraman (2015); Kim et al. (2017); Green et al. (2017).

Non selective persistent activity emerges naturally in unstructured recurrent networks provided the recurrent excitation is strong enough Wang (1999); Hansel and Mato (2001). To prevent the activity to blow up in the persistent state a non-linearity is required. Sigmoidal input-output neuronal transfer function provides the network with an appropriate stabilizing mechanism Brunel (2000). However, in that case, neurons in the persistent state will fire near saturation. Alternatively, the stabilizing non-linearity can result from the recurrent inhibition. This requires the response of the inhibitory neurons to be more sensitive to inputs than the excitatory neurons Latham et al. (2000); Brunel and Wang (2001).

On another hand, non-persistent activity selective to a stimulus characterized by a continuous angular variable can emerge from recurrent interactions. This was first investigated in Ben-Yishai et al. (1995). The "ring model" introduced in this seminal paper and its generalizations provide a classical framework to investigate the role of excitation and inhibition in orientation selectivity (see Hansel and Sompolinsky (1998); Priebe (2016) for a review). Selectivity in the ring model stems from feature specific strong excitatory connectivity stabilized by inhibition. Combined with the non-linearity of the input-output transfer function of the neurons, the network undergoes a Turing instability, as the external input increases, leading to a line of bump attractors Goldberg et al. (2004). In this mechanism, the bump attractors are not persistent: if the external input which represent the stimulus is withdrawn the bump is abolished.

Compte et al. Compte et al. (2000) showed that persistence and selectivity to direction can emerge in a network with structured connectivity. With this model they studied the involvement of slow excitation via N-methyl-D-aspartate receptor (NMDAR) channels, and showed that a decrease in the percentage of NMDAR channels out of the excitatory synapses can destabilize the representation in the network.

In this paper we investigate how the interplay between excitation, inhibition and the nonlinearities of input-output neuronal transfer function gives rise to direction selective and persistent delay activity. To this end we consider networks of excitatory and inhibitory neurons with feature specific connectivity. Neurons are modeled as rate units or integrate-and-fire elements. We combine analytical calculation with numerical simulations to characterize the network stable states as a function of the external input, the interactions strength and the spatial modulation. We show that the transient network dynamics during the switch-on and the switch off of the persistent selective state depends on the nature of the non-linearities in the input-output transfer function of the neurons.

3.2 RESULTS

3.2.1 Nonlinearities in the neuron transfer function shape the steady states configuration

The frequency-current neuron transfer function (TF), describes the relationship between the input current to a neuron and its output firing rate. It can be estimated experimentally through the recording of actions potentials elicited by a neuron subjected to a protocol of current injection (Cardin et al., 2008; Carvalho and Buonomano, 2009). With the aim to study the effect of this nonlinearities on the steady states configuration and the transient behavior between them, we implement a firing rate model of visuospatial working memory with structured ring connectivity consistent with the columnar architecture of the PFC (Goldman-Rakic, 1995; Rao et al., 1999) (model described in Appendix A.1 and Figure C.1. This specific connectivity, in which neurons

with similar tuning properties are more connected, is supported by electrophysiological data showing the horizontal connections of the superficial layers II-III of the dlPFC (Levitt et al., 1993; Kritzer and Goldman-Rakic, 1995; González-Burgos et al., 2000). The ring network connectivity used in several PFC models (Camperi and Wang, 1998; Compte et al., 2000; Gutkin et al., 2001; Hansel and Mato, 2013; Wimmer et al., 2014), provides a mathematical framework accounting for orientation selectivity (Ben-Yishai et al., 1995). At the time an external tuned input reaches the network, it will specifically activate some neurons whose preferred orientations are aligned to that of the input, giving rise to a bump of activity. The nonlinearities in the TF together with the structured connectivity will give rise to a bump attractor dynamics accounting for selective persistent activity (Camperi and Wang, 1998). In this section we use a reduced version of the general model consisting of one effective population where all the rate units share a common TF and time constant (see Appendix A.3). Hence, the synaptic activity dynamics is described as follows:

$$\tau \frac{dm(\theta,t)}{dt} = -m(\theta,t) + g(C + I_{rec}(\theta,t) + I_{stim}(\theta,t)) \tag{3.1}$$

where τ is the synaptic time constant, θ is the neuron PD which is distributed evenly between $-\pi$ and $+\pi$ and is defined as the direction of maximal response of a neuron that receives a tuned I_{stim} centered at θ. $m(\theta,t)$ is the activity of the rate unit with PD θ, $g(x)$ stands for the TF linking instantaneously total synaptic input with the firing rate, C is the external background input, $I_{stim}(\theta,t)$ is the external stimulation and $I_{rec}(\theta,t)$ is the recurrent input arriving to the postsynaptic rate unit with PD θ which, in the limit of large network, is defined as:

$$I_{rec}(\theta,t) = \frac{1}{2\pi}\int_{2\pi} J(\theta-\xi)m(\xi,t)d\xi \tag{3.2}$$

where $J(\theta-\xi)$ is the connectivity function between postsynaptic neuron with PD θ and the presynaptic neuron with PD ξ, finally $m(\xi,t)$ refers to the activity of the presynaptic neuron with PD ξ. We choose the connectivity function to be $J(\theta-\xi) = J_0 + J_1\cos(\theta-\xi)$ where J_0 and J_1 are connectivity parameters controlling the strength of the homogeneous connectivity and the angular modulation, respectively.

To simplify the analysis and calculations we choose piecewise linear TF to describe the nonlinearities. The chosen TF is defined as follows: for inputs between zero and a threshold T, the slope is α; for inputs larger than T, the slope is β; and zero for negative inputs (Appendix A.3). Thus, the relationship between α and β will determine the nature of the nonlinearity. That is to say, $\beta > \alpha$ describes an expansive nonlinearity whereas $\beta < \alpha$ describes a compressive nonlinearity (inset Figure 3.1 A and D). For the sake of simplicity we take $\alpha = 1$. In order to get an insight about the configuration of the steady states in both cases (expansive and compressive nonlinearity in the TF) we analyse the FPs of the system considering $I_{stim}(\theta) = 0$. Due to the symmetry of the network we can consider, without loss of generality, the steady states profiles which are symmetric around $\theta = 0$. Thus, the steady states equation derived from eqs. (3.1) and (3.2) is:

$$m(\theta, t) = g(C + J_0 m_0 + J_1 m_1 \cos(\theta)) \tag{3.3}$$

where m_0 and m_1 correspond to the zeroth and first Fourier components of the activity, describing the mean and the angular modulation of network activity, respectively. This parameters are defined as follows:

$$m_k = \frac{1}{2\pi} \int_{2\pi} m(\theta) \cos(k\theta)\, d\theta, \quad \text{with} \quad k \in \{0, 1\} \tag{3.4}$$

In absence of external input, it exists an homogeneous solution, where all the population presents the same activity, $m(\theta) = m_0$ (since $m_1 = 0$). The dynamics associated to this homogeneous steady state corresponds to the spontaneous activity in the PFC in the pre-cue and post-response periods. The homogeneous activity can be determined by solving the following equation:

$$m_0 = g(C + J_0 m_0) \tag{3.5}$$

The stability conditions of the FPs solutions are given by $J_0 g'(I) < 1$ and $J_1 g'(I) < 2$ where $g'(I)$ stands for the derivative of the TF at input I (i.e. 0, α or, β depending on the total input $I(\theta) = C + J_0 m_0 + J_1 m_1 \cos(\theta)$) (Appendix A.3). The stability of the homogeneous FP is ensured if both conditions are fulfilled. The first stability condition points out that an increase of the homogeneous

connectivity J_0 will destabilize the homogeneous FPs. On the other hand, the second stability condition, points that an increase of the angular modulation of the connectivity J_1, will destabilize the homogeneous FP (Figure 3.1 B and E and Figure C.3). Even more relevant is the fact that in both stability conditions, if the slope of the TF increases, the homogeneous FP becomes unstable.

The bifurcation diagrams show the stable states mean activity (m_0) and its angular modulation (m_1, color coded) under variations of the background input C for the expansive and compressive nonlinearity (Figure 3.1 A and D, respectively). In the expansive nonlinearity scenario ($\beta > 1$), for small C, the homogeneous is the only stable state. Increasing C, the slope of the TF increases to a value β, leading to a destabilisation of the homogeneous state and the apparition of a bump state. In this case, for large values of C the bump is the only stable state. Note that there is a small region of bistability where homogeneous and bump states coexist, being the bump state the one displaying larger activity (Figure 3.1 A). Notice that the dynamics associated with the bump steady states corresponds to the characteristic persistent selective activity in the PFC during the delay period. On the other hand, in the compressive nonlinearity scenario ($\beta < 1$), for small values of C, the bump is the only stable state. By increasing C, the slope of the TF decreases to a value β, leading to a stabilisation of the homogeneous state and thus to the generation of a bistable region (Figure 3.1D and Figure C.3A). In this particular case, the bump state exists only in a specific range since it destabilises at large values of C. As in the expansive nonlinearity, a region of bistability also exists but here the bump state is the one displaying smaller values of activity.

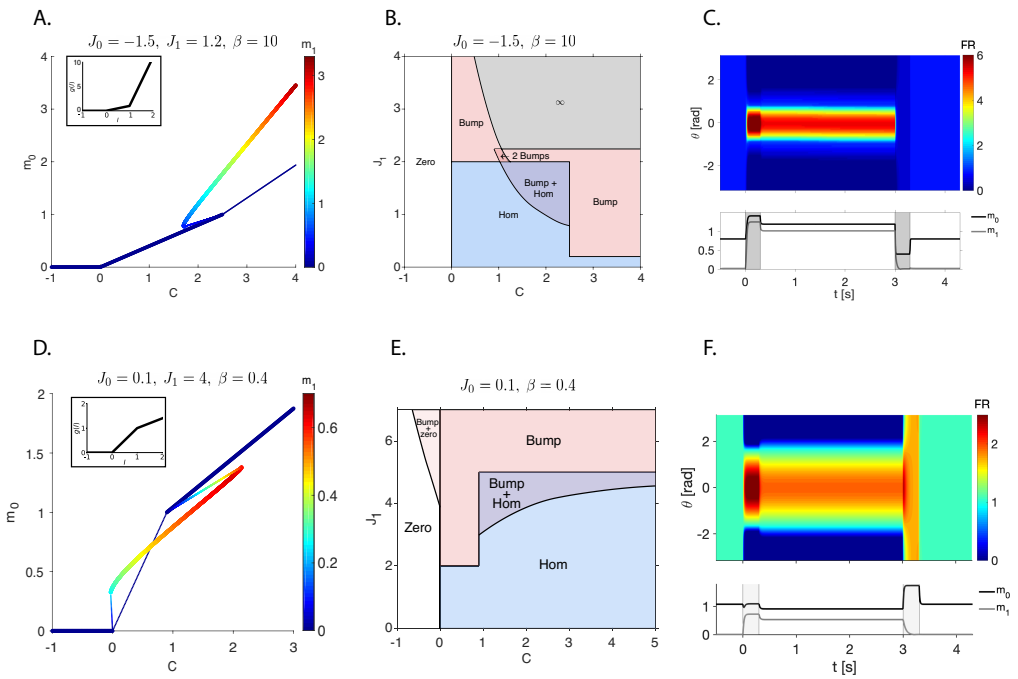

Figure 3.1. Steady state analysis and simulations of the firing rate model for expansive and compressive nonlineatities in the TF. A. Mean activity (m_0)-background input (C) bifurcation diagram describing the fixed points for different values of C for an expansive nonlinearity in the TF. The angular modulation of the activity m_1 is color-coded. The chosen parameters are $\alpha = 1$, $\beta = 10$, $J_0 = -1.5$ and $J_1 = 1.2$. Inset: $g(I)$ function or TF describing an expansive nonlinearity. **B.** $J_1 - C$ phase diagram describing all the stable steady state dynamics that can be found at different values of the angular modulation of the connectivity J_1 and C. **B.** Upper panel; Activity time-series for each rate unit in the network subjected to a modelled ODR trial consisting in a tuned input with duration $300\,ms$ centered at $\theta = 0$ presented at time $t = 0\,s$ representing the visual cue and an homogeneous input at $t = 3\,s$ with same duration representing the response period activity. With fixed background input $C = 2$. The activity amplitude for each unit ($m(\theta)$ ≡FR) is color-coded. Lower panel; Population activity time series showing the time evolution of the mean activity m_0 (black) and angular modulation m_1 (gray). External stimulation periods are enclosed in the shaded areas.**D,E** and **F**: same for the compressive nonlinearity. The chosen parameters are $\alpha = 1$, $\beta = 0.4$ and $J_0 = 0.1$. In **D** $J_1 = 4$ and in **F** $C = 1.5$.

3.2.2 Two different mechanisms for terminating persistent activity with global excitation: long *VS* short stimulus.

We performed simulations of the single network emulating one single trial under the ODR task protocol. We set the network to operate in a bistable regime by choosing a proper value of the background input (C). At time $t = 0\,s$ we apply a transient tuned input of $300\,ms$ duration centered at $\theta = 0$ which represents a visual input from sensory afferents to the PFC upon presentation of a visual cue. The first input elicits a bump state which persists over time since it is a FP. At time $t = 3\,s$ a homogeneous input is applied in order to "switch off" the bump state by transitioning to the homogeneous state. As a direct consequence of the difference in the steady states configuration between the expansive and compressive nonlinearity, the mechanisms relying on steady states allowing to transition from the bump to the homogeneous state (hereinafter referred as B-H transition) will also differ. Here, we refer to slow mechanisms when transient external inputs have a duration much longer than that of the synaptic time constant (τ). In the expansive nonlinearity case, the B-H transition can only be performed by a global inhibitory input (Figure 3.1 C). Whereas in the compressive nonlinearity the B-H transition can be performed by means of a global excitatory input (Figure 3.1 F). In the case that the external input duration is shorter than the synaptic time constant (τ), transitions between stable states that coexists for a certain value of C, can also be performed. In contrast to previous mechanisms relying on steady states, this mechanism is based on the dynamics. Since the input duration is shorter than the system time constant, the stable states will not be perturbed. In both expansive and compressive nonlinearities, owing to the bistable regime, the B-H transition can be performed by a strong short excitatory input (Figure 3.2 A and C). This phenomenon can be better understood by studying the dynamics depicted by the activity components m_0 and m_1 (Equation (3.4)) in the $m_0 - m_1$ phase plane. Using a tuned input centered at $\theta = 0$ to elicit the activation of a bump of activity symmetric around $\theta = 0$, enables to describe the activity trajectory in the $m_0 - m_1$ space since the second Fourier component $m_2 = \frac{1}{2\pi} \int_{2\pi} m(\theta) \sin(\theta)\, d\theta$ will be zero due to the sinus term in the integral. For

a certain value of C ensuring bistability, m_0 and m_1 nullclines are plotted in the space (thin grey and black lines, respectively; Figure 3.2 B and E). FPs are described by the crossings between m_0 and m_1 nullclines. Filled and empty dots describe stable and unstable FPs, respectively. The FP on the m_0 axis corresponds to the homogeneous state ($m_0 = 0$) and the upper (lower) FP correspond to the bump centered at $\theta = 0$. The thick black line depicts the stable manifold of the unstable FP which separates the basin of attraction (BOA) of the stable homogeneous and bump state. The activity trajectory plotted on the $m_0 - m_1$ space describes the fast dynamics caused by the external inputs. The trajectory of the activity under a tuned input centered at $\theta = 0$ applied to the network at time $t = 0\,s$, is described by the green line. Note that the trajectory begins at the homogeneous FP and after the stimulus removal it relaxes towards the bump FP. Next, a global strong short excitatory input is applied at time $t = 3\,s$. The input drives the trajectory away from the bump FP due to an increase of the m_0 component, crossing the separatrix and penetrating into the homogeneous BOA (blue* line in Figure 3.2 B and D). Thus, when the input is withdrawn the trajectory evolves towards the homogeneous FP.

How strong and short has to be this stimulus to effectively switch from one stable state to the other? To get some insight we calculated, for a given stimulus duration, the minimal input current (Figure 3.3 B and D) required to perform either H-B or B-H transition (Figure 3.3, green and red curve, respectively). We also show the minimal input charge expressed as the minimal input current multiplied by the duration of the stimulus (Figure 3.3 A and C). Any input equal to or greater than the curve enables the state transition. We chose the input owed to perform the H-B transition to be tuned and centered at $\theta = 0$ and the input owed to perform the B-H transition to be homogeneous. In order to interpret these phase diagrams we have to underscore that the input duration at the vicinity of the synaptic time constant ($\tau = 10\,ms$) delimits two different regimes where different transition mechanisms take place. For short input durations ($\tau \gg \Delta t_{stim}$) the state transitions rely on the dynamics and to understand the behavior of the transitions we have to study the trajectories in the activity space (Figure 3.2 B and D). Conversely, for long input durations ($\tau \ll \Delta t_{stim}$) the state transitions rely on the attractors and to explain the transitions we have to study the steady states under the external input as a bifurcation parameter (Figure 3.1 A and D).

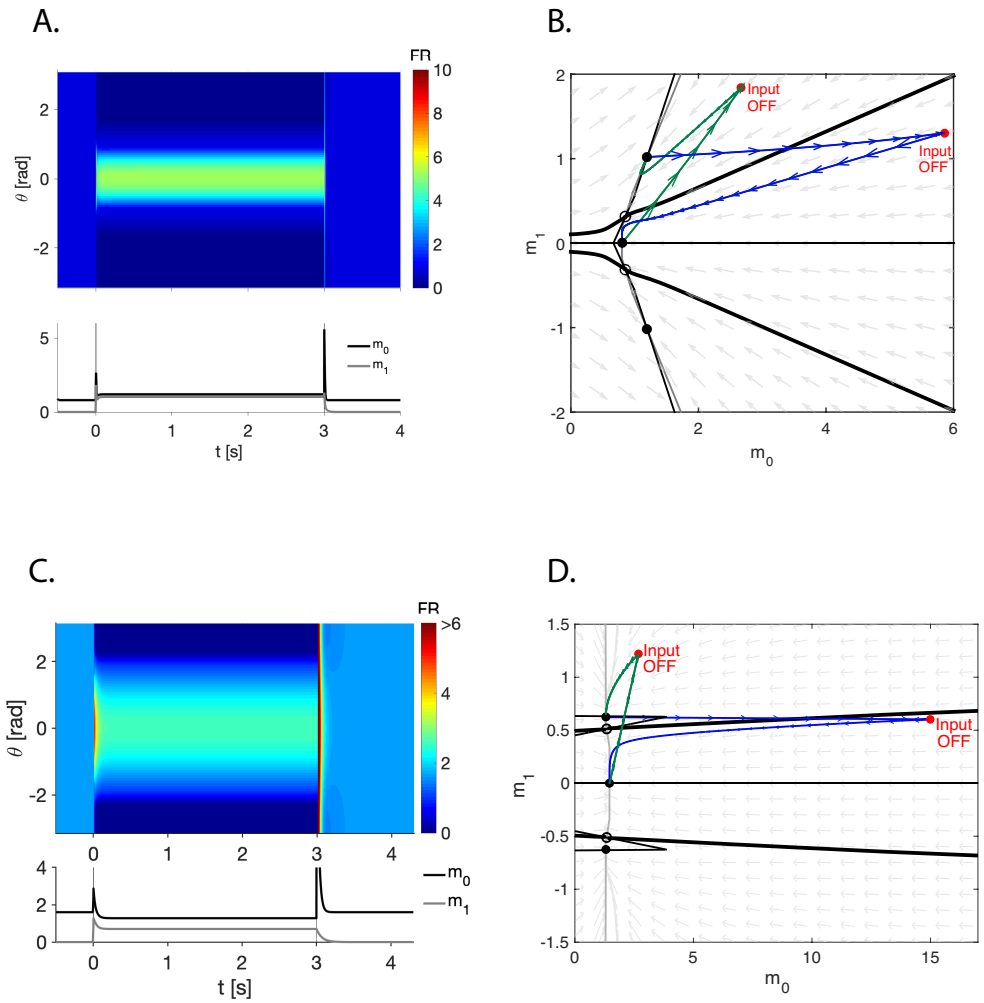

Figure 3.2. Short intense excitatory input terminates persistent activity. A. Dynamics of the network in response to a $3\,ms$ tuned excitatory pulse at $t = 0\,s$ centered at $\theta = 0$, and a $0.5\,ms$ homogeneous excitatory pulse at $t = 3\,s$. Parameters are as in Figure 3.1 C. **B.** Dynamics of m_0 and m_1 from **A** plotted on the $m_0 - m_1$ phase plane (see text). Green: trajectory during switch-on ($0 < t < 3$ ms). Blue: trajectory during switch off ($3 < t < 3.0005$ s). Thin black and grey lines are the m_0 and m_1 nullclines, respectively. Thick black line depicts the separatrix. Stable and unstable FPs are represented by a solid and empty black dot, respectively. Red dots represent the extinction of the external stimulation. **C** and **D**. The same for the compressive nonlinearity case. Parameters as in Figure 3.1 F. In all the simulations we used $dt = 0.01\,ms$; $\tau = 10\,ms$.

This being set, for the expansive nonlinearity, the B-H transition curve shows a constant value of charge for short input duration indicating that the input current needed to cross the separatrix scales with the duration of the input (Figure 3.3A, red). As the stimulus duration approaches to τ, the curve for both input charge and current depict asymptotic behavior (Figure 3.3 A and B). This was expected since in the expansive nonlinearity the B-H transition cannot be performed by a long positive input as seen in bifurcation diagram (Figure 3.1 A). The H-B transition curve, in green, shows a constant value of charge for small input duration below , while for large input duration, the input current is the one that reaches a constant value. This result shows that for input duration shorter than τ, the charge needed to cross the separatrix tends to a constant value. In other words, the minimal input current scales with the duration, again indicating that the mechanism relies on the dynamics. On the other hand, for large input duration, the input current needed to perform the transition tends to a constant value, i.e. the transition rely on the steady states. For the compressive nonlinearity, the main differences are in the B-H transition (Figure 3.3 C and D, in red). For short input durations, the minimal input charge tends to a constant value but two orders of magnitude larger than the charge needed to perform the B-H transition in the expansive nonlinearity case. This difference is mainly caused by the fact that the separatrix in the activity space for the compressive nonlinearity is almost parallel to the m_0 axis (Figure 3.2 D). Thus, for the activity to penetrate into the homogeneous BOA a large input charge is needed. Conversely, for large input durations, the B-H transition reaches a constant value of input current (Figure 3.3 D, red). In this case, the difference arises from the steady states configuration (Figure 3.1 D) in which the B-H transition can be performed by a global excitatory input. Thus, the constant value of the minimal input to perform the B-H transition can be computed from the bifurcation diagram as the difference between the initial background input and the largest value of input at which the bump state exists. This value is around 0.15 for $C = 2$ (see Figure 3.1 D and top inset Figure 3.3 D).

Putting all together we have shown two different regimes defined by the synaptic time constant with distinct mechanisms for transition between states by means of a global excitation. The dynamical regime ($\tau \gg \Delta t_{stim}$) allows the B-H transition to be performed by a strong excitatory input for both expansive and compressive nonlinearity while the steady state regime ($\tau \ll \Delta t_{stim}$) only allows

the B-H transition to be performed by a global excitatory input in the compressive nonlinearity. Taken together, the compressive nonlinearity seems to be a suitable substrate underlying termination of persistent selective activity by a global excitatory input.

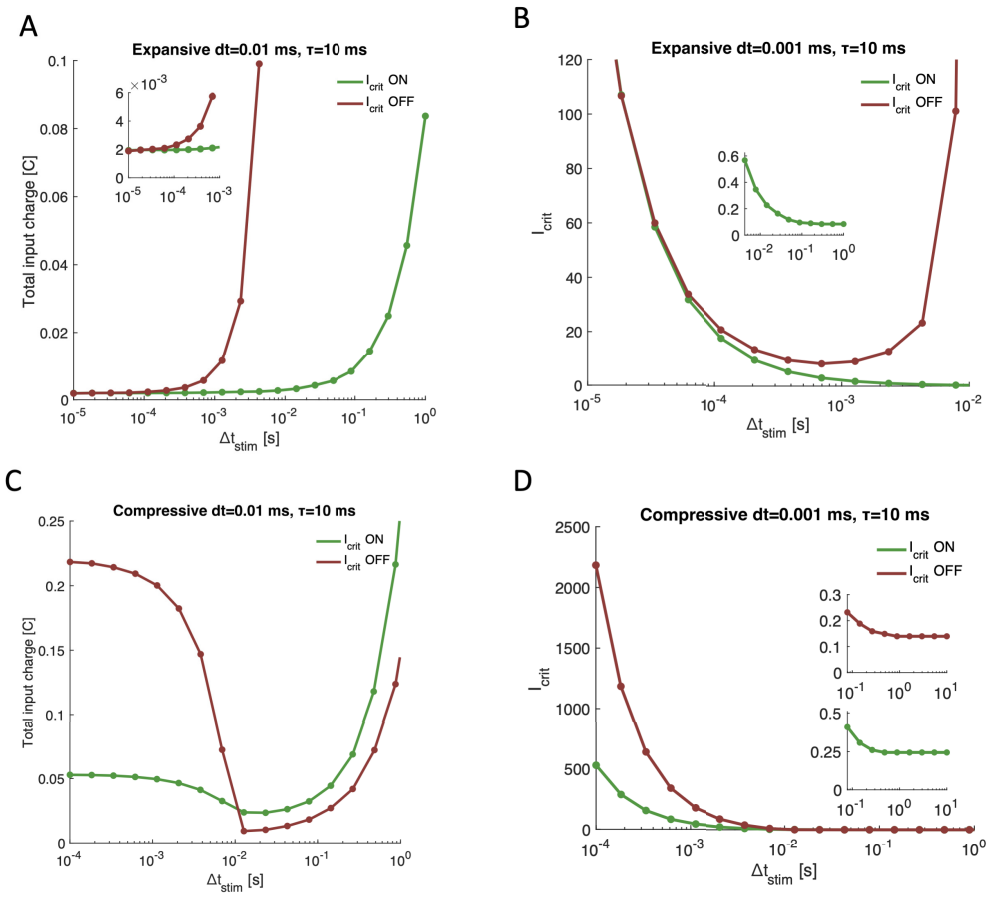

Figure 3.3. Synaptic time constant delimits two different regimes with distinct transition mechanisms. A. Minimal input charge ($I_{crit} \cdot \Delta t_{stim}$) needed to perform a state transition plotted against the duration of that input, Δt. Homogeneous to bump state transition (H-B) in green and bump to homogeneous state transition (B-H) in red. Inset: zoom for durations below $1\,ms$. **B.** Same as in panel **A** but the y-axis is representing the minimal input charge I_{crit}. Inset: zoom for durations above $10\,ms$. **C** and **D**. Same as in **A** and **B**, respectively, for the compressive nonlinearity case.

3.2.3 The interplay between excitatory and inhibitory populations elicits an effective saturation.

With the aim of studying the mechanisms for termination of persistent selective activity emerging from the compressive nonlinearity in a more biologically plausible model, we consider a network of two different populations one accounting for excitatory and the other inhibitory neurons. We implement the previous model relaxing the assumption that both populations have the same TF. In this case, the synaptic dynamics of the excitatory and inhibitory populations will be different and thus, described by separated equations of the form of eq. (3.1). The total input arriving to a rate unit in the absence of external stimulus is described as follows:

$$I(\theta) = C + (J_E * m_E)(\theta) - (J_I * m_I)(\theta) \tag{3.6}$$

where J_X is the connectivity function, m_X is the activity of the population $X \in \{E, I\}$ and the operator $*$ is described as $(f * g)(\theta) = \frac{1}{2\pi}\int_{2\pi} f(\theta - \xi)g(\theta)d\xi$. The two last terms in eq. (3.6) represent the excitatory and inhibitory recurrent interaction, respectively. In the steady state the activity for each population is described by $m_X(\theta) = g_X(I(\theta))$ with $X \in \{E, I\}$. Substituting this expression in eq. (3.6) yields the total input in the steady state:

$$I^*(\theta) = C + \frac{1}{2\pi}\int_{2\pi} F_0(I^*(\xi))d\xi + \frac{1}{2\pi}\int_{2\pi} F_1(I^*(\xi))\cos(\theta - \xi)d\xi \tag{3.7}$$

where $F_i(I) \stackrel{\text{def}}{=} J_i^E g_E(I) - J_i^I g_I(I)$ with $i = 0, 1$. Notice here that for $g_E(I) = g_I(I)$, these equations match the ones from the previous section. In addition, the Fourier components of the activity defined in eq. (3.4) also applies. In this general model, it also exists an homogeneous state where $m_1^X(I) = 0$ and the activity of a neuron in population X is described by:

$$m_X(\theta) = g_X(C + I_0) \tag{3.8}$$

where $I_k \stackrel{\text{def}}{=} J_k^E m_k^E - J_k^I m_k^I$. With the definition of I_0 and eq. (3.8) we can write

another equation for I_0:

$$I_0 = F_0(C + I_0) \tag{3.9}$$

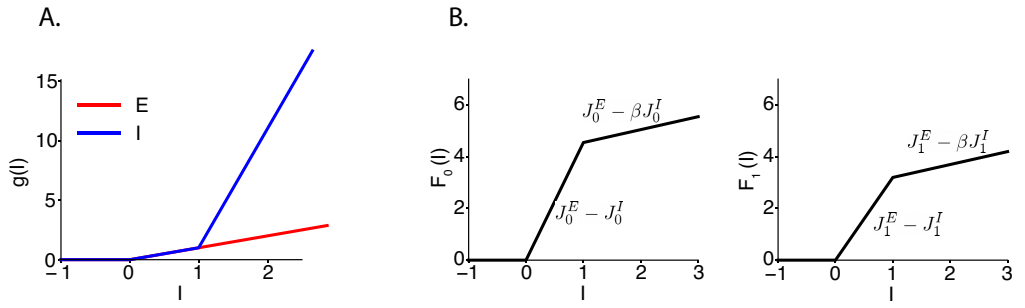

Figure 3.4. Two population mechanisms to terminate the bump activity through global excitation. A. Excitatory (red) and inhibitory (blue) neurons' TFs in the acceleration difference model. The slope of the inhibitory TF is 1 for $0 < I < 1$ and β for $I > 1$. **B.** Example of $F_0(I)$ (left) and $F_1(I)$ (right) in the acceleration difference model. The slope of the first segment ($0 < I < 1$) is $J_i^E - J_i^I$ and of the second segment ($I > 1$) is $J_i^E - \beta J_i^I$ ($i = 0, 1$).

The activity in the steady state can be calculated by solving eq. (3.9) and substituting I_0 into eq. (3.8). The stability conditions for the homogeneous state are $F_0'(I) < 1$, describing the rate stability and $F_1'(I) < 2$, describing the spatial stability (Appendix A.4). Since $F_i'(I) = J_i^E g_E'(I) - J_i^I g_I'(I)$, in order to obtain a similar scenario as in the compressive nonlinearity in Figure 3.1 D, the slope of the excitatory TF has to be larger than the slope of the inhibitory TF for small inputs, but given a input threshold, the inhibitory slope has to be larger than the excitatory to obtain a effective saturation. This can be accomplished by two different mechanisms: creating a difference in the TF acceleration and/or creating a difference in the TF threshold between the excitatory and inhibitory neurons (Figure C.4). One simple example of the acceleration difference is to use a threshold linear TF for the excitatory neurons with slope equal to 1 and a piecewise linear TF with an increase in the slope at a given threshold T (Figure 3.4 A). To show the effective saturation we plot F_0 and F_1 as a function of the input which represents the components of the network effective TF (Figure 3.4 B). For inputs between 0 and 1 the slope of $F_0(I)$ and $F_1(I)$ will be

$J_0^E - J_0^I$ and $J_1^E - J_1^I$, respectively. While for inputs above 1, the slope of $F_0(I)$ and $F_1(I)$ will be $J_0^E - \beta J_0^I$ and $J_1^E - \beta J_1^I$, respectively. Depending on the connectivity parameters J_i^X and the slope parameter β, we can obtain different scenarios of effective saturation, for instance, the slope of the segment for $I > 1$, can be smaller than the slope of the segment $0 < I < 1$ or even negative. Notice that in the reduced model F_0 and F_1 are proportional to the TF, thus the compressive mechanism exploits the nonlinearity close to the saturation and the neurons display an activity level close to the saturation rate both in the homogeneous state and close to the center of the bump. Interestingly, unlike in the reduced model, F_0 and F_1 depend on the interplay between the excitatory and inhibitory TF and the connectivity, thus the same mechanism elicited by the compressive nonlinearity can be obtained without involving the saturating part of the neuronal TF.

For the acceleration difference mechanism described in Figure 3.4 and for a given connectivity parameters, we obtain the bifurcation diagram of the mean excitatory activity (m_0^E) and the spatial modulation of the activity (m_1^E, color coded) under variations of the background input C. Similar to the bifurcation in the compressive nonlinearity (Figure 3.1 D), there is a bounded region of C where a bump state exists and for larger values of C it destabilises and the only stable state is the homogeneous. In a relatively large region of C the bump state coexists with the upper homogeneous state. There is also a small region of bistability between the zero stable state and the bump and a small region where the bump is the only stable state. By studying the steady states as a function of the spatial connectivity parameter of the inhibitory population (J_1^I) we can say that the spatial stability of the homogeneous state is guaranteed provided that $J_1^E - J_1^I < 2$ and $J_1^E - \beta J_1^I < 2$ (i.e. $1 > J_1^E - 2J_1^I < \beta$). This two stability conditions are represented by the horizontal lines separating the homogeneous from the bump state in the $J_1^I - C$ phase diagram (Figure C.4). We performed simulations of the two populations network under the same stimulation protocol used in Figure 3.1 C and F. As expected from the bifurcation diagram of the acceleration difference mechanism, the termination of the persistent selective activity can be performed by a global excitatory input of duration larger than the synaptic time constant . Interestingly, for the same mechanism exposed in the previous section, the B-H transition can also be mediated by an input of

duration shorter than the synaptic time constant (Figure 3.5 B and C). Diagrams showing the minimum input charge and intensity to perform the B-H steady state transition show a qualitatively similar scenario for long stimulus as in the diagrams for compressive nonlinearity (Figure 3.5 D). In both cases the stimulus intensity saturates to a constant value as stimulus duration increases above well above $\Delta t_{stim} > 1\, ms$ (Figure 3.5 D, inset). Regarding the minimal input charge, as the duration of the stimulus is reduced to a value of the same order of magnitude than the synaptic time constants ($\tau_E = 5$ ms and $\tau_I = 1$ ms) the minimal charge needed to do the switch off increases to a maximum around 2 ms. This behavior represents the effect of having two population with different time constants. Since excitatory time constant ($\tau_E = 5$ ms) is larger than the inhibitory (τ_I =1 ms), for stimulus duration of the order of τ_E the inhibitory firing rate increases faster than the excitatory since the external input is integrated faster for the inhibitory population and thus larger current is needed to increase the excitatory firing rate (and thus m_0^E) in order to cross the separatrix. This effect is maximum for stimulus duration larger than the τ_I and smaller than the τ_E. Once the stimulus duration becomes much smaller than τ_I this effect is attenuated and the minimal charge decreases to a saturation value around 0.1 ms. Conversely to the behavior in the compressive diagrams where the charge saturates at a very large values for very small stimulus ($\tau \gg \Delta t_{stim}$), in the effective saturation scenario, the charge decreases towards small values. This difference may be a reflex of the steeper separatrix in the $m_0^E - m_1^E$ space related to the almost horizontal separatrix in the compressive nonlinearity case. The same analysis for the threshold difference as well as for the combination of both effective saturation mechanisms can be found in Figure C.4 and Figure C.5.

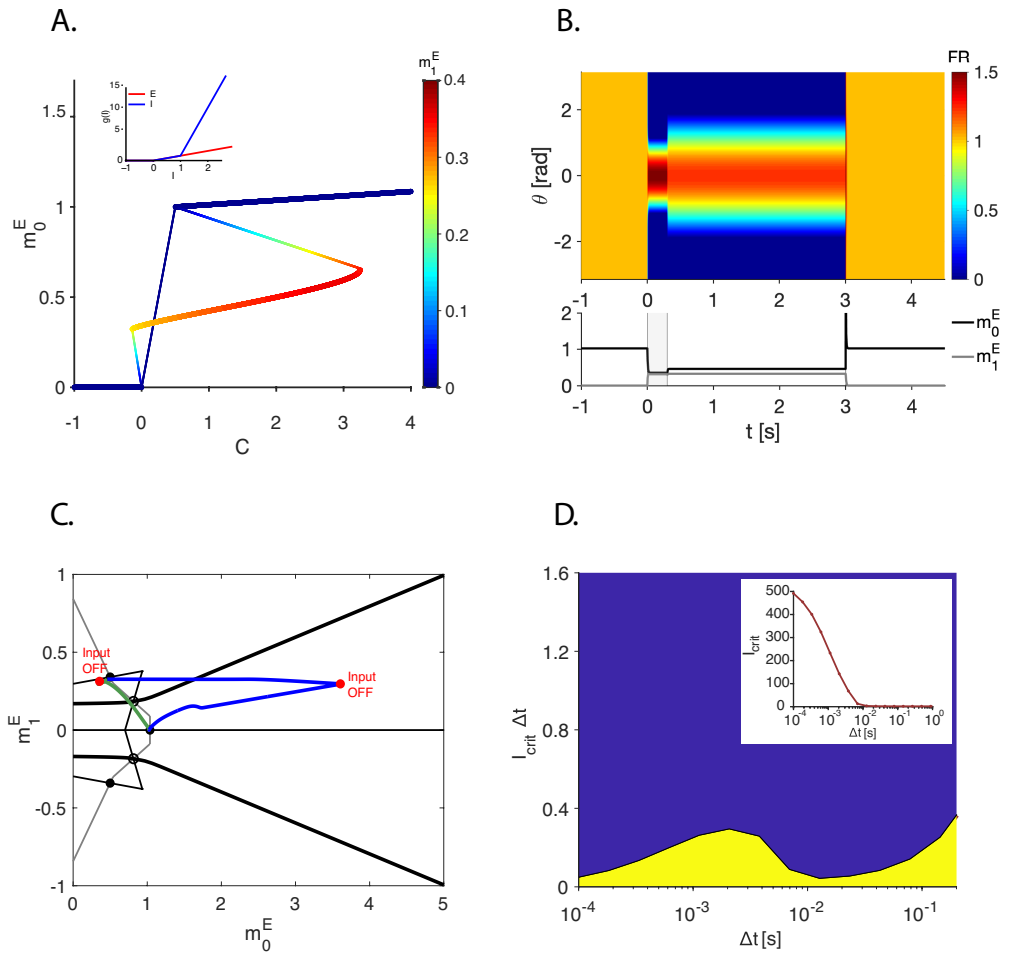

Figure 3.5. Two population mechanisms to terminate the bump activity through global excitation. A. Bifurcation diagram of m_0^E-C in the two population network in the acceleration difference case. Color code: first Fourier component of the activity, m_1^E. Inset: the TFs of the excitatory (red) and the inhibitory (blue) neurons. **B.** Dynamics of $m_E(\theta)$ (top, color code) and the time course of m_0^E and m_1^E (bottom) in response to a $0.3\,s$ tuned input at $t = 0\,s$ centered at $\theta = 0$ and a $0.3\,ms$ homogeneous pulse a $t = 3\,s$. **C** Dynamics of m_0^E and m_1^E from **B** plotted on the m_0^E-m_1^E phase plane. Green: response to the switch-on input. Blue: response to switch off input. For further details see caption in Figure 3.2. **D**. Minimal input charge ($I_{crit} \cdot \Delta t$) needed to perform the switch off transition plotted against the duration of that input, Δt. In yellow, the area where the switch off cannot be performed. In blue, the region where the switch off can be performed. To be compared with Figure 3.3 C red curve. Inset: Minimal input I_{crit}, as a function of its duration Δt_{stim}. To be compared with Figure 3.3 D red curve. Parameters used in this figure: $J_E^0 = 5$, $J_I^0 = 4.5$, $J_E^1 = 5$, $J_I^1 = 1.1$ and $\beta_I = 10$. $\tau_E = 5$ ms and $\tau_I = 1$ ms and $C = 1.5$ (in **B,C,D**)

3.2.4 Effective saturation within a network of spiking neurons.

Towards a more realistic model providing time courses of synaptic interactions, subthreshold dynamics and spatiotemporal spiking patterns, we developed a two population spiking network model of current based leaky integrate and fire neurons with structured connectivity (Figure 3.6 A and B). The model consists in two populations of leaky integrate-and-fire neurons ($N_E = N_I = 12000$) connected all-to-all through a "Mexican's hat" connectivity profile (Figure 3.6 B). We take the excitatory to excitatory synaptic weight profile ($J_{EE}(\theta)$) to be the same as the excitatory to inhibitory ($J_{EI}(\theta)$). And the same applies for the inhibitory population (i.e. $J_{II}(\theta) = J_{IE}(\theta)$). As in the rate model, each neuron is labeled by its PD, and the evolution of the membrane potential $V_X(\theta, t)$ in the subthreshold regime for a neuron in the postsynaptic population $X \in \{E, I\}$ is described by:

$$C_X \dot{V}_X(\theta, t) = g_l^X (V_X(\theta, t) - v_l^X) + I_{rec}^{EX}(\theta, t) - I_{rec}^{IX}(\theta, t) + I_{bg}(\theta, t) + I_{stim}(\theta, t) \tag{3.10}$$

where C_X is the membrane capacitance, g_l^X and v_l^X are the leak conductance and the membrane reversal potential, respectively. I_{rec}^{YX} is the recurrent input from presynaptic neurons in population Y to a postsynaptic neuron in population X where $X, Y \in \{E, I\}$. I_{bg} is a background input representing synaptic noise and I_{stim} is a stimulus dependent input. Whenever the voltage membrane V_X reaches the firing threshold v_t^X a spike is emitted and the membrane potential is reset to v_r^X, without a refractory period. For more details about the parameters, recurrent and background input Appendix B.

In this model, the effective saturation can also be implemented by the same mechanisms as in the rate model: acceleration difference, threshold difference or the combination of both (Figure C.6). Since in the spiking model we cannot choose explicitly the TF, we shape the neuronal TF by choosing specific parameters of background noise and membrane time constant. In general, the external noise shapes the expansive nonlinearity close to the firing threshold of the TF

(Hansel and van Vreeswijk, 2002) while the membrane time constant shapes the acceleration or slope of the TF. Likewise, the membrane time constant is controlled by the intrinsic properties of the neuron such as the membrane capacitance and the leak conductance ($\tau_m^X = C_X/g_l^X$). We find the parameters of external noise and membrane capacitance which provide a threshold linear TF for the excitatory neurons and an accelerating TF for the inhibitory neurons, both with the same activation threshold (Figure 3.6 C, inset). By injecting an external input, increased and decreased adiabatically, we explore all the stable states recreating the rate bifurcation diagram for the excitatory activity in Figure 3.5 (Figure 3.6 C). As in the rate model with acceleration difference, we found a bounded region of the external input (I_{ext}) for which the bump state exists. In addition, there is also a region of coexistence between the bump state and a homogeneous state which arises at larger values of external input. The main difference with the bifurcation diagram of the rate model is that the spiking network displays smoother transitions between the steady states due to the smoothness of the TFs nonlinearities. By simulating the bifurcation diagram at different values of the inhibitory spatial modulation parameter of the connectivity (J_1^I), we obtain the $J_1^I - I_{ext}$ phase diagram (Figure 3.6 D). We found qualitatively the same states configuration as in the rate model in the acceleration difference case (**??**), although in the rate model the boundary lines separating the bump and homogeneous states and the bistable regime from the bump alone are straight lines. The differences emerging in the spiking network could have its origin in the noise and the use of a continuous TF in contrast to the rate model with no noise and piecewise linear TFs. Same analysis for threshold difference and the combination of both mechanisms can be found in **??**.

3.2.5 Oscillatory dynamics govern the steady state transitions

Setting the network to operate into the bistable regime, the homogeneous state can be destabilised by applying a tuned input eliciting a transition towards the bump state. On the other hand, the bump state can be terminated by applying a long global excitatory input as expected from the bifurcation diagram in Figure 3.6

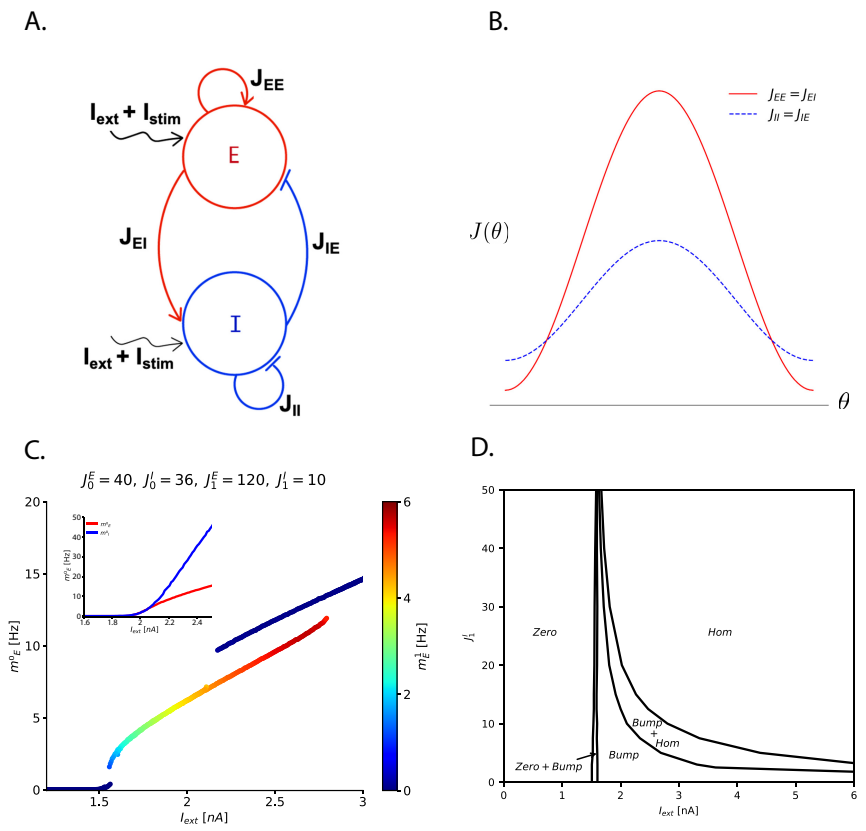

Figure 3.6. Two population network of leaky integrate and fire neurons operating in the acceleration difference model. **A.** Schematic diagram of the two populaition spiking network model. Inhibitory neurons ($N_I = 12000$) in blue and excitatory neurons in red ($N_E = 12000$). Inhibitory and excitatory recurrent inputs depicted by a blue flat arrow and a red arrow, respectively. **B.** Connectivity profile of neuron with PD $\theta = 0$. Inhibitory and excitatory connectivity is depicted in dotted blue and red, respectively. For simplicity we consider $J_{EE} = J_{EI}$ and $J_{II} = J_{IE}$. **C** Bifurcation diagram of m_0^E-C in the two population spiking network in the acceleration difference case. Color code: first Fourier component of the activity, m_1^E. Inset: the TFs of the excitatory (red) and the inhibitory (blue) neurons. **D.** $J_1^I - C$ phase diagram describing all the stable steady states that can be found at different values of the angular modulation of the connectivity J_1^I and C.

C (Figure 3.7 A). Likewise, and as seen in the two populations rate model, the B-H transition can also be mediated by a short strong homogeneous input which

restores the homogeneous state by synchronising the network and inducing a subsequent silent period (Figure 3.7 B). The simulations of the spiking network show that during the homogeneous state the mean firing rate of the excitatory population is around 12 Hz, and 30 Hz for the inhibitory population (Figure 3.7 A top and bottom panels, respectively) while the rate of the excitatory neurons in the center of the bump increases up to 25 Hz, and 60 Hz for the inhibitory neurons. Notice that , in all cases, during the bump state, the population mean firing rate (m_0^X) decreases respect to its value in the homogeneous state but the individual firing rate of neurons in the bump increases by 2-fold approximately.

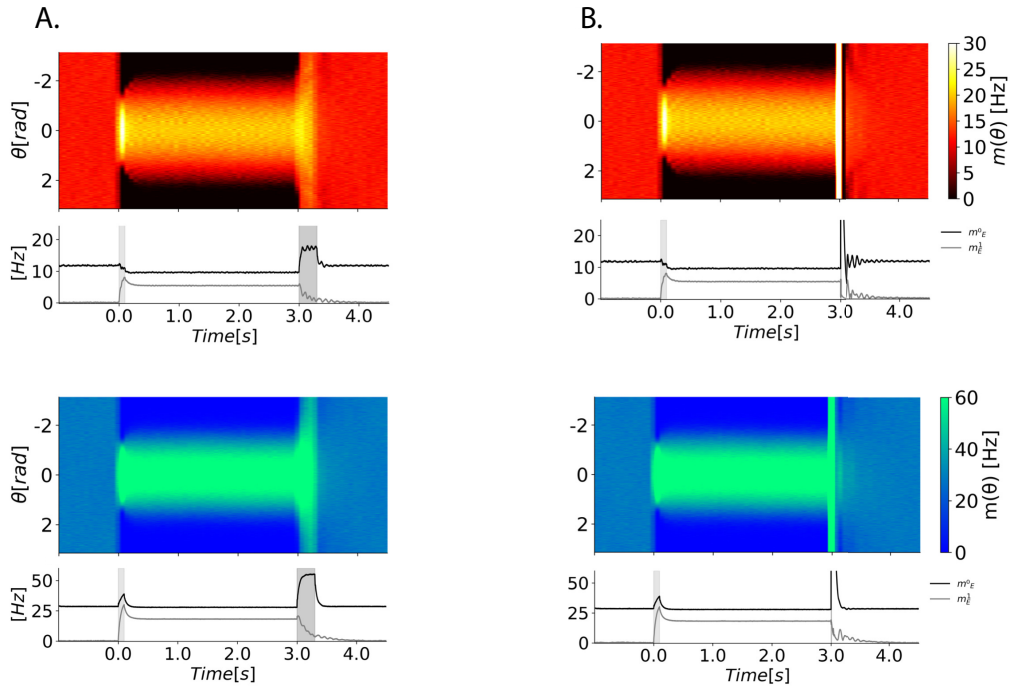

Figure 3.7. Dynamics of the network of leaky integrate and fire neurons. A. Dynamics of the network subjected to an ODR stimulation protocol: tuned input with duration $0.3\,s$ centered at $\theta = 0$ at $t = 0\,s$ and a homogeneous input with duration $0.3\,s$ at $t = 3\,s$. Top pannel: mean firing rate for each excitatory neuron computed in windows of $75\,ms$ and below the time evolution of the activity components m_0^E and m_1^I. Bottom: same for the inhibitory population. **B.** Same as in **A** but with an homogeneous input of duration $10\,ms$ at $t = 3\,s$.

One striking difference of the spiking model regarding the rate model is that the trajectories of the activity components m_0^X and m_1^X of the spiking network

describe oscillatory patterns in the vicinity of the steady state transitions, most notably in the B-H transition. The origin of these oscillatory patterns is attributable to the interaction amid several dynamical processes with different time constants. Explicitly, in this model, together with the structured connectivity, the voltage membrane and the interaction between excitatory and inhibitory recurrent synaptic inputs operating at different time scales account for the oscillatory spatiotemporal patterns. In order to unravel how the voltage membrane and the recurrent synaptic input dynamics interact giving rise to oscillatory patterns we studied the spike traces of the excitatory population close to the B-H transition (Figure 3.8). When a global external input is applied during a long period (e.g. 300 ms), the mean firing rate of the population (m_0^X) increases rapidly and fluctuates around a constant value. On the other hand, the spatial selectivity (m_1^X) decays to zero also depicting a fluctuation (Figure 3.8 A top). Notice that these two parameters describe a low-pass filtered trace of the population spiking activity. Taking advantage of working with a spiking model, we can display the spike raster plot of the excitatory population during the transition period (Figure 3.8 A bottom). From the raster, one can see a sequential activation (SA) of all the population which is initiated by certain neurons at the center of the bump and is extinguished after the activation of neurons with PD opposite to the bump. Before the first SA is finished, a group of neurons from the center of the bump starts the next SA cycle. In the time between two SA onsets certain neurons belonging to the bump reactivate. To identify which neurons are implicated either in the SA or bump reactivation, we track neurons which spike at certain times referred to the stimulus onset. Units which spike 5 ms before the stimulus onset are implicated in the initiation of the subsequent SA (Figure 3.8 A bottom). Since these units spiked right before the stimulus onset, their voltage membrane is close to the reset potential and needs around 50 ms to reach the spiking threshold, matching the period of the SA. Tracking the units which have spike 1, 10 and 20 after the stimulus onset we can see that they maintain the same order of activation in the following SA but eventually this sequence fades over time due to the recurrent interaction and the background noise. Tracking back the first neurons to activate after the stimulus onset (at 1 ms after) we found that they are neurons from the bump which spiked around 50 ms before. Our hypothesis to explain this pattern is that neurons belonging to the bump

which have not spiked lately, have been receiving stronger excitatory inputs from nearby neurons which have spiked thus, when the stimulus is applied, these neurons with the voltage membrane close to the firing threshold are the first to spike. These neurons will be followed by other neurons more distal from the center of the bump which, due to the structured connectivity, have been receiving smaller recurrent excitatory input and will need slightly more time to integrate the stimulus and fire. Thus, the recurrent synaptic input which is dependent on the structured connectivity would be the origin of this SA pattern. Moreover, the role of the voltage membrane dynamics, together with the broad inhibition, is to create an effective refractory period which depends on the intensity of the stimulus. Neurons which have spiked a few milliseconds ($< 50\,ms$) before the stimulus onset and hence belonging to the bump, won't be able to spike at the very beginning of the stimulation despite they have been receiving strong inputs from nearby neurons. Because they have spiked recently, their voltage membrane will be close to the reset potential at the stimulus onset. As previously stated, these groups of neurons will reactivate the bump activity during the progression of the SA (Figure 3.8 A bottom, e.g. pink dots). We validate this hypothesis by analyzing the effect of shuffling the voltage membrane values of all the neurons at the onset of the stimulus and by setting to zero the recurrent inputs during the first half of the duration of the stimulus (Figure 3.8 C top and bottom, respectively). The shuffling of V_m removes the reactivation of the bump during SA and then, neurons from the bump are the firsts to spike when the stimulus is applied since the shuffling removes temporarily the effective refractory period created by the membrane voltage dynamics. Another consequence is that the SA becomes prolonged in time although the period of the onset is still the same as in the control situation.This causes a desynchronisation of the activity created by the recurrent inputs from the inhibitory population. By setting to zero the recurrent inputs the spiking pattern results in an alternation between a SA, which is barely a synchronous activation, and the bump reactivation. Neurons have lost their synaptic memory, hence the effect of the structured connectivity is somehow lost. The dynamics is entirely driven by the membrane voltage dynamics. This results in an almost synchronous activation where neurons far from the bump fire just a few milliseconds after, because their V_m is lower than that of some neurons in the

bump. The absence of inhibition results in a regular and more frequent spiking pattern where the SA period has almost doubled. After 150 ms from the stimulus onset, the recurrent inputs are restored and inhibition increases the SA period but the information from the previous bump completely fades out. In addition, the broad inhibition introduces variability in the synaptic inputs of excitatory neurons resulting in a less synchronised SA.

The scenario for short stimulus although seems different at a glance, they share similar features. The mean firing rate (m_0^X) increases rapidly in steps and reaches a maximum value at the end of the stimulation. Since the duration of the stimulus is shorter than the time constant of the low-pass filter used to compute the firing rate, the values cannot reach a constant value. On the other hand, the spatial selectivity (m_1^X) displays a frequency modulated oscillation which frequency diminishes and cease to oscillate when the stimulus is withdrawn (Figure 3.8 B top). The raster plot shows a spiking synchronization right after the stimulus onset followed by a reactivation of the bump. First neurons to activate following the arrival of the stimulus (Figure 3.8 B, bottom; 0.2 ms from onset) are those which have fired about 40 ms ago together with the neurons more distal to the center of the bump (almost silent). The ones that participate in the reactivation of the bump are those neurons which have spiked 1-3 ms before the stimulus onset (Figure 3.8 B bottom, green and yellow). In contrast with the long stimulus case in which the stimulus intensity was set to 1 nA, to elicit the B-H transition through a short stimulus, the intensity has to be larger (50 nA in Figure 3.8 B and D). In this case, the external stimulus prevails over the recurrent input obtaining a similar scenario as setting to zero the recurrent interaction. The periodic alternation between homogeneous activation and bump states is, as in the previous example, due to the heterogeneities of the voltage membrane. When shuffling V_m the bump reactivation is completely lost and a SA pattern arises at the end of the stimulation due to the small effect of the recurrent synaptic activity which has slow dynamics compared to the stimulus duration (Figure 3.8 D top). The withdrawn of the recurrent inputs intensifies the effect of the external stimulus and, since inhibition is shut down, the frequency of the spiking pattern is doubled and the homogeneous activation becomes even more synchronous (Figure 3.8 D bottom).

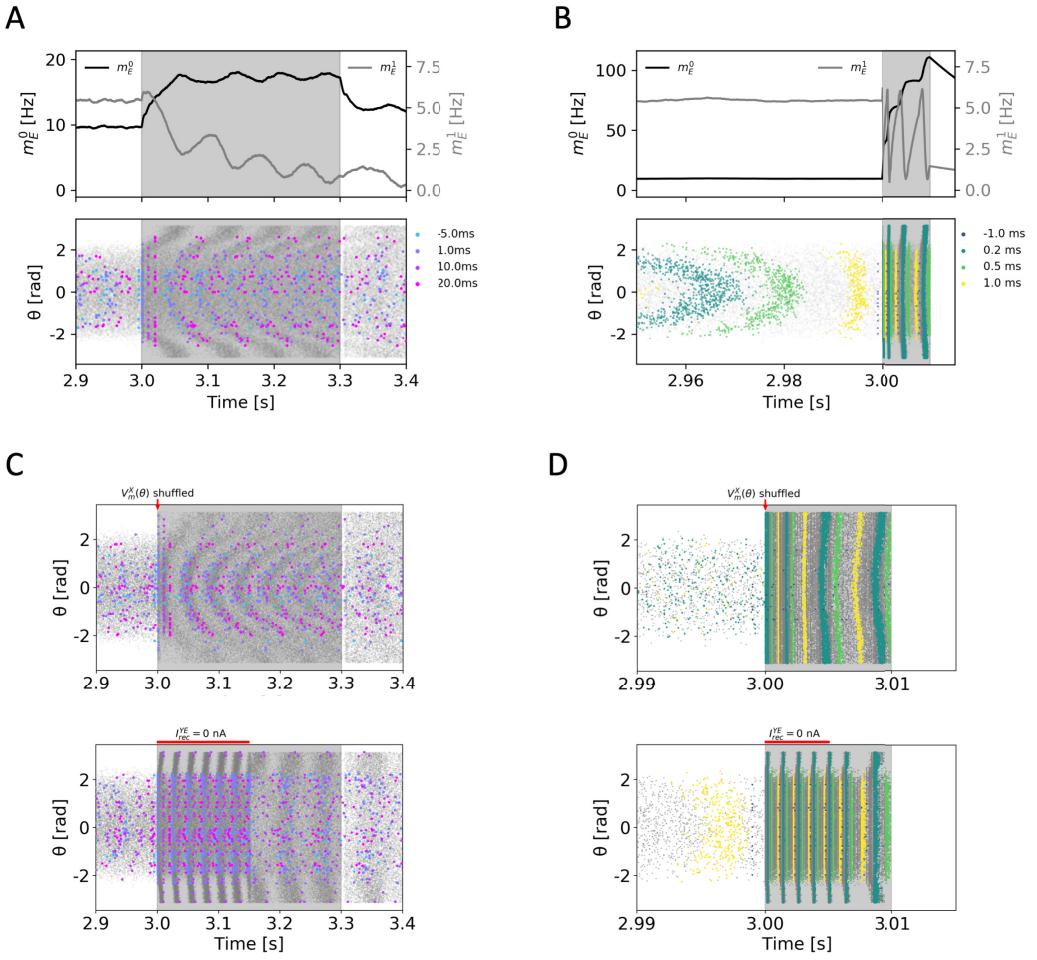

Figure 3.8. Oscillatory patterns in the steady states transition arise from the interplay between the membrane and the synaptic current dynamics. A. Detail of the B-H transition in Figure 3.7 A top (excitatory population). Top: time course of the activity components m_0^E and m_1^E. (black and grey, respectively). Bottom: temporal raster plot. Neurons spiking at -5, 1, 10 and 10 ms after the stimulus onset are tracked by colored spikes. **B.** Same as in **A**for the short stimulus in Figure 3.7 B top. In this case, the neurons spiking at -1, 0.2, 0.5 and 1 ms are tracked. **C** Temporal raster plots of the B-H transition. Top: Shuffling of the membrane voltage at the stimulus onset disrupts the bump reactivation. Bottom: Setting to zero the recurrent inputs to the excitatory population during the first half of the stimulus (red line) attenuates the sequential activation pattern. **D**. Same as in **C** for the short duration stimulus.

Now that we understand the origin of the oscillations arising in the B-H transition by a global excitatory stimulus, one might ask if the final state of the oscillation at the end of the stimulus affects the transition. In other words, will the transition depend whether the network is in a bump reactivation or in a SA at the end of the stimulus? In order to answer this question we explore the stimulus charge needed to elicit the B-H transition through a global excitatory input at different stimulus duration as in Figure 3.5 D (Figure 3.9 A). Surprisingly, we found that, for a certain stimulus duration between 10 and 100 ms, the intensities of the stimulus that allows the B-H transition are quantified in regions (Figure 3.9 A, blue areas). These regions are separated by the yellow areas representing the regions of charge at which the B-H transition cannot be performed (Figure 3.9 A, yellow areas). We simulate the network transitions subjected to a certain stimulus duration and intensity and studied the oscillatory pattern in the $m_0^E - m_1^E$ activity space as well as the spiking pattern. For stimulus duration in the range of 10 to 100 ms, if the stimulus is withdrawn when the network is in a homogeneous state ($m_1^X \simeq 0\,Hz$) the transition to the homogeneous state takes place. Conversely, if the network was in a reactivation of the bump state at the end of the stimulus (i.e large m_1^X) the transition cannot be performed. As in the rate model, two different mechanisms account for the B-H transitions: transitions mediated by stimulus duration shorter than the time constant of the system rely on the dynamics whereas transitions mediated by longer stimulus duration rely on the steady state configuration. To explain the quantification of the transitions for these relative short stimulus durations, we can use the analysis of the $m_0^E - m_1^E$ activity space in the rate model (Figure 3.5 C). From it, we can infer that in the spiking model there should also exist a separatrix between the bump and homogeneous BOA, and the crossing of this separatrix will determine the materialization of an effective steady state transition. Since in the spiking model the trajectory of the activity under the action of external stimulus depicts an oscillatory behavior respect to the spatial modulation m_1^X, its value at the end of the stimulus will determine the success of the transition. For very large stimulus intensity, above the top yellow area, the transition is always taking place regardless on m_1^X. This can also be explained by the shape of the separatrix in the $m_0^E - m_1^E$ activity space in the rate model that grows linearly as the population firing rate m_0^X increases (Figure 3.5 C). Under a very strong stimulation, m_0^X will increase and since the

oscillation in m_1^X is restricted to a span between 0 and a certain maximum value, eventually the trajectory will cross the separatrix (Figure 3.9 B, right). In the case that the stimulus duration is larger than the time constants of the system, the transition is always achieved provided that the network receives a stimulus with intensity above a certain constant value determined by the steady states. In Figure 3.9 A the intensity is normalized by the duration, thus representing the total charge transferred into the network which grows linearly with the increase of the duration implying that the critical intensity is a constant value. This is expected since, as well as in the rate model, starting in a certain I_{ext} in the bifurcation diagram in Figure 3.6 C, to perform the transition from bump to homogeneous state the current needed is fixed. Moreover, for stimulus duration shorter than 10 ms, in general, the oscillatory pattern is not able to cross the separatrix multiple times thus there only exists a value of input intensity above which the B-H transition occurs. Notice that considering only the lower yellow region in Figure 3.9 A the result is qualitatively the same as in the rate model (Figure 3.5 D).

Worth mention that these oscillatory pattern is attenuated as the synaptic decay time constants (τ_d^X) are increased well above the value of the membrane time constant ($\tau_m^E = 40\,ms$ and $\tau_m^I = 10\,ms$). We expect that in the limit where the synaptic time constants are much larger than the membrane time constant the spiking network will behave as the rate model since in the rate model we assumed that the membrane dynamics is much faster than the synaptic dynamics (Figure C.7).

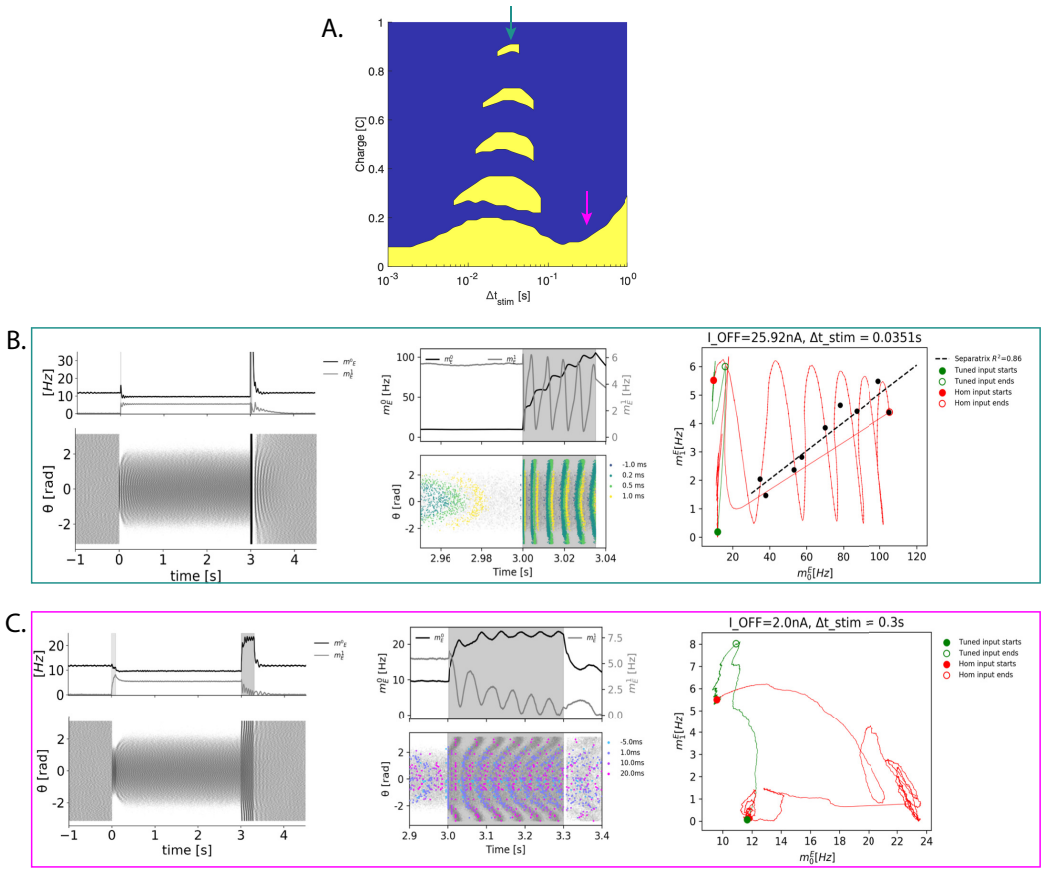

Figure 3.9. Oscillatory dynamics determines the termination of persistent activity. A. Input charge for the bump to homogeneous transition as a function of stimulus duration (Δt_{stim}). As in Figure 3.5 D, yellow areas depicts the values at witch the transition cannot be performed while in blue the values at which the B-H transition is performed. **B.** Analysis of the dynamics for stimulus duration of 35 ms. Left: Time course of m_0^E and m_1^E and the temporal raster plot of the excitatory population. Center: As in Figure 3.7 top, detail of the activity dynamics and spikes in the peri-stimulus period (stimulation period shaded in gray). Right: Projection of the activity trajectory on the $m_0^E - m_1^E$ phase plane. In green the trajectory for the switch-on and in red the trajectory of the switch off (B-H transition). Estimated separatrix by plotting the end of the trajectory at the time in which the stimulus is terminated for the values of charge at the boundaries of the yellow areas. **C** Same as in **B** for a 0.3 s stimulus duration.

3.2.6 Oscillatory dynamics in more realistic spiking model.

At this point we wonder whether this oscillatory dynamics will appear in a more realistic model in which the membrane and synaptic parameters are constrained by biological data. To address this question, we used membrane parameters from McCormick et al. (1985) and Tripathy and Gerkin (2012). On the other hand, excitatory synaptic recurrent inputs are a combination of the contribution of NMDA and AMPA currents and the ratio of the contribution of NMDA and AMPA to the postsynaptic current is different in the E-I and in the E-E synapses (see Appendix B.3.3 for more details). In this new model, we obtained an acceleration difference between the excitatory and inhibitory TF by changing the background noise (I_{bg}) and the reset voltage after the spike (v_r^X). For a given set of connectivity parameters we found a bistable region between an homogeneous and a bump state (Figure 3.10 A). We simulated the network dynamics subjected to a short switch off stimulus and we analyzed the trajectory in the $m_0^E - m_1^E$ space (Figure 3.10 B, C). We found qualitatively similar oscillatory trajectory in the $m_0^E - m_1^E$ as in the model of the previous section. We also compute the diagram of the input charge as a function of the stimulus duration and found regions in which the bump to homogeneous transition does not take place (Figure 3.10 D, yellow). As in the previous model, an increase of the synaptic decay time constant will attenuate the oscillatory pattern and thus the yellow areas in the charge *VS* stimulus duration will fade out since the trajectory will tend to cross less times the separatrix (Figure C.8)

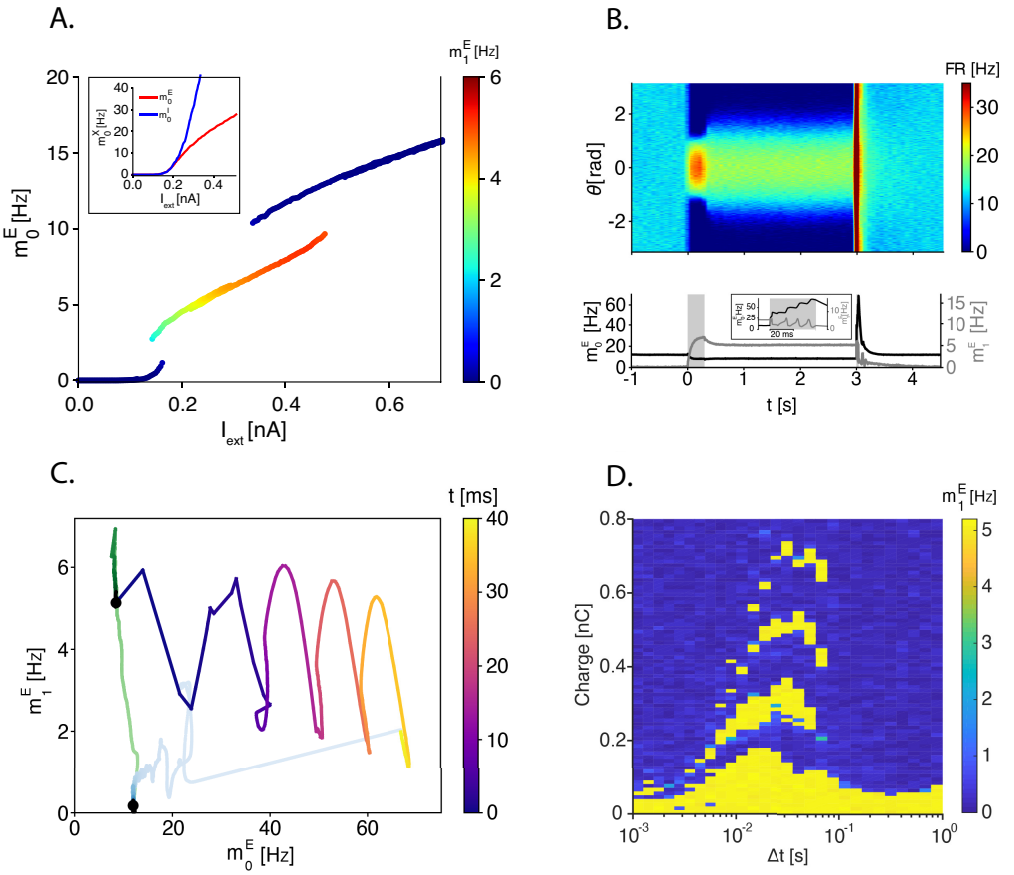

Figure 3.10. Oscillatory dynamics in a biological realistic spiking network model. A. Bifurcation diagram of m_0^E (y-axis) and m_1^E (color code) with I_{ext} as bifurcation parameter. Obtained by simulated annealing. Inset: TFs of the excitatory (blue) and inhibitory (red) neurons. **B.** Dynamics of the excitatory population of the LIF network subjected to a 300 ms tuned input at t=0 s and a 40 ms homogeneous input at t=3 s. Parameters as in figure **??**. Top: average firing rate of the neurons calculated in 75 ms time bins. Bottom: Dynamics of the zero- and first order Fourier components of the average firing rate (m_0^E and m_1^E, respectively). Inset: Detail of the response to the 40 ms homogeneous input at t=3 ms. **C** Trajectory of m_0^E and m_1^E from A plotted on the $m_0^E - m_1^E$ phase plane. Green: trajectory during switch on ($0 < t < 0.3$ s). Time course of the trajectory during 40 ms switch off color coded ($3 < t < 3.04$ s). Light Blue: trajectory during relaxation after the 40 ms stimulus. **D.** Stimulus input charge plotted against the duration of that input. Regions in blue represent where switch off is performed and regions in yellow where the switch off is not performed. Model parameters are specified in Appendix B.6.1, Tables B.3 and B.4.

3.3 DISCUSSION AND CONCLUSION

Experimental data from different animals performing working memory tasks shows that the extinction of a memory encoded as a persistent selective activity is coupled by a broad increase of the spiking activity (Funahashi et al., 1989, 1990, 1991; Inagaki et al., 2018, 2019). This pattern can be considered as a hallmark of the transition between two network states.

In this work we showed that selective persistent activity can emerge in a single recurrent network with specific connectivity. With sufficiently strong and spatially modulated recurrent excitation the network dynamics possess a continuous set (ring) of attractors. In these attractors the activity is ‘‘bumpy’’ and encodes for the cue direction. In order to account for the selective persistent activity observed experimentally, one must find the conditions under which a ring attractor can coexist with a homogeneous activity state. Such multistability requires appropriate nonlinearities. Using a simple modeling framework we investigated in a systematic manner, several mechanisms based on different f-I curve nonlinearities. We show that they differ by their bifurcation structures, the roles played by excitation and inhibition, as well as on the dynamics of the memory erasing after the delay period. We have shown how compressive nonlinearities in the neuronal transfer function as well as an effective saturation mechanism, that arise from the interaction between two populations, allows the extinction of the bump state through global excitation. Furthermore, we have implemented the effective saturation mechanism in a network of spiking neurons and found that, due to the intrinsic oscillatory dynamics not provided by the rate model, the transitions between the bump and homogeneous steady states are quantified for a certain region of the stimulus duration.

A simple and intuitive mechanism which can underlie the reset of the persistent activity at the end of the delay period, is a global inhibition of the excitatory neurons triggered by direct inhibition of the excitatory cells or by feedforward inhibition (e.g. Hansel and Mato (2001); Wang (1999); Roudi and Latham (2007)). Physiologically, it is plausible that in DR tasks, either the response, the go cue or the reward trigger an increase in inhibition. In experiments, however, many neurons display a transient increase of their firing rate at the end of the delay

period, even when the direction of the cue is away from the center of their tuning curve Funahashi et al. (1989); Constantinidis et al. (2001). This suggests a global excitation of a large fraction of these neurons. Indeed, these could be due to inhibitory neurons receiving excitation from outside the network, but nonetheless it is valuable to explore other possible switch off mechanisms. The origin of the activity increase in the PFC associated with the extinction of persistent activity is still a matter of discussion. One hypothesis links this global increase of spiking activity to feedback-loop activity from subcortical areas related to the saccadic responses (Alexander et al., 1986). Neurons from substantia nigra show activity suppression during saccades, leading to a disinhibition of the superior colliculus which triggers the saccadic response (Hikosaka and Wurtz, 1983). Suppression of nigral neurons also leads to an disinhibition in the thalamic activity that could excite PFC neurons (Buee et al., 1986; Guo et al., 2017). Apparently, this signature is restricted to rewarded saccadic responses related to the task since spontaneous saccades in the inter-trial intervals did not show a marked increase of the spiking activity (Funahashi et al., 1991). In Funahashi et al. (1989) it is shown how the delay period activity from both excitatory and inhibitory populations is terminated immediately after the saccadic response. All these evidence uphold the role of the post saccadic activation of neurons in PFC in the termination of the persistent selective activity.

In this work we describe two different mechanisms for switching-off the persistent state by global excitation. For stimulus duration much larger than the synaptic timescale, the switch off relies on the bifurcation structure. Conversely, for stimulus duration shorter than the synaptic timescale, the switch off rely on the dynamics and specifically on the activity trajectory crossing of the separatrix.

The mechanism mainly discussed involves switch off by a global excitatory pulse. This can only be implemented when the network response to input is saturating. In the case where the excitatory and inhibitory TFs have a similar shape, this requires the mechanism to make use of the saturating part of the TF. This means that during the excitatory pulse all the neurons are driven to the saturating part of the TF; there, the efficacy of recurrent excitation close to the center of the bump is no longer sufficient to fend off the inhibition arriving from the rest of the network, even though the activity around the center is higher than in the

edges, and the bump is obliterated. When the excitatory and inhibitory TF differ from each other, the decrease in the relative efficacy of the excitation can arise not because neurons arrive at the saturating part of the TF, but from a rise in the efficacy of the inhibition. In the model of Compte et al. Compte et al. (2000) the bump was also terminated by a homogeneous excitatory pulse, and saturating NMDA synapses in this model could potentially be the source of the increase in inhibition efficacy which causes the bump extinction. A similar mechanism for memory erasure through global excitation is thoroughly described in a network dominated by inhibition in Brunel and Wang (2001).

The other discussed mechanism uses a short and strong stimulus of excitation delivered to the whole network in order to extinguish the bump state. This stimulus causes a transient increase in the network average activity, which is followed by a transition back to the homogeneous state. In terms of dynamics, the mechanism underlying this transition relies on the activity trajectory crossing the separatrix in the phase space. Notice that this mechanism allows the switch off regardless of the nature of the nonlinearity in the TF and thus the bifurcation structure. In the spiking network model studied in Gutkin et al. (2001); Laing and Chow (2001), a brief global excitation resets the activity back to its baseline level. This is because the brief pulse causes simultaneous firing in all the neurons. All the neurons enter then into a refractory period where they are unable to respond to inputs. Since the duration of this refractory period is larger than the time constant of the synapses (they use fast AMPA-like synapses), when the neurons are recovered from the refractory period, the synaptic memory trace has faded. However, this is not the case in the model studied here, since the spike dynamics are not captured in the rate model and in the spiking model described in our work, the timescale of the recurrent excitation is larger than the effective refractory period.

Another way to switch off the selective persistent state by a short stimulus is to use an anti-tuned stimulus (Figure C.2). In this case the input will create a bump of activity in the opposite direction making the activity trajectory to approach the lower half-plane in the phase space. If this input is withdrawn when the trajectory is somewhere inside the homogeneous BOA, the switch off will succeed.

The effective saturation mechanism proposed here is based on the fact that as the

background input increases the efficacy of the inhibition increases with regard to the excitation. That is, the gain of the inhibitory TF becomes larger than that of the excitatory. Thus, the interplay between the inhibitory and excitatory neuronal TFs leads to a saturation of two effective functions of the input resulting in a similar scenario as in the compressive nonlinearity in the single population model. A striking difference is that the effective saturation mechanism does not exploit the saturation part of the TF since the origin of the bistability emerges from the difference of the gain between excitatory and inhibitory TFs. This effective saturation mechanism is well supported by experimental data showing the differences between excitatory and inhibitory neuronal TFs. In fact, the architecture of the cortex and its dynamics rely on the interaction between excitatory and inhibitory neurons. This segregation of excitatory and inhibitory neurons is mainly based on differences in their spiking patterns which, in turn, depends on their intrinsic membrane properties. Neurons with a regular spiking pattern with adaptation are typically considered as putative excitatory neurons while neurons with a fast spiking pattern without adaptation are considered as putative inhibitory neurons (Connors and Gutnick, 1990). These differences in the firing pattern have a direct impact in the spike frequency response to an injected current i.e. the TF. Several experimental studies agree that excitatory neuron TF presents a smaller slope than that of the inhibitory (Nowak et al., 2003; Arsiero et al., 2007; Schiff and Reyes, 2012). Other less pronounced characteristics are that excitatory neurons have lower spiking threshold and the slope is mainly linear in contrast to the inhibitory neurons which present a larger spiking threshold and an accelerating slope (Schiff and Reyes, 2012). These evidences about the differences between the excitatory and inhibitory neurons give biological significance to the effective saturation mechanism which is a suitable substrate allowing the termination of persistent selective activity by global excitation. We have shown that either creating a difference in the acceleration or a difference in the threshold between excitatory and inhibitory TFs is sufficient to elicit effective saturation. In a more realistic scenario mimicking the experimental TFs, we also have shown that by creating both a difference in acceleration and in the threshold the effective saturation can be achieved. Alternative substrates disregarding intrinsic cell properties, can also induce an effective saturation giving rise to bistability. Short term synaptic

plasticity is an immediate candidate providing nonlinearities which are a source of bistability (Del Giudice et al., 2003; Romani et al., 2006; Mongillo et al., 2008; Hansel and Mato, 2013). In fact, the model in Compte et al. (2000) introduces saturating NMDA synapses which could be the origin of the bistability and it would explain the fact that the persistent selective activity can be extinguished by a global excitatory input. A possible explanation generated from our work is that the external homogeneous input induces the saturation of the NMDA synapses leading to an increase of the inhibition efficacy which will cause the extinction of the bump.

Our analytical investigations were performed in the rate-model framework (e.g. Ben-Yishai et al. (1995); Ermentrout (1998); Roudi and Latham (2007); Roxin et al. (2005); Wilson and Cowan (1972)). This framework can be derived from first principles from the dynamics of a spiking network model if one assumes that the single neuron spiking dynamics is much faster than the synaptic dynamics. In that case the firing rate of a neuron can be well approximated as an instantaneous function of its total input, namely, its f-I curve (see equation A.1). If one also assumes that the synaptic current induced by a presynaptic spike decay exponentially, the network follow the classical dynamics of a rate model in which the "rate" variables correspond to the synaptic inputs.

A consequence of this interpretation is that the behaviors of the rate model and the spiking model, which tend to be in good correspondence when the synaptic time constants in the spiking model are large, show some discrepancies when the synapses are not slow enough. In the model presented here, although the general properties of spiking network behavior matched those in the rate-model network, one such inconsistency occurred: in the spiking model, for an inhibitory synaptic decay time constant of 5 ms, the stability of the bump broke down unless the excitation time constant was 4-5 times as large, whereas in the rate model, equal excitation and inhibition time constants were sufficient for bump stability. It should be noted though, that this discrepancy was made much smaller when both inhibition and excitation time constants were enlarged regarding the membrane time constants in the spiking network, i.e. approaching to the assumption taken in the rate model. This destabilization of the bump has also been reported in the spiking models considered by Compte et al. Compte et al. (2000). In fact, in

Wang (1999); Compte et al. (2000) they show how slow and saturating NMDAR channels can stabilize a state of asynchronous persistent activity. Although this mechanism is supported by recent studies showing that persistent activity can be eliminated by blocking NMDAR channels Wang et al. (2013); van Vugt et al. (2020), it is not clear whether this is a result of the reduction of the synaptic time scales or in the overall level of excitation. In fact, in a study by Hansel and Mato Hansel and Mato (2001) is shown that slow excitation is not compulsory provided the inhibitory-to-inhibitory interactions are strong enough.

On the other side, the spiking network model is not restricted to this assumption and captures the interaction between the synaptic and membrane dynamics. As a direct result of this interaction, in the presence of an external input the network displays an oscillatory spatiotemporal pattern consisting in an alternation between synchronous activation and bump reactivation. Surprisingly, for input durations in the range of the time constants governing the network dynamics, the steady state transitions are quantified in specific regions of intensity of the input. For stimulus duration below the shorter than the smallest time constant, which in our model corresponds to the synaptic current decay of the inhibitory population (5 ms), the interaction between the membrane time constants (40 ms and 10 ms for the excitatory and inhibitory neurons, respectively) and decay synaptic time constant (40 ms and 5 ms for excitatory and inhibitory synapses, respectively) becomes less pronounced since the dynamics will be mainly determined by the fast synaptic timescales, therefore the amplitude of the m_1^X oscillation is not large enough to cross the separatrix multiple times. In the opposite situation, when the stimulus duration is long ($>$100 ms) the dynamics is strongly dominated by the steady states and the oscillations in m_1^X are attenuated by the slow dynamics thus, becoming irrelevant in the states transition. One possible hypothesis to explain this phenomenon is that the subcritical Hopf bifurcation that gives rise to bistability described in the bifurcation analysis gives birth to a limit cycle when an external input is applied and relaxes to a stable state when the input is removed. For specific duration of the external stimulation the frequency and amplitude of the limit cycle allows the multiple state transitions depending on the intensity of the stimulus. For very short inputs the amplitude of the limit cycle is reduced, while for long stimulus the frequency of the limit cycle diminishes. This hypothesis is supported by the

activity trajectories in the $m_0^E - m_1^E$ space, which amplitude in m_1^E is smaller for short stimulus and the frequency of the limit cycle oscillation is smaller for long stimulus. In addition, the relaxation to the homogeneous state once the stimulus is withdrawn depicts a typical trajectory of an unstable limit cycle that collapses to a stable fixed point. Similar oscillatory patterns are shown in citeLaing2017 and in Schmidt and Avitabile (2020) arisen from the different timescales orchestrating the dynamics of a network of excitatory and inhibitory populations. Several works have underlie the role of oscillatory brain activity in WM (Koene and Hasselmo, 2007; Jokisch and Jensen, 2007). While fast oscillations on the gamma band represents maintenance of memories, slower oscillations in the alpha band are related to memory clearance (Jokisch and Jensen, 2007). Previous findings in theoretical works support that oscillations in the $\alpha - \beta$ frequency range are related to memory extinction (Dipoppa and Gutkin, 2013; Schmidt et al., 2018). In fact in Schmidt et al. (2018), they describe a mechanism where these oscillations create transient periods of spike synchrony which destabilise the persistent activity. In our work we found qualitatively similar results where low frequency oscillatory dynamics showing a competition between synchronous activation and bump reactivation mediate the memory extinction. Further experimental research involving electrophysiological recordings from large populations of neurons in behaving animals together with optogenetic manipulation may be required in order to reveal the nature behind the formation and fading of transient memories. In summary, we propose novel mechanisms for termination of persistent selective activity through global excitation. Besides, we successfully implemented these mechanisms in a more realistic network of spiking neurons capturing the oscillatory dynamics neglected in the rate model. Interestingly, the model predicts that, for relatively short stimulus, these oscillations determine the extinction of the persistent activity. Taken together, this results shed light on the neural network mechanisms controlling the gating between brain dynamics in WM.

4 Models of Aberrant Prefrontal Cortex Dynamics

4.1 INTRODUCTION

In this Chapter we exploit the previous spiking network model describing the dynamics of the PFC during a WM task to study how dysfunctions in the network could lead to aberrant dynamics that could be the origin of cognitive impairments typically encountered in mental disorders. As we reviewed in section 2.4 the PFC seems to be one of the most affected regions in mental illness such as Schizophrenia among others. Thought disorder including hallucinations, and cognitive symptoms such as impaired WM and learning described in subjects with Schizophrenia are strongly associated with malfunction of the PFC. The PFC, in synchrony with sensorimotor areas and subcortical areas, is in charge to orchestrate the perception-action cycle by linking past sensory information and prospective behavior. Thus, alterations of the PFC function could potentially lead to an abnormal perception of the reality leading to hallucinations, as well as impairments in WM. One of the strong theories on the etiology of the Schizophrenia revolves around the a hypofunction of the NMDA receptors (NMDAR) (Gao and Snyder, 2013). The main evidences supporting this theory came from studies in which NMDAR antagonists (such as ketamine) could replicate the symptoms of patients with Schizophrenia in health patients (Krystal et al., 1999). Moreover, drugs enhancing the NMDAR function mitigate negative and cognitive symptoms. Besides, persistent blockade of NMDAR in animals

develop pathologic features of Schizophrenia such an impairment on the WM tasks (Aura and Riekkinen Jr, 1999; Driesen et al., 2013; Coyle, 2012). Additionally, neuroimaging and anatomical studies of Schizophrenic brains have described a decrease in the PFC volume as well as smaller cell bodies and dendritic spine loss in pyramidal neurons from layer III of the dlPFC (Hoftman et al., 2017). Layer III of the dlPFC is thought to be where the PFC microcircuits reside. This circuits are composed by networks of recurrent excitatory pyramidal cells strongly connected and exciting each other via AMPA and NMDA receptors on the dendritic spines. This circuit leads to persistent activity that is shaped by lateral inhibition from GABAergic cells creating the tuning properties seen in working memory experiments (Arnsten, 2011). The hypothesis of NMDAR hypofunction suggest that glutamate neurotransmitter no longer activates NMDAR receptors originating slow/long excitatory postsynaptic currents (EPSCs). Instead AMPAR, which have shorter decay time constant (≈2-5ms), would be the main contribution to the EPSCs (Rubio et al., 2012). In order to explore how different contributions of NMDA receptors could alter the PFC dynamics we perform an exploratory study analyzing the role the synaptic time constants on the spatiotemporal dynamics and on the stability of the steady states. We found that, for physiological values of the inhibitory synaptic decay time constant (≈5-10 ms in the $GABA_A$ receptor), the excitatory synaptic decay time constant has to be slower in order to avoid synchrony and oscillatory patterns of activity. As already noticed by Wang (1999), this results suggest that slow dynamics of NMDA plays an important role stabilizing the irregular firing present in both the homogeneous and bump state. In the second part we explore how the network size of the model affect to the stability of the steady states of the system. We found that a reduction in the number of neurons could lead to spontaneous transitions giving rise to the appearance of bumps of activity in absence of external stimulations. This results could suggest a mechanism for the origin of the hallucinations in the PFC circuits.

4.2 RESULTS

4.2.1 NMDAR hypofunction leads to aberrant spatiotemporal patterns.

We use the spiking model presented in section 3.2.4 where the time course of the postsynaptic currents (PSCs) are described with a double exponential function:

$$s_X(t - t_{sp}) = \frac{1}{\tau_d^X - \tau_r^X}\left[e^{-(t-t_{sp})/\tau_d^X} - e^{-(t-t_{sp})/\tau_r^X}\right] \tag{4.1}$$

where t_{sp} is the time of the spike and τ_r^X and τ_d^X are the rise and decay synaptic time constants of the population X, respectively. The normalization factor ensures that the total area of the PSC of an individual spike is equal to 1. In this model we consider a single pair of rise and decay time constant for the EPSCs that could be interpreted as an effective time constant accounting for a weighted contribution of NMDA and AMPA PSCs. While the τ_r^E will be fixed to 1 ms, τ_d^E will be modified from 2 ms to 128 ms. A total blockade of NMDA currents will be represented by a total contribution of the AMPA currents i.e τ_d^E =2-5 ms (Hestrin et al., 1990; Spruston et al., 1995; Angulo et al., 1999). On the other hand, a total blockade of AMPA will be represented by a total contribution of NMDA currents i.e τ_d^E =50-150 ms (Hestrin et al., 1990; Sah et al., 1990; Bear et al., 2020). For the inhibitory postsynaptic currents (IPSCs) we consider that the total current is due to $GABA_A$ receptor with physiological values of τ_d^I around 5-10 ms (Salin and Prince, 1996; Xiang et al., 1998; Gupta et al., 2000). Nevertheless, we are going to modify the τ_d^I values ranging from 2 to 128 ms in order to explore how a variation in the inhibitory synaptic currents could affect the dynamics.

In this section we use the network model with leaky integrate and fire neurons presented in the previous chapter. The model describes bistability due to the effective saturation mechanism (Figure 4.1). In this case, in contrast to acceleration difference showed in section 3.2.4 we used a threshold difference in which both TFs have the same gain but the excitatory TF have lower spiking threshold than the inhibitory (Figure 4.1 inset). At lower input the inhibitory population will activate and at a certain input value the inhibitory population will activate creating

an effective reduction of the excitation resulting in an effective saturation.

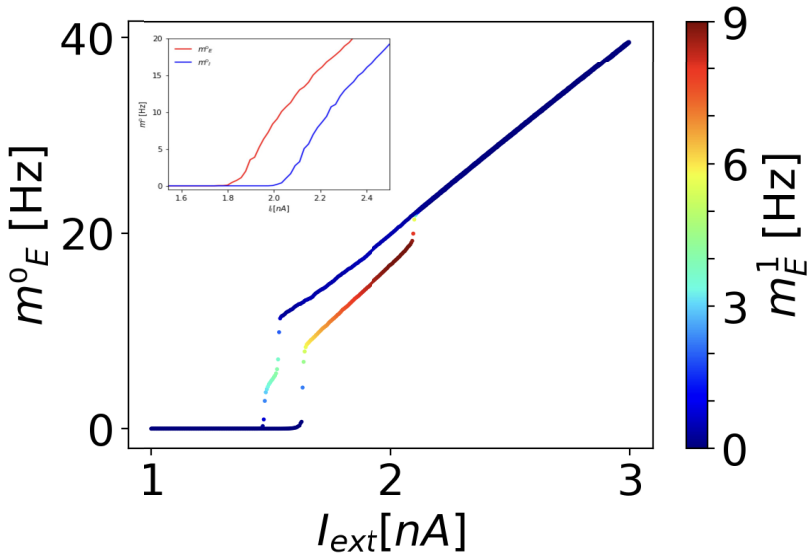

Figure 4.1. Bifurcation diagram of the threshold difference LIF model. Bifurcation diagram of m_0^E-C in the two population network in the threshold difference case. Color code: first Fourier component of the activity, m_1^E. Inset: the TFs of the excitatory (red) and the inhibitory (blue) neurons. Parameters as described in Appendix B.6.2.

We explore how different ratios of excitatory and inhibitory decay time constants affect the spatiotemporal patterns and the stability of the homogeneous and the bump state. To do so, we start to simulate the dynamics setting the network to work in the homogeneous or bump state and as the time progresses we change the decay time constants. In order to characterize the spatiotemporal dynamics we quantify the neuronal synchrony as described in Hansel and Sompolinsky (1992):

$$\chi(N) = \sqrt{\frac{\sigma_V^2}{\frac{1}{N}\sum_{i=1}^{N}\sigma_{V_i}^2}} \tag{4.2}$$

where σ_V^2 is the variance of the time fluctuations of the subthreshold membrane potential. This value is normalised by the population average of the variance of time fluctuations of single cell membrane potentials $\sigma_{V_i}^2$. The χ is thus a

normalised synchrony measure being 1 the fully synchronised state and small values of order $O(N^{-1/2})$ correspond to the asynchronous state.

In order to distinguish between traveling waves and standing bumps, which both show similar value of synchrony, we quantify the variation in time of the center of the bump:

$$\sigma_\theta^2 = \left\langle |\arctan\left(\frac{m_2(t)}{m_1(t)}\right)|^2 \right\rangle_t - \left[\left\langle |\arctan\left(\frac{m_2(t)}{m_1(t)}\right)| \right\rangle_t\right]^2 \tag{4.3}$$

where $\langle ... \rangle_t$ denotes a time average over a large time period and $m_1(t)$ and $m_2(t)$ are the first and second Fourier components described in sections 3.2.1 and 3.2.2, respectively. Small values of the variance of the bump argument indicates that the center of the bump is not changing in time pointing either to a oscillatory bump or a uniform bump while a large value will indicate that there is no bump (asynchronous uniform, travelling wave, standing wave, oscillatory uniform).

Another useful measure to characterize the spatiotemporal patterns is the mean spatial modulation of the activity that is computed as:

$$m_s = \left\langle \sqrt{m_1^2(t) - m_2^2(t)} \right\rangle_t \tag{4.4}$$

A large value will indicate that the activity is strongly modulated (bump state) while a small value will indicate that the activity is mostly uniform (homogeneous state). We set the network at the top homogeneous branch in a region of coexistence depicted in Figure 4.1 ($I_{ext} = 1.7\, nA$). The homogenous state displays an asynchronous uniform spiking pattern in the top right region above the white transition area represented by small values of the synchrony measure χ (Figure 4.2 A). Inhibitory synaptic decay time constant below 8ms show small values of σ_θ (in red) indicating that the homogeneous state destabilizes towards a bump state (Figure 4.2 C). For ratio τ_d^E/τ_d^I=1 there is a strong synchrony corresponding to an oscillatory uniform (OU) state for inhibitory time constant below 64 ms (Figure 4.2 A and Table 4.1). For τ_d^I=2 ms we found a very rich variety of patterns: for ratio equal to 1 we found a standing wave (SW) pattern while for ratio equal to 2 a travelling wave (TW) pattern arises (Table 4.1). Notice that in the boundary there is coexistence between

different spiking patterns.

Next, we set the network at the bump state in the region of coexistence (Figure 4.1 at $I_{ext} = 1.7\,nA$). The bump state shows an uniform bump (UB) activity pattern at the top right area of the phase diagrams (Figure 4.2 B and D). In the boundary region we found patterns of oscillatory bump (OB)) and in the bottom left corner both SW as well as TW as described in the homogeneous case. This result indicates that for small time constants the bistability no longer exists.

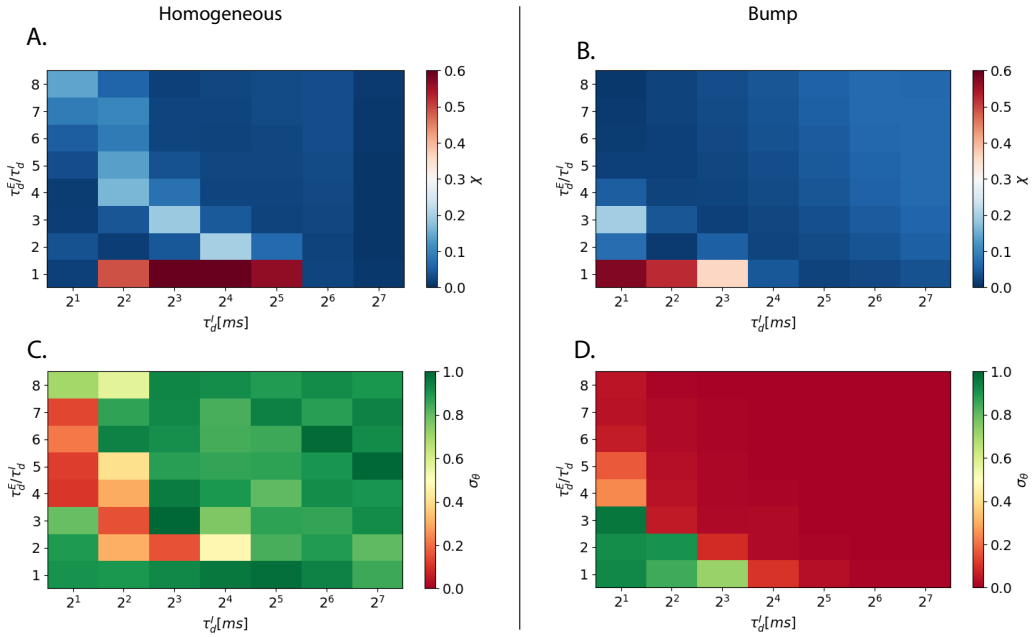

Figure 4.2. Phase diagrams of the synaptic decay time constants. A.and **B** Phase diagrams of the ratio τ_d^E/τ_d^I as a function of τ_d^I and color coded the value of (χ) for each condition. **A** for the homogeneous state and **B** for the bump state. **C** and **D** Phase diagrams of the ratio τ_d^E/τ_d^I as a function of τ_d^I and color coded the value of the standard deviation of the bump's argument (angle pointing the center of the bump) (σ_θ). **C** for the homogeneous state and **D** for the bump state.

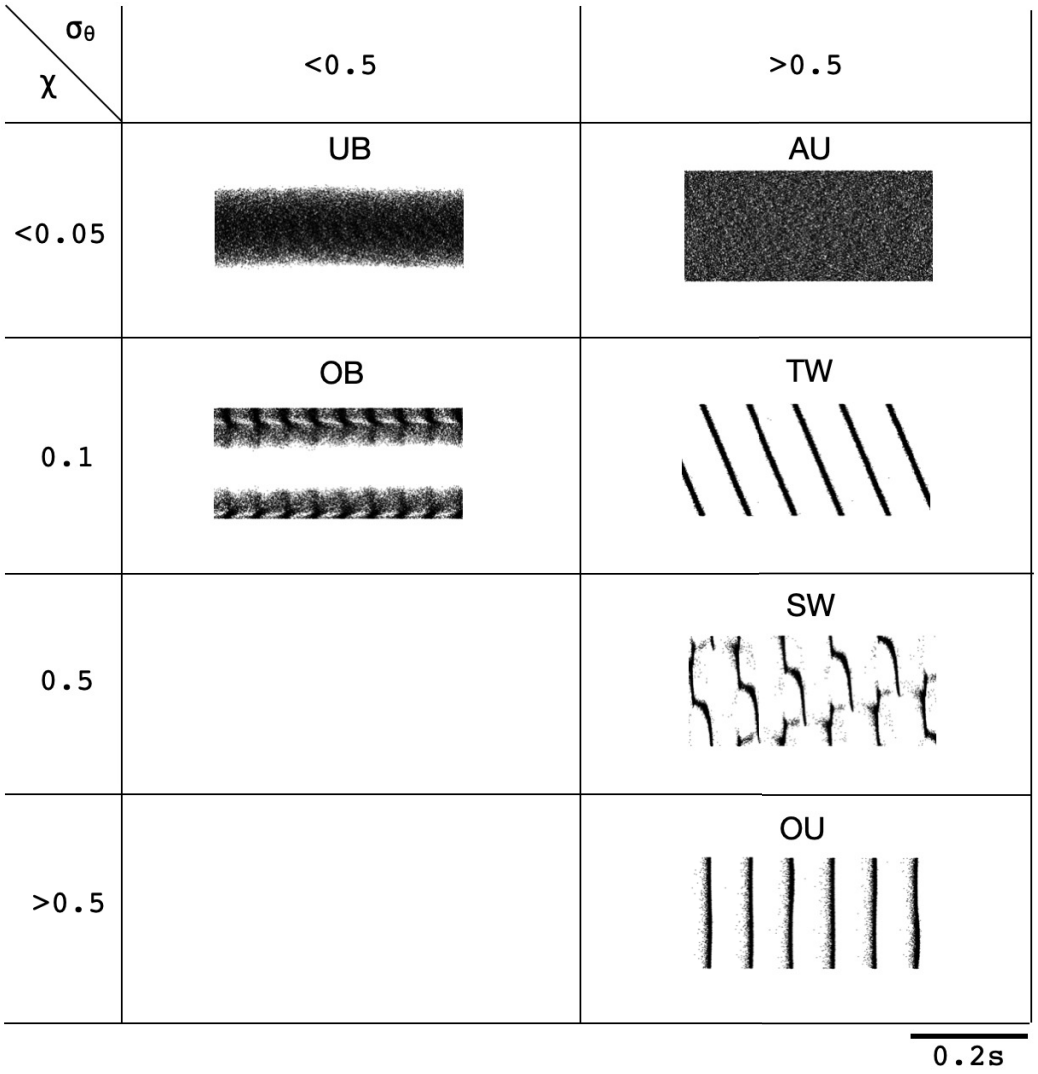

Table 4.1 – Distinct spatiotemporal patterns classified by the synchrony (χ) and temporal standard deviation of the bump's argument (σ_θ). UB: uniform bump; OB: oscillatory bump; AU: asynchronous uniform; TW: traveling wave; SW: standing wave; OU: oscillatory uniform.

4.2.2 Network finite size effects and spontaneous transitions.

As we previously seen in section 3.2.4 a possible way to obtain bistability in the spiking network model is to use a difference in the acceleration of the TFs. That is, at a certain threshold the inhibitory TF display a larger gain thus resulting in an effective saturation (Figure 4.3)

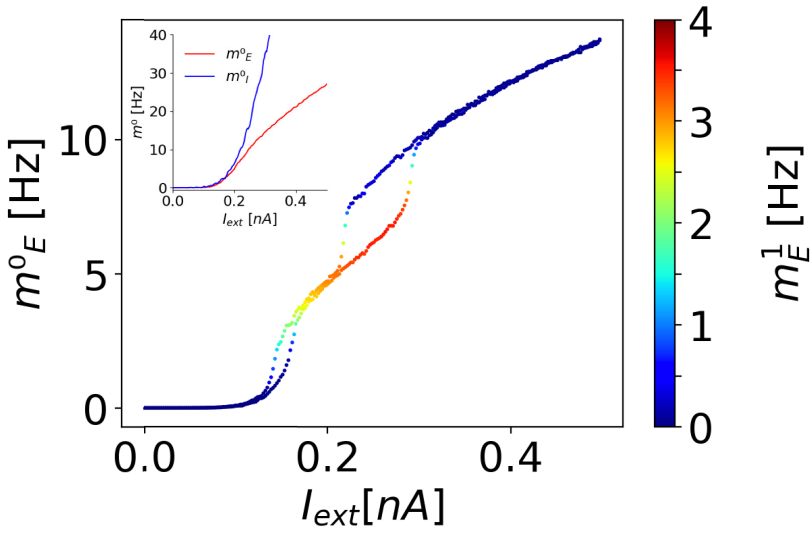

Figure 4.3. Bifurcation diagram of the LIF network in the acceleration difference model. **A.** Bifurcation diagram of m_0^E-C in the two population network in the acceleration difference case. Color code: first Fourier component of the activity, m_1^E. Inset: the TFs of the excitatory (red) and the inhibitory (blue) neurons. Parameters as described in Appendix B.6.1, Tables B.3 and B.4 but with $J_0^{EX} = 25$, $J_0^{IX} = 18$, $J_1^{EX} = 50$ and $J_1^{IX} = 12.5$.

The models used in this section is the similar as used in section 3.2.6 in which we consider excitatory synaptic currents as a combinations of currents shaped by NMDAR and AMPAR dynamics. Moreover, we choose physiological values of the synaptic time constants as described in the previous section 4.2.1. Furthermore, we choose a larger contribution of the NMDA current in E-E synapses (Thomson, 1997) based on the NMDA/AMPA ratio of the EPSCs area (Gonzalez-Burgos et al. (2008), NMDA/AMPA=2.7) and larger AMPA contribution in the E-I synapses Wang and Gao (2009), NMDA/AMPA=0.5)(For further details see Appendix B.3.3).

In this model, we found that a decrease in the mean excitatory connectivity (reduction of the J_0^E parameter) and/or a reduction in the number of neurons in the network lead to spontaneous transitions from the upper homogeneous state to bump state (H-B transition). On the other hand, a decrease in the inhibitory spatial connectivity (J_1^I) tends to increase the region of bistability making it more robust against spontaneous transitions (as predicted from the $J_1^I - C$ phase diagram in section 3.2.4 Figure 3.6 D). Here we are going to study the mechanisms underlying the spontaneous H-B transition upon variations in the number of neurons.

We hypothesize that the transition takes place due to the fluctuations of the mean synaptic inputs and the proximity of the unstable state and the stable state. At some point the fluctuations of the mean inputs are large enough to cross the unstable state (or separatrix in the $m_0^X - m_1^X$ activity plane, Figure 4.4). In order to estimate the time needed to this transition to take place we consider the abstract idea that the probability of one neuron to cross the separatrix is p, thus the probability of the activity of all the N neurons in the network to cross the separatrix (i.e. global increase of m_1^X) at the same time is the joint probability i.e. p^N, considering the events are independent from each other. Thus, the expected value of realizations to see at least once that $O(N)$ neurons crossed the separatrix will be p^{-N}. Hence, we hypothesize that the time expected to see a transition will scale exponentially with the number of neurons in the network N.

To test this hypothesis we set the network in the homogeneous state in the bistable region (Figure 4.4 A and B $I_{ext} = 0.27\,nA$). Then, we run several simulations of the network with different number of neurons and store the time needed to see a spontaneous transition. We named this variable time to transition (t2t). Then we plot thet2t obtained for different network size. We found that the t2t can be well fitted by an exponential function of the form ap^{-N} (Figure 4.5, $R^2 = 0.99$ with $p = 8 \cdot 10^{-7}$). This confirms our hypothesized mechanism for the spontaneous transition based on the crossing of the separatrix due to the proximity between the homogeneous fixed point and the unstable fixed point (on the separatrix).

Another type of spontaneous transition can happen if you set the network at the upper homogeneous state close to the transition towards the bump (Figure 4.3 and Figure 4.4 A and B, $I_{ext} = 0.25\,nA$). Due to the fluctuations of the mean

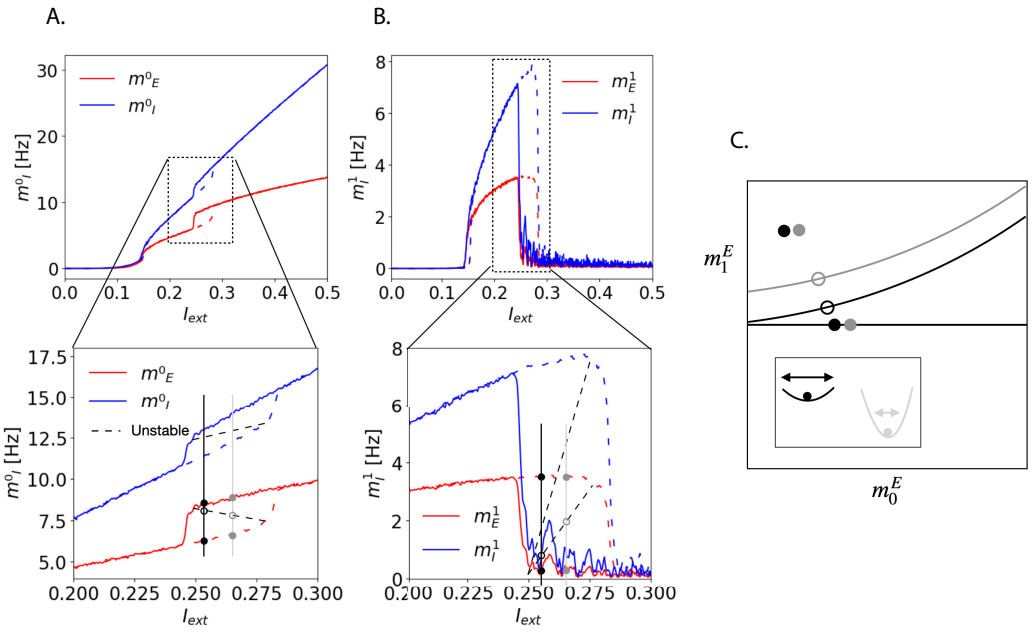

Figure 4.4. Spontaneous transition mechanism. A. Top: Bifurcation diagram of m_0^X-C in the two population network in the acceleration difference case. Excitatory population in red and inhibitory population in blue. Bottom: detail of the bistable region in **A**. Expected unstable state inferred from the rate bifurcation diagram depicted as a black dashed line. For two different values of I_{ext}(black and gray) solid dot indicates stable fixed point and empty dot indicates unstable fixed point. **B.** Same for the m_1^X-C bifurcation diagram. **C.** Schematic representation of the separatrix and fixed points in the $m_0^E - m_1^E$ space. Notice that a decrease in the I_{ext} (form grey to black situation) produces a shift in the separatrix and fixed points making the homogeneous fixed point to be closer to the separatrix. Inset: abstract representation of the potential energy well of the homogeneous fixed point in the two different situations. The expected amplitude of the fluctuations of the activity at the fixed point is depicted by a double ended arrow.

input, the network could receive global inputs below $I_{ext} = 0.25\,nA$. In this region the homogeneous state is spatially unstable, hence a transition towards the bump state will occur. In this case, the probability for this to take place will be directly proportional to the size of the fluctuations. From the central limit theorem the amplitude of these fluctuations will scale as $1/\sqrt{N}$. Thus we expect that the time to transition will scale as $\sqrt{N}$. To test this, we set the network in the homogeneous state very close to the edge of the upper homogeneous state

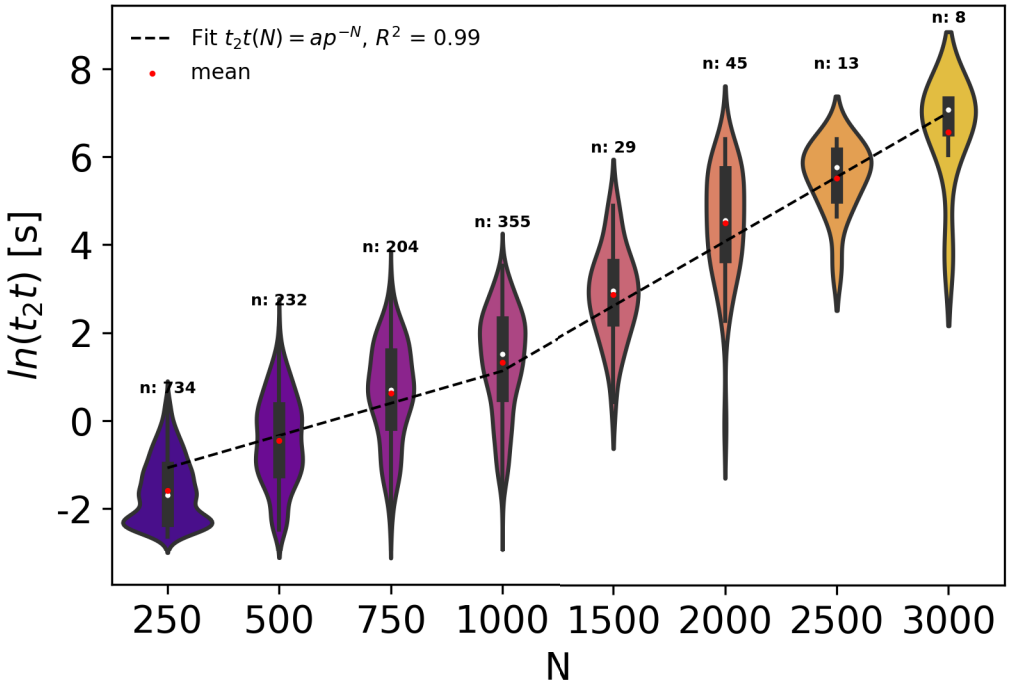

Figure 4.5. Time to spontaneous transition due to the crossing of the separatrix at different network sizes. Violin plot of the distribution of the natural logarithm of t2t for different network sizes. Median: white dot; Mean: red dot; Interquartile range: black bar in the center of the violin plot. Dashed line: least squares regression to a linear model $\ln(t2t) = \ln(a) - N\ln(p)$ with $R^2 = 0.99$ and $p = 8 \cdot 10^{-7}$. Minuscule n above each violin plot represents the number of simulations in the distribution for each network size.

($I_{ext} = 0.25\,nA$) and we simulate the network with different sizes and record the time needed to spontaneously transition to the bump state (t2t). In this case, the t2t can be well fitted with a power law function of the form N^a with $a = 1/2$ (Figure 4.6, $R^2 = 0.99$ with $p = 5 \cdot 10^{-7}$). Notice that in this case we can explore larger network sizes since the rate of increase of t2t with N is smaller. In this case, the t2t for the same network size are smaller than for the previous case. This is because in the previous case the fluctuations in the mean synaptic recurrent inputs were not large enough to make the population activity crossing the separatrix since for $I_{ext} = 0.27\,nA$ the separatrix is relatively far from the homogeneous fixed point (Figure 4.4 C, in gray). For $I_{ext} = 0.25\,nA$ the proximity between the

separatrix and the homogeneous fixed point is small enough to be crossed due to the fluctuations in the mean synaptic recurrent inputs (Figure 4.4 C, in black).

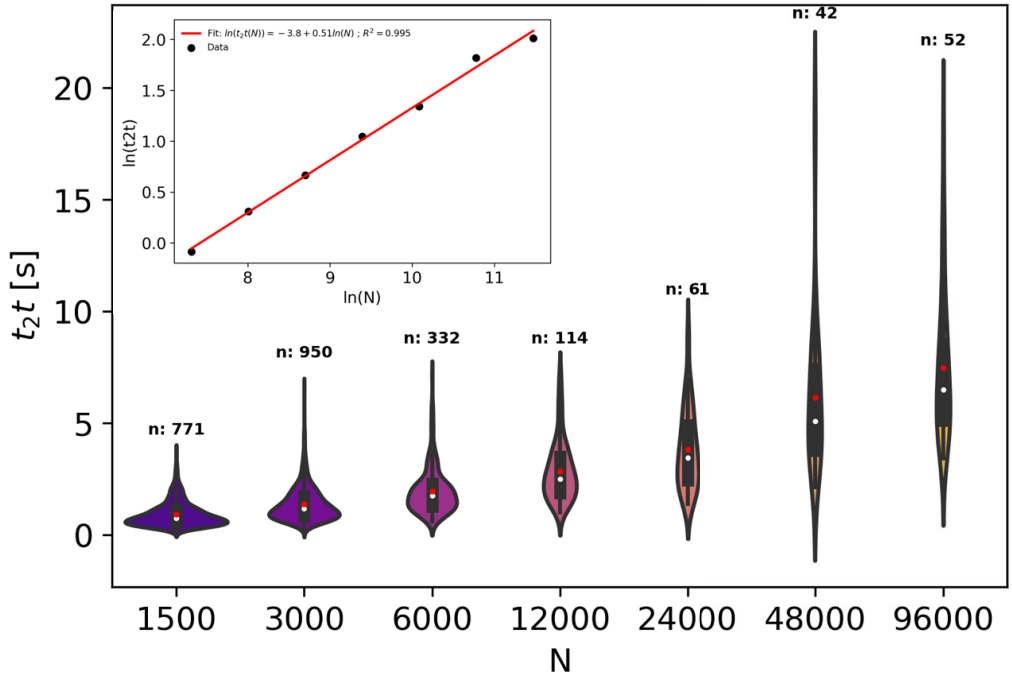

Figure 4.6. Time to spontaneous transition due to fluctuations in the mean synaptic inputs for different network sizes. Violin plot of the distribution of t2t for different network sizes. Median: white dot; Mean: red dot; Interquartile range: black bar in the center of the violin plot. Minuscule n above each violin plot represents the number of simulations in the distribution for each network size. Inset: least squares regression to a linear model $\ln(t2t) = \ln(b) + a\ln(N)$ with $R^2 = 0.99$ and $p = 5 \cdot 10^{-7}$.

4.3 DISCUSSION and CONCLUSION

In this work we first explore how the synaptic decay time constants leads to a rich variety of spatiotemporal patterns of activity. We found that asynchronous or uniform activity requires slower excitation than inhibition and a decrease of the synaptic excitatory time constant will result in synchronization of the activity and a disruption of the bistability. In the second part of this work we study the

mechanisms underlying the spontaneous transitions between the homogeneous state and the bump state. The results suggest that, under a certain circumstances, a reduction in the number of neurons will increase exponentially the probability for this spontaneous transitions to take place. Spatiotemporal patterns arising in a bump attractor model have been well characterized in a previous work from Roxin et al. (2005). In this work they studied the role of delays in a firing rate model and they presented a detailed phase diagram of the distinct patterns arising upon variation of the connectivity parameters. They also provide a spiking network model which reproduces most of the patterns described in the rate model. The only pattern that could not be reproduced was the traveling wave. However in our model with two populations, for inhibitory time constant equal to 2 ms and excitatory time constant 4 ms we found robustly the travelling wave pattern. Our results show that variations in the synaptic time constants are qualitatively similar to have delays in the activity. We found that at physiological values of the inhibitory time constant attributed to $GABA_A$ receptor (5-10ms), a reduction in the NMDA contribution by decreasing the synaptic time constant (decrease of ratio τ_d^E/τ_d^I) leads to oscillatory patterns (OU and OB). This transition is depicted in the phase diagrams in Figure 4.2 A and B, by a light blue region that separates uniform activity in the top right from other patterns in the bottom left corner. Similar results where described in Compte et al. (2000) in which a decrease of the NMDAR contribution elicited oscillatory activity in the memory state (bump) with a frequency peak at 40 Hz (γ range). This synchronization in the spiking activity is due to the E-I loop in the case of fast synaptic transmission. The mechanism for this oscillation is well described in Hansel and Mato (2003); Ledoux and Brunel (2011): a fast increase in excitatory activity will produce an increase of the inhibition that will result in a decrease on the excitation and a consequent decrease of the inhibition leading to an increase of the excitation thus repeating the loop. Thus an increase of excitatory synaptic time constants (increase of NMDA contribution) will attenuate the oscillations by reducing the fast increase of excitation as shown by Wang (1999). In another attractor model of spiking neurons presented by Laing and Chow (2001) also describe how a progressive reduction of the synaptic time constant could lead to a destabilization of the bump state towards a travelling wave. Overall these results suggest that NMDA hypofunction could potentially result in oscillatory activity and other

complex spatiotemporal patterns that could explain the cognitive disfunction in Schizophrenic patients as well as the development of positive symptoms such as hallucinations. Indeed the blockade of NMDAR is one of the possible mechanisms of hallucinations since dissociative drugs acting as a NMDAR blockers are known to produce hallucinations (Rolland et al., 2014).

On the second part we describe a mechanism for spontaneous transition that could explain spontaneous generation of internal representations in absence of external stimulus. The mechanisms rely on the proximity of the unstable state or separatrix and/or with the fluctuations on the mean recurrent input of each neuron. Under a given set of connectivity parameters ensuring bistability, we show that a reduction on the number of neurons of the network will result in an exponential increase of the probability of a spontaneous transition to occur. This situation occurs when the region of bistability is relatively small, making the homogeneous state to be closer to the unstable state. This can be accomplished by an increase of the inhibitory spatial connectivity parameter J_1^I or a reduction of J_1^E. This could happen because of a disruption of the connectivity in the PFC microcircuits.

The results presented in this section suggest that a reduction on the number of neurons or just a reduction on the inputs received due to the loss of spines, as described in layer III of the PFC in Schizophrenic brains (Hoftman et al. 2017), could lead to spontaneous generation of internal representations of information that is not triggered by external stimulus. It is important to underline that in our model the connectivity is normalized by the network size thus, a reduction in the number of inputs by decreasing the number of neurons leads to an increase in the connectivity between the neurons. This makes that the total synaptic input to a postsynaptic neuron remains unaltered when changing the network size. Although this assumption is speculative, a facilitation mechanism based in synaptic plasticity could explain an increase of the connectivity when the total number of synaptic connections is reduced. Overall, the proposed mechanism could explain the origin of hallucinations. This hypothesis is reinforced by the fact that there is a significant decrease in the functional connectivity in PFC suggesting a reduction of the number of neurons activated in the PFC recurrent networks involved in WM (Perlstein et al., 2001; Chari et al., 2019). On the other

hand, around 20% of patients with Alzheimer's disease show visual hallucinations (Sinclair et al., 2019). It is well known that in Alzheimer, there is a considerable connectivity loss (Rose et al., 2000). This could result in a reduction of the number of synaptic inputs neurons in PFC, leading to spontaneous transitions.

In the second mechanism described, a decrease of the external input arriving to the network leads to a situation in which the homogeneous state is very close to be spatially unstable. Moreover, in this case, the unstable state is very close to the homogeneous state. Thus, fluctuations in the mean synaptic inputs of a neuron can elicit an spontaneous transition towards the bump state. From these results we can hypothesize that a reduction of the external input to the PFC circuits probably coming from subcortical areas through the thalamus could set the network to operate in a regime very similar to the one described by our model leading to spontaneous internal representations. Indeed, the cognitive dysmetria theory postulated by Andreasen et al. (1996) point out that a disruption of PFC-thalamus-cerebellum loop could be the origin of most of Schizophrenic symptoms including hallucinations. The disruption of this loop could result in a reduction of excitatory inputs from thalamus to dlPFC as described in section 2.4.2 (Figure 2.13 A). A fMRI study found that Schizophrenic patients showed a decrease of functional connectivity between cerebellum and PFC and cerebellar stimulation can restore normal levels of cerebellar-PFC functional connectivity leading to a reduction of Schizophrenic symptoms (Brady Jr et al., 2019). In fact, in Glantz and Lewis (2000) they show evidence that patients with Schizophrenia show a reduction in excitatory inputs in pyramidal neurons of layer III in the dlPFC suggesting that this could explain cognitive disorders related to the disease.

Altogether these results present different mechanisms for the generation of spontaneous internal representations in PFC circuits that could be understood as network mechanisms of hallucinations. Small changes in the connectivity and/or decrease of external inputs to the PFC can result in spontaneous transitions that will be more frequent upon a reduction of the number of neurons in the circuit.

5 Cognitive impact of cerebellar non-invasive stimulation in a patient with schizophrenia [1]

Cerebellum plays a role in the regulation of cognitive processes. Cerebellar alterations could explain cognitive impairments in schizophrenia. We describe the case of a 50 years old patient with schizophrenia whom underwent cerebellar transcranial direct current stimulation (tDCS). In order to study the effect of cerebellar stimulation on cognitive functions, the patient underwent a neuropsychological assessment and an EBC protocol. Although the effect of brain stimulation cannot be only assessed in a single-case study, our results suggest that cerebellar stimulation may have an effect on a broad range of cognitive functions typically impaired in patients with schizophrenia, including verbal episodic, short term and working memory. In addition to neuropsychological tests, we evaluated the cerebellar function by performing EBC before and after tDCS. Our data suggest that tDCS can improve EBC. Further clinical trials are required for better understanding of how cerebellar stimulation can modulate cognitive processes in patients with schizophrenia and healthy controls.

[1] Charles Laidi, Carole Levenes, Alex Suarez-Perez, Caroline Février, Florence Durand, Noomane Bouaziz and Dominique Januel. Published in Frontiers in psychiatry, 2020.

5.1 INTRODUCTION

The cerebellum is involved in a broad range of cognitive functions, including working memory, emotion processing and social cognition (Li et al., 2013). In humans, the cerebellum represents 10 % of the brain volume but contains more than 50 % of its neurons (Li et al., 2013). The posterior lobe of the cerebellum is involved in cognition and connected to associative regions such as the prefrontal cortex, whereas the anterior cerebellum is known to modulate sensory-motor cortical activity (Stoodley et al., 2012). As alterations of the sensory-motor parts of the cerebellum lead to motor dysmetria, abnormalities in the posterior cerebellum may have implications for cognitive dysmetria. Andreasen et al. (1998) have proposed that abnormalities in the posterior cerebellum may explain some symptoms of schizophrenia. Schizophrenia is a severe mental disorder characterized by the association of positive, negative and cognitive symptoms. Cognitive symptoms, that often precedes the illness, have an impact on the quality of life and on the functioning of the patients (Green and Harvey, 2014). Deficits in working memory, attention, processing speed, visual and verbal procedural learning have been documented in schizophrenia (Gupta et al., 2018). However, pharmacological interventions only have a very limited impact on cognitive deficits in schizophrenia. Although non-pharmacological interventions, such as cognitive remediation can improve cognitive deficits in patients, there is a clear need for new interventions to target cognitive symptoms in schizophrenia. Neuropsychological tests are commonly used to assess cognitive functions in patients with schizophrenia. These tests require full cooperation of participants, which can be difficult in a population of patients with schizophrenia suffering from motivational deficits (Waltz and Gold, 2015).

Eyeblink conditioning (EBC) does not rely on motivation of the subject. It is based on a simple reflex pathway and mesures associative learning (Takehara-Nishiuchi, 2016). EBC is a form of classical conditioning that consists of pairing a stimulus (conditioned stimulus (CS), auditory in our study) with an unconditioned stimulus (US) (airpuff in our study) that induces an eyeblink reflex. In delay-type EBC, a tone CS precedes and co-terminates with a corneal airpuff US that elicits an unconditioned response (UR). Over repeated pairings, the CS induces a conditioned response (CR) that precedes and reduces the US.

McCormick et al. first showed that electrolytic lesions of the ipsilateral cerebellum completely prevented the acquisition and retrieval of the delay EBC (McCormick et al., 1982; Lincoln et al., 1982). The abundant literature based on lesion, reversible inactivation, genetic manipulation, electrical stimulation, optogenetics, electrophysiology, and brain-imaging studies show that the cerebellum is necessary and sufficient for acquisition, expression, and extinction of EBC provided that the interval between CS and US stay in the range of one second (see review in Freeman and Steinmetz (2011) and Li et al. (2019)). In accordance with animal research, EBC is a relevant method to investigate cerebellar dysfunction in schizophrenia disorders (Parker et al., 2017). Non-invasive brain stimulation techniques are commonly used in healthy adults and patients with neuropsychiatric disorders to investigate brain mechanisms or to enhance cognitive, behavioral, social and emotion processes (Finisguerra et al., 2019). Transcranial direct current stimulation (tDCS) is a form of neuromodulation delivering a low direct constant current over two electrodes placed on the scalp. Applied to the cerebellum, tDCS can deliver an electric field reaching the cerebellum at a strength within the range of values for modulating activity in the cerebellar neurons (Ferrucci et al., 2015). In healthy subjects, two studies (Zuchowski et al., 2014; van der Vliet et al., 2018) reported an effect of cerebellar tDCS on EBC. Interestingly, van der Vliet et al. (2018) reported an interaction between the effect of cerebellar tDCS on EBC and the BDNF Val66Met polymorphism, previously involved in cognitive deficits in schizophrenia (Notaras et al., 2015). We describe the case of a 50 years old patient with schizophrenia whom underwent posterior cerebellar tDCS. We report neuropsychological testing and monitoring of cerebellar function with EBC before and after 1 week of stimulation in the posterior cerebellum.

5.2 RESULTS

The patient was a 50 years old man suffering from schizophrenia. During the stimulation period, the patient was stabilized under a treatment of intramuscular haloperidol (150 mg every four weeks) and Zopiclone (7.5 mg/day). There was no change in the patient medication during the assessment and stimulation protocol.

The patient was married with two children and was discontinuously working in the construction sector. He was mostly complaining from auditory hallucinations: the patient reported that he was hearing at least once a day a male voice that was giving him orders. Written consent was obtained from the participant prior to the study. The patient underwent 5 days of tDCS stimulation. The post-tDCS EBC session was performed 5 days after tDCS; 7 days separated the two EBC sessions. The cerebellum was stimulated using a NeuroConn DC Stimulator (NeuroConn GmbH) with two 5x7 cm conductive-rubber electrodes placed over the cerebellum, 1 cm below the inion (anode) and on the right arm (cathode). Stimulation was administered during two sessions of 25 minutes (separated by one hour), including 5 seconds of ramp-up and 5 seconds of ramp-down, with an intensity of 2 mA (for a total of 10 stimulation sessions). The patient was stimulated during 5 consecutive days for a total of 10 sessions. We chose this stimulation protocol based on a previous modelling study Parazzini et al. (2014) and on the work of Ferrucci et al. (2015). Clinical assessment included the Positive And Negative Symptom Scale (PANSS) (Leucht et al., 2005) and the Auditory Hallucination Rating Scale (AHRS), before and after stimulation. Neuropsychological assessment explored key cognitive functions typically impaired in patients with schizophrenia: episodic memory, executive and attentional functions. We selected neuropsychological tests with no test/retest effect in order to compare neuropsychological outcomes before and after stimulation (Lee et al., 2018; Stoykova et al., 2013). The patient underwent a long term episodic memory test (French version of Free and cued recall - 16 items, Grober-Buschke, measuring anterograde episodic verbal memory using two different verbal material) (Stoykova et al., 2013), two subtests of the Wechsler Adult Intelligence Scale (WAIS-IV) (digit span and spatial memory), the stroop test (Golden version) (Golden, 1975) and the D2 Test of attention (Brickenkamp) (Lee et al., 2018). Neuropsychological assessment was repeated 2 days before and 2 days after the stimulation protocol (Figure 5.1) by a trained neuropsychologist that was not involved in the conception of the study nor in the brain stimulation.

The conditioning of the eyeblink reflex was performed with a portable human eyeblink conditioning system (San Diego Instruments). The system included an infrared (IR) reflective sensor glued together with small 1.5 mm air-delivering tubing positioned just beneath the superior eyelid of the subject. The EBC device,

Figure 5.1. tDCS Stimulation protocol and clinical assessment time-line.

comprising a portable airpuff and headset sound-delivering unit, controlled the timing and intensity of both the airpuff and the sound (pure tone). It also converted the analog IR-reflection signal to numeric and sent it to a personal computer. The IR-reflection signal was collected online with the San Diego Instruments Labview software and then analyzed offline with a custom-made routine under Python (Python Software Foundation). During the overall experiment a continuous background white noise was delivered to the subject through the headset in order to provide constant ambient noise. The conditioning stimulus (CS) consisted of a 400 ms - 1.2 kHz pure tone. We set the intensity of the CS such as it did not trigger any startle reflex nor any detectable reaction from the subject, thereby reducing the occurrence of alpha responses. The unconditioning stimulus (US) consisted of a 50 ms airpuff whose intensity (air pressure at the tip of the tube) was set to trigger painless eyeblink in 100% of the trials. Initially, the participant was exposed to five CS alone and to five US alone stimuli to establish appropriate responses to the tone and the airpuff as well as to measure the UR prior to conditioning. We also ensured that the US did not induce any startle reaction from the participant. A conditioning trial lasted one second and consisted of (successively): an initial 200 ms baseline period, a 400 ms CS that co-terminated with a 50 ms US, and a final 400 ms period during which the eyeblink was recorded. An EBC block consisted of 9 successive paired presentations of CS-US and a last trial with the CS alone. An EBC session consisted of 5 successive blocks separated by an inter-trial interval randomly ranging from 2 to 12 s. The participant was passively watching a silent movie during the task.

EBC sessions: The patient underwent two EBC sessions on the same days as the

clinical and neuropsychological assessments (Figure 5.1). The first one (pre-tDCS) was made two days before the first tDCS session, and the second one (post-tDCS) was made 2 days after the last tDCS session. Thus, 7 days separated the pre- and the post-tDCS EBC sessions.

The EBC signal was low-pass filtered using a 4th order Butterworth filter with a 10 Hz cutoff frequency and was offset-subtracted by deducing to the trace the averaged baseline. To estimate the percentage of CRs, we discarded trials for which a spontaneous blink occurred during the baseline. CRs were detected in a time window between 330 and 400 ms after the CS onset with a threshold of 5 times the standard deviation above the baseline. We visualized each trace separately afterwards to verify that the detection of CRs was correct. Clinical and neuropsychological characteristics before and after cerebellar stimulation are reported in Table 5.1. The patient did not report any side effects, except from a slight itching in the beginning of the first session of the second day of stimulation. A careful inspection of the scalp did not evidence any cutaneous lesion. The patient did not report any headache after brain stimulation. Clinical symptoms remained stable during the stimulation protocol. Notably there was no changes in the PANSS score and the Auditory Hallucination Verbal Scale. There was a global improvement in a large part of neuropsychological measurements (episodic memory, executive and attentional functions) before and after stimulation (Table 5.1).

We found an improvement in the long-term episodic memory, assessed with the free/cued recall 16 items (Table 5.1) test. There was an increase of performance in the delayed free recall: after the stimulation, the score of the participant was in the normal range (19 to 20 percentile) *vs* < 1 percentile before stimulation. Likewise, there was a strong improvement in the first attempt of the free recall after stimulation: the score of the participant was in the normal range (11 to 12 percentile) after stimulation *vs* 1 to 2 percentile before stimulation. Two different lists of words were proposed to the participant before and after stimulation in order to avoid a test-retest effect (Stoykova et al., 2013).

	Before stimulation	After stimulation
CLINICAL ASSESSMENT		
PANSS score:		
- Total	57	58
- Positive	22	24
- Negative	15	14
- General psychopathology	20	20
AHRS	22	23
NEUROPSYCHOLOGICAL ASSESSMENT		
Free and cued recall 16 items:		
- Recall 1		
- Free recall (pc)	5 (1–2)	7 (11–12)
- Total recall (pc)	8 (1)	11 (1–5)
- Recall 2		
- Free recall (pc)	5 (<0.1)	7 (3–4)
- Total recall (pc)	10 (<1)	13 (5)
- Recall 3		
- Free recall (pc)	7 (<1)	8 (1–2)
- Total recall (pc)	9 (<1)	12 (1)
- Delayed recall		
- Free recall (pc)	7 (<1)	11 (19–20)
- Total recall (pc)	9 (<1)	11 (1)
WAIS-IV:		
- Digit span		
- Direct order	5	6
- Indirect order	4	5
- Increasing order	6	6
- Total	15	17
- Spatial memory		
- Direct order	6	9
- Indirect order	6	6
- Total	12	15
Stroop-test golden version*:		
- Reading (pc)	74 (4–5)	82 (8–12)
- Denomination (pc)	62 (12)	61 (8–12)
- Interference (pc)	30 (5–8)	33 (12)
D2 test of attention (Brickenkamp):		
- GZ (pc)	187 (<0.1)	238 (0.5)
- % of errors (pc)	3.21 (50–75)	6.72 (25–50)
- KL (pc)	73 (4.5)	86 (8.1)

*Pc, percentiles; PANSS, Positive And Negative Symptoms Scale; AHRS, Auditory hallucinations rating scale; CGI, Global clinical impression; WAIS IV, Wechsler Adult Intelligence Scale (WAIS-IV); GZ, quantitative performance index; KL, concentration performance index; *age corrected scores.*

Table 5.1 – Clinical and neuropsychological characteristics before and after cerebellar stimulation.

In two tests measuring short term and working memory (digit span and spatial memory), there was an improvement in the performance of the participant in both the direct and indirect order, suggesting an effect of stimulation in both the short term and working memory.

We found an increase in the Stroop test performance. In particular, the reading performance improved from 4-5 percentile (pathological range) before stimulation to 8-12 percentile (normal range) after stimulation. In addition there was an improvement in the interference condition after stimulation (12 percentile) as compared to before stimulation (5 to 8 percentile).

Last, we measured the selective attention with the D2 test of attention (Brickenkamp). Again, there was an increase in both the quantitative performance index (GZ) and the concentration performance index (KL). In the KL index, the subject scored in the normal range after cerebellar stimulation *vs* in the pathological range ($<$0.5 percentile) before stimulation.

Results from the EBC assessment before and after stimulation are reported in Figures 5.2 and 5.3. Before tDCS, the averaged block response amplitudes of the URs remained unchanged over the pre-tDCS session (Figure 5.2 A), very few CRs were detected and peak latencies of averaged block signals remained constant during the overall session (Figure 5.3). Thus, 45 CS-US pairings were insufficient to induce EBC, at the end of this session only 43 % of the trials displayed CRs (Figure 5.3 A). This result is in agreement with previous EBC evaluation in patients with schizophrenia (Golden, 1975). After tDCS, the patient was rapidly conditioned and reached a final value of 83 % of CRs (Figure 5.3 A), as expected for a normal EBC session. Accodingly after tDCS, the averaged amplitude of the URs decreased from block to block (Figure 5.2 D) while it was stable before tDCS, This progression of EBC can also be observed by monitoring the first peak latency of the responses from block to block. Before tDCS it remained stable but rapidly decreased over the blocks after tDCS (Figure 5.3 B). Those features indicate the shift of the eyeblink timing towards the CS, which corresponds to a progressive change from *reflexive* towards *predictive* behavior. Thus, before tDCS the patient could not be conditioned over the EBC session while after cerebellar tDCS he displayed progressive conditioning from block to block.

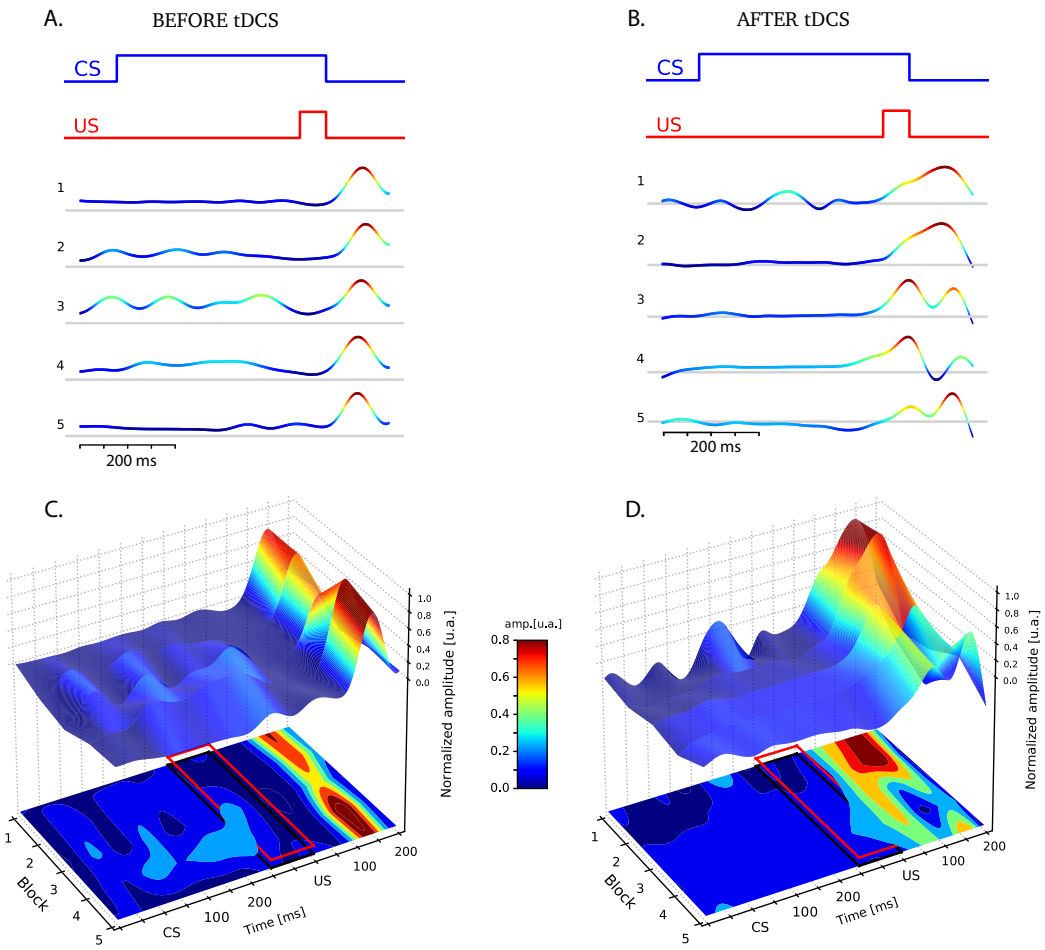

Figure 5.2. EBC sessions before (A and C) and after (B and D) tDCS in the schizophrenic patient.A and **B**: Each block represents the average signal over the 9 CS-US trials, for EBC experiments respectively before (**A**) and after (**B**) tDCS. The EBC protocol is depicted at the top of panels **A** and **B**. **C** and **D**: Plots of the IR-reflected signal for each block before (**C**) and after (**D**) tDCS. Red rectangles delimit the area where CRs were detected in Figure 5.3 A. Notice that the unconditioned response (UR) peak amplitude and latency decreases block after block after tDCS (**D**), while before the tDCS the UR peak amplitude remains constant among all the blocks, and the latency shows no trend (**C**). The color code corresponds to the normalized response amplitude of the IR-reflected signal.

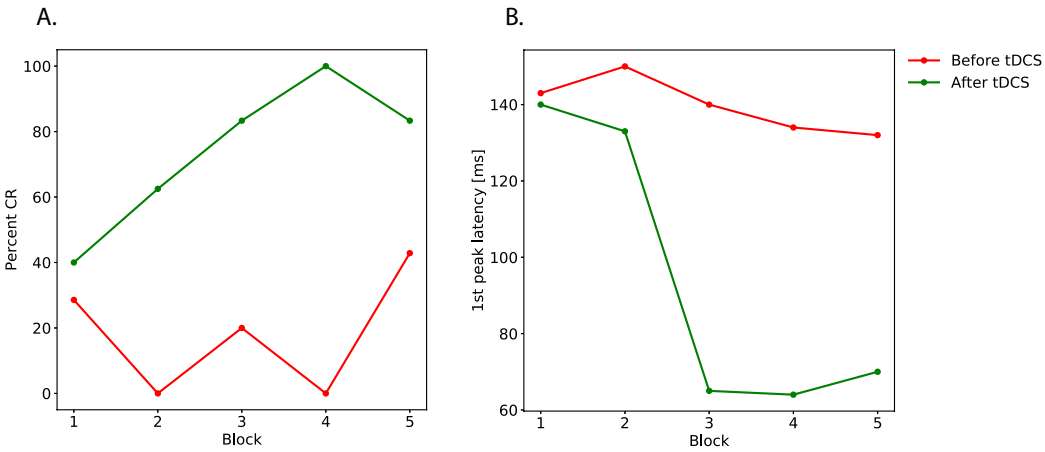

Figure 5.3. Percentage of CRs and latency of the first peak progression from block to block before (red) and after (green) tDCS. A. Percentage of CRs from each block measured in the red rectangle indicated on Figure 5.2 C and D. **B.** First peak latency calculated from the data averaged in blocks in Figure 5.2 A and C. Time zero for latency calculation corresponds to the onset of the US.

5.3 DISCUSSION and CONCLUSION

We describe the case of a 50 years old patient with schizophrenia whom underwent a non-invasive cerebellar stimulation protocol. Data from clinical and psychometric evaluations including long term verbal memory, executive and attention functions were collected before and after the stimulation, as well as data from a cerebellar-dependent eyeblink conditioning protocol.

Although we did not report changes in the positive or negative symptoms of schizophrenia before and after stimulation, there was a global improvement in psychometric measurements after stimulation. We also found an improvement in the performance in selected attention, long/short term memory, working memory and response inhibition; cognitive domains known to be altered in patients with schizophrenia. There was a clear improvement of EBC after stimulation. In the absence of data on healthy subjects in our conditions, we can not exclude any retest effect ("saving") in this improvement (Hoffland et al., 2012). This is however unlikely given that in our data: 1) no clear cut EBC could be observed at

the end of the pre-tDCS session, and 2) conditioning gradually appeared during the post-tDCS EBC session, starting from an absence of predictive response (and thus showing no saving, Figure 5.3). In addition, and contrary to control subjects, patients with schizophrenia have been shown not to improve their performance during consecutive EBC sessions (Bolbecker et al., 2009). Those lines of evidence therefore support the view that the improvement of EBC after tDCS was due to the stimulation itself and not to any retention of the first EBC session. Based on the abundant literature in both humans and animals (Thompson and Steinmetz, 2009; Mauk et al., 2014; Timmann et al., 2010; Freeman and Steinmetz, 2011), our data indicates that the cerebellar function of the patient was basally impaired as previously described in schizophrenia (Kent et al., 2015). Our EBC assessment is consistent with several studies that reported an effect of non-invasive cerebellar stimulation on associative learning measured with EBC (Zuchowski et al., 2014; Monaco et al., 2014). More importantly, it points out cerebellar tDCS as a powerful tool to significantly improve cerebellar function in schizophrenia. EBC has proven to be a relevant method to investigate cerebellar dysfunction in neuropsychiatric disorders. Disentangling motivational aspects of cognitive deficits can be challenging in patients with schizophrenia. This is however important since the deficits in motivation commonly present in patients with schizophrenia can bias classic neuropsychological tests. EBC does not require active participation of the subject. In our paradigm, the patient was watching a silent movie during the experiment; in newborns, EBC can even be performed during sleep (Fifer et al., 2010; Tarullo et al., 2016). Thus, the outcome of EBC are unlikely to be related to motivational deficits in patients with schizophrenia.

Gupta et al. (2018) found in a double-blind crossover study an effect of cerebellar tDCS on procedural learning in a population of non clinical psychosis (NCP) population. The authors reported greater rate of motor learning in NCP population after active stimulation. We used a different stimulation protocol with the cathode electrode (return electrode) placed on the right arm, whereas Gupta et al. (2018) placed the electrode on the midline of the scalp. Although there is no consensus on the placement of the return electrode (Ferrucci et al., 2015), we chose this location based on a modelisation study (Parazzini et al., 2014) to target the posterior region of the cerebellum.

Brady Jr et al. (2019) reported in a population of patients with schizophrenia, an improvement of negative symptoms after transcranial magnetic stimulation (TMS) related to dorso-lateral prefrontal cortex-to-cerebellum connectivity. However the authors did not investigate the effect of cerebellar stimulation on cognitive symptoms.

In healthy subjects, there is evidence that non-invasive cerebellar stimulation can modulate working memory, motor control, learning and emotional processing (Ferrucci et al., 2015). These results are in line with our results where non-invasive cerebellar stimulation had an effect on verbal memory, executive and attention function. The participant did not report any significant side effects after 2 sessions of stimulation during 5 days, which is in line with previous studies showing the feasibility and good tolerance profile of cerebellar tDCS (Ferrucci et al., 2015).

Our results supports several strengths. To the best of our knowledge, this case is the first to report the effect of tDCS on cognition (including associative learning measured with EBC) in schizophrenia. We carefully selected psychological measurements with no test/retest effects, which suggests that the cognitive improvement is related to the stimulation. In addition, there was no significant change in the positive and negative symptoms, suggesting again that the change in cognition are not related to a change in the symptoms of schizophrenia. Our work suggests that eyeblink conditioning can be used to assess the effect of cerebellar stimulation.

Several limitations should be considered before interpreting our results. Because we only investigated the effect of stimulation in a single patient, our study remains purely qualitative. The posterior part of the cerebellum is connected to multiple regions in the associative cortex and it is difficult to target a specific domain of cognition with cerebellar brain stimulation. We were not able to measure other cognitive domains, such as social cognition that could also be modulated by cerebellar stimulation (Stoodley et al., 2017). However, our goal was to propose an original cognitive evaluation by combining a classic neuropsychological assessment and EBC.

In conclusion, our results suggests that cerebellar tDCS stimulation can have an impact on cognitive impairments in patients with schizophrenia. We suggest that eyeblink conditioning, known as a relevant method to investigate cerebellar

dysfunction in neuropsychiatric disorders, could be used to assess the impact of stimulation on the cerebellum in patients with schizophrenia. Further clinical trials are required to address the potential therapeutic potential of tDCS in schizophrenia.

Part III

EPILOGUE

6 Discussion

Working memory or, maintaining a mental representation of past immediate information to drive coherent actions, is critical feature in the realisation of complex cognitive functions such as planning, decision making, learning, thinking etc. Experimental evidence point that neural correlates of working memory involves persistent selective activity in the PFC. Main accepted mechanisms of persistent activity rely on strong recurrent excitation in PFC whereas selectivity or spatial tuning arises from columnar PFC architecture together with lateral inhibition (Rao et al., 1999; Goldman-Rakic, 1995; González-Burgos et al., 2000). However, the mechanisms underlying network bistability remain elusive. On the other hand, mental illnesses such Schizophrenia present PFC disfunction and particularly WM deficits (Lee and Park, 2005). Although main theories of the Schizophrenia etiology underline the role of PFC dysconnectivity with subcortical areas as well as imbalance in the glutamate, dopamine and GABA systems, the mechanisms underlying WM deficits in Schizophrenia remain unknown (Andreasen et al., 1996; McCutcheon et al., 2020). In this thesis we argue the idea that understanding the neurobiology of the dlPFC and WM provides important knowledge in order to unravel the etiology of mental illnesses such Schizophrenia. With the aim to shed some light into these questions, in this thesis, we present novel mechanisms for network bistability based on the nonlinearities in the neuronal input-output transfer functions. With this model, we study how changes of biological parameters seen in Schizophrenia can shape the spatiotemporal dynamics. Finally we presented a Schizophrenia treatment based on cerebellar stimulation.

We asked ourselves in the prologue what network mechanisms underlie WM in

PFC. We saw that experimental evidence supports the theory that visuospatial location of an object is transiently maintained through persistent selective activity in specific regions in the PFC. We also reviewed several works that propose different mechanisms for persistence and selectivity within a single network of neurons based on continuous bump attractors (Camperi and Wang, 1998; Compte et al., 2000; Gutkin et al., 2001). In Camperi and Wang (1998) they used bistable neurons in order to have network bistability, however in the last part of the results they propose that a nonlinearity in the input-output neuron TF could be a suitable substrate underlying network bistability. In our work we propose a bump attractor network model of rate neurons to exhaustively study how the typical nonlinearities found in the experimental neuron TF can shape the network bistability. The firing rate model presented describes the temporal dynamics of the activities of the synapses of the neuron with $PD = \theta$. We found that the saturation of the TF can generate a bistability that allows to switch off the bump state by means of a global excitatory input. This is relevant since in electrophysiological experiments there is a global increase of activity at the end of the delay period. This suggest that an external excitatory input from other cortical or subcortical could drive the network from the memory state (bump state) to spontaneous activity (homogeneous state). The main problem is that neurons during WM do not operate in its saturation firing rates indicating that the saturation of the TF is not a biological realistic substrate for the bistability. To overcome this problem we showed that in a model comprising excitatory and inhibitory populations, a similar bistability can be obtained by an effective saturation of the network TF. This mechanism in which inhibition overcome excitation at a certain point subserves a mechanism to switch off the bump state through global excitation without requirement of the saturation part of the single neuron TF. We implemented the effective saturation mechanism in a network of spiking neurons and verify a qualitatively similar scenario in which switch off of the bump state through global excitation. The spike time dynamics allowed to study in detail the state transitions. We found that for a given range of switch off stimulus duration an oscillatory pattern determines the state transition.

Bistability arising from neuron TF nonlinearities cannot be applied to balanced networks since the main feature of this networks is that the mean firing rate have a linear dependence with the input independent of the nonlinearities in the

neuron TF (van Vreeswijk and Sompolinsky, 1996). In this case, the effect of the nonlinearities in single neuron TF is not represented in the network level. A possible bistability mechanisms in balanced systems rely on synaptic nonlinearities. In the literature some examples of synaptic nonlinearities in balanced networks rely on short term synaptic plasticity (Mongillo et al., 2012). In another work, Hansel and Mato (2013) present a balanced network of integrate-and-fire neurons with synaptic plasticity displaying persistent selective activity reproducing the irregular firing found in WM experiments. However, in these works the switch off by global excitation is not directly addressed but the results in Hansel and Mato (2013) suggest that synaptic plasticity could provide a compressive nonlinearity enabling the bump termination by a global excitatory input.

Future experimental studies will be necessary in order to demonstrate the impact of neural nonlinearities in the genesis of bistability in the PFC circuits. A possible experiment setup could rely on optogenetic modification of membrane properties in order to modify the neuronal TF and study its impact in the PFC dynamics during WM tasks.

In Chapter 4 we wanted to explore how changes in changes in the spiking network model such as variations in the synaptic current dynamics or variations in the network size could affect to the bistability and the dynamics of the system. We saw in the literature review (Section 2.4) that in Schizophrenia, the PFC is strongly affected and WM is impaired. One of the main hypothesis of the origin of Schizophrenia involves a hypofunction of the NMDAR. We thus proposed to study how variations in the contribution of NMDA currents can alter the dynamics in the PFC circuits. By exploring different excitatory synaptic time constants representing relative contribution between AMPA and NMDA postsynaptic currents we found a vast number of spatiotemporal patterns previously described in a network model with delays (Roxin et al., 2005). In general, for synaptic time constants smaller than the membrane time constants, the bistability is disrupted. We also found that for physiological values of inhibitory synaptic time constants, a decrease in the contribution of NMDA currents leads to a synchronisation of the activity showing oscillatory patterns as described previously in (Compte et al., 2000). Besides, we described in the

literature review (Section 2.4) that pyramidal neurons in layer III in PFC show a dendritic spine loss in Schizophrenia (Glantz and Lewis, 2000). We hypothesize that this will lead to decrease in the number of synaptic inputs and this is qualitatively comparable with a reduction in the number of neurons in the network. We find that a decrease in the network size results in a major probability of spontaneous transitions to occur. Specifically, we found that a reduction in the number of neurons increases exponentially the probability of having spontaneous generation of bumps of activity in absence of external stimulation. We propose this as a possible mechanisms of hallucinations typically encountered in mental illnesses such as Parkinson and Schizophrenia which are characterized by synaptic dysconnectivity or degeneration (Burghaus et al., 2012).

In Chapter 5 we present a case report study in which we argue in favor of the hypothesis that Schizophrenic symptoms could be due to a functional dysconnectivity between cerebellum and PFC. Since first postulation by Andreasen et al. (1996), this hypothesis have been gaining strength due to increasing evidences of anatomical and functional dysconnectivity (Zhou et al., 2015; Brady Jr et al., 2019). In a paper from Parker et al. (2014) from Andreasen lab, they present the hypothesis that cerebellar stimulation could be a plausible treatment for Schizophrenia based in the idea of restoration of synchrony between subcortical structures and PFC. A recent study from Brady Jr et al. (2019) have shown that stimulation of the cerebellum by rTMS produces an increase in the functional connectivity between PFC and cerebellum that correlates with an improvement of Schizophrenic symptoms. In our work, we show that cerebellar stimulation with tDCS enhances cerebellar function measured with an EBC protocol that provides a characterization of the cerebellar function. Although PFC function was not be measured, there was an improvement in WM task as well as in other cognitive symptoms suggesting a restoration of PFC function directly associated with an improvement of cerebellar function. Despite the study was perform in one Schizophrenic patient, in conjunction with the work presented by Brady Jr et al. (2019), the results suggest its as a potential treatment for Schizophrenia.

A natural continuation of the work presented will be to incorporate the cerebellum

and thalamus in to the PFC model of WM in order to study how the inputs from subcortical areas can shape the PFC dynamics. In our model, the input from subcortical areas can be understood as the external input. As we saw in the model, the bistability region is bounded by certain values of external input, thus strong variations of this input could lead to a cessation of the bistability. We also showed that small changes in the external input could set the network to operate close to the transition thus increasing the probability of having spontaneous transitions. On the other hand, in the spiking model, modifications of the background noise which represents the synaptic bombardment, will directly impact the the firing threshold of the neuronal TF (Persi et al., 2011). Thus, changes in the background input due to a reduction of the synaptic inputs could potentially alter the neuronal TF and thus alter the network bistability.

There is increasing evidence of the role of subcortical areas in complex cognitive functions attributed to PFC such as WM. In a series of articles from Svoboda Lab they study the role of subcortical structures in the preparatory activity present in the anterolateral motor cortex (ALM) in mice during delayed response tasks (Li et al., 2015; Guo et al., 2017; Gao et al., 2018). The ALM could be considered as a counterpart of the PFC in mice and it displays persistent selective activity during the delay period of WM taks. In their works they use optogenetics to activate or inactivate specific areas and study its implication in the WM circuit dynamics. In Guo et al. (2017) they show that the thalamus display similar persistent selective activity as the ALM and inactivation of the thalamus results in impairments of WM suggesting that the circuit loop between ALM and thalamus is critical for maintaining the persistent activity. In Gao et al. (2018) they show that cerebellum is as well critical for the preparation of future movements in a WM task suggesting that the cerebellum through the thalamus, could play a role of timing the onsets for persistent selectivity generation and extinction. Moreover, another series of works from Deverett et al. have showed that Purkinje cells in cerebellum encodes task-relevant information and thus cerebellar disruption produces WM impairments (Deverett et al., 2018, 2019). However, further experiments studying the role of subcortical areas in WM are crucial to understand the PFC dynamic and more importantly, its implication in mental illnesses such Schizophrenia in which a disruption of these circuits seems to be a central cause of the symptoms.

Bibliography

Albus, M., Hubmann, W., Mohr, F., Hecht, S., Hinterberger-Weber, P., Seitz, N.-N., and Küchenhoff, H. Neurocognitive functioning in patients with first-episode schizophrenia. *European archives of psychiatry and clinical neuroscience*, 256(7): 442--451, 2006.

Alexander, G. E., DeLong, M. R., and Strick, P. L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, 9(1):357--381, 1986.

Amaral, D. G. Memory: Anatomical organization of candidate brain regions. In Plum, F., editor, *Handbook of Physiology: Section 1. The Nervous System*, pages 211--294. American Physiological Society, Bethesda, 1987.

Amari, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern*, 27(2):77--87, 1977.

Amit, D. J. and Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex*, 7(3):237--252, 1997.

Amit, D. J. and Brunel, N. Learning internal representations in an attractor neural network with analogue neurons. *Network: Computation in Neural Systems*, 6(3): 359--388, 1995.

Amit, D. J. and Mongillo, G. Spike-driven synaptic dynamics generating working memory states. *Neural Computation*, 15(3):565--596, 2003.

Amit, D. J. and Treves, A. Associative memory neural network with low temporal spiking rates. *Proceedings of the National Academy of Sciences*, 86(20):7871--7875, 1989.

Amit, D. J. and Tsodyks, M. Quantitative study of attractor neural network retrieving at low spike rates. i. substrate-spikes, rates and neuronal gain. *Network: Computation in neural systems*, 2(3):259, 1991.

Andreasen, N. C. and Pierson, R. The role of the cerebellum in schizophrenia. *Biological psychiatry*, 64(2):81--88, 2008.

Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Ponto, L., Watkins, G. L., and Hichwa, R. D. Schizophrenia and cognitive dysmetria: a positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry. *Proceedings of the National Academy of Sciences*, 93(18):9985--9990, 1996.

Andreasen, N. C., Paradiso, S., and O'leary, D. S. "cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? *Schizophrenia bulletin*, 24(2):203--218, 1998.

Andreasen, N. C., Nopoulos, P., O'Leary, D. S., Miller, D. D., Wassink, T., and Flaum, M. Defining the phenotype of schizophrenia: cognitive dysmetria and its neural mechanisms. *Biological psychiatry*, 46(7):908--920, 1999.

Angulo, M. C., Rossier, J., and Audinat, E. Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex. *Journal of neurophysiology*, 82(3):1295--1302, 1999.

Antonova, E., Sharma, T., Morris, R., and Kumari, V. The relationship between brain structure and neurocognition in schizophrenia: a selective review. *Schizophrenia research*, 70(2-3):117--145, 2004.

Arnsten, A. F. Prefrontal cortical network connections: key site of vulnerability in stress and schizophrenia. *International Journal of Developmental Neuroscience*, 29(3):215--223, 2011.

Arsiero, M., Luscher, H. R., Lundstrom, B. N., and Giugliano, M. The impact of input fluctuations on the frequency-current relationships of layer 5 pyramidal neurons in the rat medial prefrontal cortex. *J. Neurosci.*, 27(12):3274--3284, 2007.

Aura, J. and Riekkinen Jr, P. Blockade of nmda receptors located at the dorsomedial prefrontal cortex impairs spatial working memory in rats. *Neuroreport*, 10(2): 243--248, 1999.

Baddeley, A. Working memory oxford. *England: Oxford Uni*, 1986.

Barak, O. and Tsodyks, M. Working models of working memory. *Current opinion in neurobiology*, 25:20--24, 2014.

Barak, O., Tsodyks, M., and Romo, R. Neuronal population coding of parametric working memory. *Journal of Neuroscience*, 30(28):9424--9430, 2010.

Bear, M., Connors, B., and Paradiso, M. A. *Neuroscience: Exploring the Brain*. Jones & Bartlett Learning, 2020.

Ben-Yishai, R., Lev Bar-Or, R., and Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. U S A*, 92:3844--3848, 1995.

Blum, R. A. Effects of subtotal lesions of frontal granular cortex on delayed reaction in monkeys. *AMA Archives of Neurology & Psychiatry*, 67(3):375--386, 1952.

Bolbecker, A. R., Hetrick, W. P., Johannesen, J. K., O'Donnell, B. F., Steinmetz, J. E., and Shekhar, A. S. Secretin effects on cerebellar-dependent motor learning in schizophrenia. *American Journal of Psychiatry*, 166(4):460--466, 2009.

Bourgeois, J.-P., Goldman-Rakic, P. S., and Rakic, P. Synaptogenesis in the prefrontal cortex of rhesus monkeys. *Cerebral cortex*, 4(1):78--96, 1994.

Brady Jr, R. O., Gonsalvez, I., Lee, I., Öngür, D., Seidman, L. J., Schmahmann, J. D., Eack, S. M., Keshavan, M. S., Pascual-Leone, A., and Halko, M. A. Cerebellar-prefrontal network connectivity and negative symptoms in schizophrenia. *American Journal of Psychiatry*, 176(7):512--520, 2019.

Brody, B. A., Kinney, H. C., Kloman, A. S., and Gilles, F. H. Sequence of central nervous system myelination in human infancy. i. an autopsy study of myelination. *Journal of Neuropathology & Experimental Neurology*, 46(3): 283--301, 1987.

Brunel, N. Persistent activity and the single-cell frequency--current curve in a cortical network model. *Network: Computation in Neural Systems*, 11(4): 261--280, 2000.

Brunel, N. Dynamics and plasticity of stimulus-selective persistent activity in cortical network models. *Cerebral Cortex*, 13(11):1151--1161, 2003.

Brunel, N. and Wang, X.-J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of computational neuroscience*, 11(1):63--85, 2001.

Buchsbaum, M. S. and Hazlett, E. A. Functional brain imaging and aging in schizophrenia. *Schizophrenia research*, 27(2-3):129--141, 1997.

Buee, J., Deniau, J., and Chevalier, G. Nigral modulation of cerebello-thalamo-cortical transmission in the ventral medial thalamic nucleus. *Experimental brain research*, 65(1):241--244, 1986.

Burghaus, L., Eggers, C., Timmermann, L., Fink, G. R., and Diederich, N. J. Hallucinations in neurodegenerative diseases. *CNS neuroscience & therapeutics*, 18(2):149--159, 2012.

Camperi, M. and Wang, X. J. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J Comput Neurosci*, 5(4):383--405, 1998.

Cannon, T. D. How schizophrenia develops: cognitive and brain mechanisms underlying onset of psychosis. *Trends in cognitive sciences*, 19(12):744--756, 2015.

Cardin, J. A., Palmer, L. A., and Contreras, D. Cellular mechanisms underlying stimulus-dependent gain modulation in primary visual cortex neurons in vivo. *Neuron*, 59(1):150--160, 2008.

Carta, I., Chen, C. H., Schott, A. L., Dorizan, S., and Khodakhah, K. Cerebellar modulation of the reward circuitry and social behavior. *Science*, 363(6424), 2019.

Carvalho, T. P. and Buonomano, D. V. Differential effects of excitatory and inhibitory plasticity on synaptically driven neuronal input-output functions. *Neuron*, 61(5):774--785, 2009.

Chari, S., Minzenberg, M. J., Solomon, M., Ragland, J. D., Nguyen, Q., Carter, C. S., and Yoon, J. H. Impaired prefrontal functional connectivity associated with working memory task performance and disorganization despite intact activations in schizophrenia. *Psychiatry Research: Neuroimaging*, 287:10--18, 2019.

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10:910--923, 2000.

Conel, J. L. R. The postnatal development of the human cerebral cortex. vol. 1. the cortex of the newborn. 1939.

Connors, B. W. and Gutnick, M. J. Intrinsic firing patterns of diverse neocortical neurons. *Trends in neurosciences*, 13(3):99--104, 1990.

Constantinidis, C., Franowicz, M. N., and Goldman-Rakic, P. S. Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J. Neurosci.*, 21(10):3646--3655, 2001.

Constantinidis, C. and Wang, X.-J. A neural circuit basis for spatial working memory. *Neuroscientist*, 10(6):553--565, 2004.

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., and Arnsten, A. F. Persistent spiking activity underlies working memory. *Journal of neuroscience*, 38(32):7020--7028, 2018.

Coyle, J. T. Nmda receptor and schizophrenia: a brief history. *Schizophrenia bulletin*, 38(5):920--926, 2012.

Curtis, C. E. and D'esposito, M. The effects of prefrontal lesions on working memory performance and theory. *Cognitive, Affective, & Behavioral Neuroscience*, 4(4):528--539, 2004.

Darshan, R. and Rivkind, A. Learning to represent continuous variables in heterogeneous neural networks. *bioRxiv*, 2021.

Del Giudice, P., Fusi, S., and Mattia, M. Modelling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. *J. Physiol. Paris*, 97(4-6):659--681, 2003.

Deverett, B., Koay, S. A., Oostland, M., and Wang, S. S. Cerebellar involvement in an evidence-accumulation decision-making task. *Elife*, 7:e36781, 2018.

Deverett, B., Kislin, M., Tank, D. W., and Wang, S. S.-H. Cerebellar disruption impairs working memory during evidence accumulation. *Nature communications*, 10(1):1--7, 2019.

Dipoppa, M. and Gutkin, B. S. Flexible frequency control of cortical oscillations enables computations required for working memory. *Proceedings of the National Academy of Sciences*, 110(31):12828--12833, 2013.

Driesen, N. R., McCarthy, G., Bhagwagar, Z., Bloch, M. H., Calhoun, V. D., D'souza, D. C., Gueorguieva, R., He, G., Leung, H.-C., Ramani, R., et al. The impact of nmda receptor blockade on human working memory-related prefrontal function and connectivity. *Neuropsychopharmacology*, 38(13):2613--2622, 2013.

Durstewitz, D., Kelc, M., and Güntürkün, O. A neurocomputational theory of the dopaminergic modulation of working memory functions. *Journal of Neuroscience*, 19(7):2807--2822, 1999.

Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., and Alonso, A. A. Graded persistent activity in entorhinal cortex neurons. *Nature*, 420(6912): 173--178, 2002.

Ermentrout, G. B. Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.*, 61:353--430, 1998.

Esnaola-Acebes, J. M., Roxin, A., and Wimmer, K. Bump attractor dynamics underlying stimulus integration in perceptual estimation tasks. *bioRxiv*, 2021.

Ferrier, D. *The functions of the brain.* Smith, Elder, 1886.

Ferrucci, R., Cortese, F., and Priori, A. Cerebellar tdcs: how to do it. *The Cerebellum*, 14(1):27--30, 2015.

Fifer, W. P., Byrd, D. L., Kaku, M., Eigsti, I.-M., Isler, J. R., Grose-Fifer, J., Tarullo, A. R., and Balsam, P. D. Newborn infants learn during sleep. *Proceedings of the National Academy of Sciences*, 107(22):10320--10323, 2010.

Finisguerra, A., Borgatti, R., and Urgesi, C. Non-invasive brain stimulation for the rehabilitation of children and adolescents with neurodevelopmental disorders: a systematic review. *Frontiers in psychology*, 10:135, 2019.

Frankle, W. G., Lerma, J., and Laruelle, M. The synaptic hypothesis of schizophrenia. *Neuron*, 39(2):205--216, 2003.

Freeman, J. H. and Steinmetz, A. B. Neural circuitry and plasticity mechanisms underlying delay eyeblink conditioning. *Learning & memory*, 18(10):666--677, 2011.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.*, 61(2): 331--349, 1989.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *J. Neurophysiol.*, 63(4): 814--31, 1990.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.*, 65(6):1464--83, 1991.

Funahashi, S. and Inoue, M. Neuronal interactions related to working memory processes in the primate prefrontal cortex revealed by cross-correlation analysis. *Cerebral Cortex*, 10(6):535--551, 2000.

Fuster, J. M. and Alexander, G. E. Neuron activity related to short-term memory. *Science*, 173:652--654, 1971.

Fuster, J. *The prefrontal cortex*. Academic Press, 2015.

Fuster, J. M. Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *Journal of neurophysiology*, 36(1):61--78, 1973.

Gao, W.-J. and Snyder, M. A. Nmda hypofunction as a convergence point for progression and symptoms of schizophrenia. *Frontiers in cellular neuroscience*, 7:31, 2013.

Gao, Z., Davis, C., Thomas, A. M., Economo, M. N., Abrego, A. M., Svoboda, K., De Zeeuw, C. I., and Li, N. A cortico-cerebellar loop for motor planning. *Nature*, 563(7729):113--116, 2018.

Glantz, L. A. and Lewis, D. A. Reduction of synaptophysin immunoreactivity in the prefrontal cortex of subjects with schizophrenia: regional and diagnostic specificity. *Archives of general psychiatry*, 54(10):943--952, 1997.

Glantz, L. A. and Lewis, D. A. Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *Archives of general psychiatry*, 57 (1):65--73, 2000.

Goldberg, J. A., Rokni, U., and Sompolinsky, H. Patterns of ongoing activity and the functional architecture of the primary visual cortex. *Neuron*, 42(3):489--500, 2004.

Golden, C. J. A group version of the stroop color and word test. *Journal of personality assessment*, 39(4):386--388, 1975.

Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron*, 14(3):477--485, 1995.

Goldman-Rakic, P. S. et al. Working memory dysfunction in schizophrenia. *The Frontal Loves and Neuropsychiatric Illness. Washington, DC*, pages 71--82, 1994.

Goldstein, J. M., Goodman, J. M., Seidman, L. J., Kennedy, D. N., Makris, N., Lee, H., Tourville, J., Caviness Jr, V. S., Faraone, S. V., and Tsuang, M. T. Cortical abnormalities in schizophrenia identified by structural magnetic resonance imaging. *Archives of General Psychiatry*, 56(6):537--547, 1999.

González-Burgos, G., Barrionuevo, G., and Lewis, D. A. Horizontal synaptic connections in monkey prefrontal cortex: an in vitro electrophysiological study. *Cerebral Cortex*, 10(1):82--92, 2000.

Gonzalez-Burgos, G., Kroener, S., Zaitsev, A. V., Povysheva, N. V., Krimer, L. S., Barrionuevo, G., and Lewis, D. A. Functional maturation of excitatory synapses in layer 3 pyramidal neurons during postnatal development of the primate prefrontal cortex. *Cerebral Cortex*, 18(3):626--637, 2008.

Green, J., Adachi, A., Shah, K. K., Hirokawa, J. D., Magani, P. S., and Maimon, G. A neural circuit architecture for angular integration in drosophila. *Nature*, 546 (7656):101--106, 2017.

Green, M. F. and Harvey, P. D. Cognition in schizophrenia: Past, present, and future. *Schizophrenia Research: Cognition*, 1(1):e1--e9, 2014.

Guo, Z. V., Inagaki, H. K., Daie, K., Druckmann, S., Gerfen, C. R., and Svoboda, K. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*, 545(7653):181--186, 2017.

Gupta, A., Wang, Y., and Markram, H. Organizing principles for a diversity of gabaergic interneurons and synapses in the neocortex. *Science*, 287(5451): 273--278, 2000.

Gupta, T., Dean, D. J., Kelley, N. J., Bernard, J. A., Ristanovic, I., and Mittal, V. A. Cerebellar transcranial direct current stimulation improves procedural learning in nonclinical psychosis: a double-blind crossover study. *Schizophrenia bulletin*, 44(6):1373--1380, 2018.

Gur, R. E., Cowell, P. E., Latshaw, A., Turetsky, B. I., Grossman, R. I., Arnold, S. E., Bilker, W. B., and Gur, R. C. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. *Archives of general psychiatry*, 57(8):761--768, 2000.

Gutkin, B. S., Laing, C. R., Colby, C. L., Chow, C. C., and Ermentrout, G. B. Turning on and off with excitation: the role of spike-timing asynchrony and synchrony in sustained neural activity. *J Comput Neurosci*, 11(2):121--134, 2001.

Hansel, D. and Mato, G. Existence and stability of persistent states in large neuronal networks. *Phys. Rev. Let.*, 86(18):4175--4178, 2001.

Hansel, D. and Mato, G. Asynchronous states and the emergence of synchrony in large networks of interacting excitatory and inhibitory neurons. *Neural Comput*, 15(1):1--56, 2003.

Hansel, D. and Mato, G. Short-term plasticity explains irregular persistent activity in working memory tasks. *J. Neurosci.*, 33(1):133--149, 2013.

Hansel, D. and Sompolinsky, H. Synchronization and computation in a chaotic neural network. *Phys. Rev. Lett.*, 68(5):718--721, 1992.

Hansel, D. and Sompolinsky, H. Modeling feature selectivity in local cortical circuits. In Koch, C. and Segev, I., editors, *Methods in Neuronal Modeling: From Synapse to Network*, chapter 13. MIT Press, Cambridge, MA, 2nd edition, 1998.

Hansel, D. and van Vreeswijk, C. How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *J. Neurosci.*, 22:5118--5128, 2002.

Hansel, D., Mato, G., and Meunier, C. Synchrony in excitatory neural networks. *Neural computation*, 7(2):307--337, 1995.

Hebb, D. O. The first stage of perception: growth of the assembly. *The Organization of Behavior*, 4:60--78, 1949.

Hestrin, S., Nicoll, R., Perkel, D., and Sah, P. Analysis of excitatory synaptic action in pyramidal cells using whole-cell recording from rat hippocampal slices. *The Journal of Physiology*, 422(1):203--225, 1990.

Hikosaka, O. and Wurtz, R. H. Effects on eye movements of a gaba agonist and antagonist injected into monkey superior colliculus. *Brain research*, 272(2): 368--372, 1983.

Hitzig, E. *Untersuchungen über das Gehirn: Abhandlungen physiologischen und pathologischen Inhalts*. A. Hirschwald, 1874.

Hoff, A. L., Svetina, C., Shields, G., Stewart, J., and DeLisi, L. E. Ten year longitudinal study of neuropsychological functioning subsequent to a first episode of schizophrenia. *Schizophrenia research*, 78(1):27--34, 2005.

Hoffland, B. S., Bologna, M., Kassavetis, P., Teo, J. T., Rothwell, J. C., Yeo, C. H., van de Warrenburg, B. P., and Edwards, M. J. Cerebellar theta burst stimulation impairs eyeblink classical conditioning. *The Journal of physiology*, 590(4): 887--897, 2012.

Hoftman, G. D., Datta, D., and Lewis, D. A. Layer 3 excitatory and inhibitory circuitry in the prefrontal cortex: developmental trajectories and alterations in schizophrenia. *Biological psychiatry*, 81(10):862--873, 2017.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.*, 79(8):2554--2558, 1982.

Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10):3088--3092, 1984.

Hubel, D. H. and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture of the cat's visual cortex. *J. Physiol. (Lond.)*, 160:106--154, 1962.

Hubel, D. H. and Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.(Lond.)*, 195:215--243, 1968.

Hulse, B. K. and Jayaraman, V. Mechanisms underlying the neural computation of head direction. *Annual review of neuroscience*, 43:31--54, 2020.

Ilg, W., Christensen, A., Mueller, O. M., Goericke, S. L., Giese, M. A., and Timmann, D. Effects of cerebellar lesions on working memory interacting with motor tasks of different complexities. *Journal of neurophysiology*, 110(10):2337--2349, 2013.

Inagaki, H. K., Inagaki, M., Romani, S., and Svoboda, K. Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *Journal of Neuroscience*, 38(17):4163--4185, 2018.

Inagaki, H. K., Fontolan, L., Romani, S., and Svoboda, K. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature*, 566(7743): 212--217, 2019.

Ingvar, D. and Franzén, G. Distribution of cerebral activity in chronic schizophrenia. *The Lancet*, 304(7895):1484--1486, 1974.

Jacobsen, C. F. Functions of frontal association area in primates. *Archives of Neurology & Psychiatry*, 33(3):558--569, 1935.

Jacobsen, C. F., Elder, J. H., and Haslerud, G. M. Studies of cerebral function in primates. 1936.

Jentsch, J. D. and Roth, R. H. The neuropsychopharmacology of phencyclidine: from nmda receptor hypofunction to the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology*, 20(3):201--225, 1999.

Jokisch, D. and Jensen, O. Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *Journal of Neuroscience*, 27(12):3244--3251, 2007.

Jonas, P., Major, G., and Sakmann, B. Quantal components of unitary epscs at the mossy fibre synapse on ca3 pyramidal cells of rat hippocampus. *The Journal of physiology*, 472(1):615--663, 1993.

Jones, E. G. and Powell, T. P. S. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, 93:793--820, 1970.

Kamali, A., Kramer, L. A., Frye, R. E., Butler, I. J., and Hasan, K. M. Diffusion tensor tractography of the human brain cortico-ponto-cerebellar pathways: a quantitative preliminary study. *Journal of Magnetic Resonance Imaging*, 32(4): 809--817, 2010.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., and Mack, S. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

Kegeles, L. S., Abi-Dargham, A., Zea-Ponce, Y., Rodenhiser-Hill, J., Mann, J. J., Van Heertum, R. L., Cooper, T. B., Carlsson, A., and Laruelle, M. Modulation of amphetamine-induced striatal dopamine release by ketamine in humans: implications for schizophrenia. *Biological psychiatry*, 48(7):627--640, 2000.

Kelly, R. M. and Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *Journal of neuroscience*, 23(23):8432--8444, 2003.

Kent, J. S., Bolbecker, A. R., O'Donnell, B. F., and Hetrick, W. P. Eyeblink conditioning in schizophrenia: a critical review. *Frontiers in psychiatry*, 6: 146, 2015.

Kiehn, O. Plateau potentials and active integration in the 'final common pathway'for motor behaviour. *Trends in neurosciences*, 14(2):68--73, 1991.

Kim, S. S., Rouault, H., Druckmann, S., and Jayaraman, V. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849--853, 2017.

Kleppe, I. C. and Robinson, H. P. Determining the activation time course of synaptic ampa receptors from openings of colocalized nmda receptors. *Biophysical Journal*, 77(3):1418--1427, 1999.

Knierim, J. J. and Zhang, K. Attractor dynamics of spatially correlated neural activity in the limbic system. *Annual review of neuroscience*, 35:267--285, 2012.

Koenderink, M. T., Uylings, H., and Mrzljak, L. Postnatal maturation of the layer iii pyramidal neurons in the human prefrontal cortex: a quantitative golgi analysis. *Brain research*, 653(1-2):173--182, 1994.

Koene, R. A. and Hasselmo, M. E. First-in--first-out item replacement in a model of short-term memory based on persistent spiking. *Cerebral Cortex*, 17(8): 1766--1781, 2007.

Kraepelin, E. Die erforschung psychischer krankheitsformen. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 51(1):224--246, 1919.

Kritzer, M. F. and Goldman-Rakic, P. S. Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.*, 359(1):131--143, 1995.

Krystal, J. H., D'Souza, D. C., Petrakis, I. L., Belger, A., Berman, R. M., Charney, D. S., Abi-Saab, W., and Madonick, S. Nmda agonists and antagonists as probes of glutamatergic dysfunction and pharmacotherapies in neuropsychiatric disorders. *Harvard review of psychiatry*, 7(3):125--143, 1999.

Kubota, K. and Niki, H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of neurophysiology*, 34(3):337--347, 1971.

Lafon, B., Rahman, A., Bikson, M., and Parra, L. C. Direct current stimulation alters neuronal input/output function. *Brain stimulation*, 10(1):36--45, 2017.

Lahti, A. C., Koffel, B., LaPorte, D., and Tamminga, C. A. Subanesthetic doses of ketamine stimulate psychosis in schizophrenia. *Neuropsychopharmacology*, 13 (1):9--19, 1995.

Laing, C. R. and Chow, C. C. Stationary bumps in networks of spiking neurons. *Neural computation*, 13(7):1473--1494, 2001.

Larsen, J. K. and Divac, I. Selective ablations within the prefrontal cortex of the rat and performance of delayed alternation. *Physiological Psychology*, 6(1):15--17, 1978.

Latham, P. E., Richmond, B., Nelson, P., and Nirenberg, S. Intrinsic dynamics in neuronal networks. i. theory. *Journal of neurophysiology*, 83(2):808--827, 2000.

Laubach, M., Amarante, L. M., Swanson, K., and White, S. R. What, if anything, is rodent prefrontal cortex? *eneuro*, 5(5), 2018.

Ledoux, E. and Brunel, N. Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. *Frontiers in computational neuroscience*, 5:25, 2011.

Lee, J. and Park, S. Working memory impairments in schizophrenia: a meta-analysis. *Journal of abnormal psychology*, 114(4):599, 2005.

Lee, P., Lu, W.-S., Liu, C.-H., Lin, H.-Y., and Hsieh, C.-L. Test--retest reliability and minimal detectable change of the d2 test of attention in patients with schizophrenia. *Archives of Clinical Neuropsychology*, 33(8):1060--1068, 2018.

Leucht, S., Kane, J. M., Kissling, W., Hamann, J., Etschel, E., and Engel, R. R. What does the panss mean? *Schizophrenia research*, 79(2-3):231--238, 2005.

Levitt, J. B., Lewis, D. A., Yoshioka, T., and Lund, J. S. Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J. Comp. Neurol.*, 338(3):360--376, 1993.

Levy, R. and Goldman-Rakic, P. S. Association of storage and processing functions in the dorsolateral prefrontal cortex of the nonhuman primate. *Journal of Neuroscience*, 19(12):5149--5158, 1999.

Lewis, D. A. and Lieberman, J. A. Catching up on schizophrenia: natural history and neurobiology. *Neuron*, 28(2):325--334, 2000.

Lewis, D. A., Curley, A. A., Glausier, J. R., and Volk, D. W. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends in neurosciences*, 35(1):57--67, 2012.

Li, D.-b., Yao, J., Sun, L., Wu, B., Li, X., Liu, S.-l., Hou, J.-m., Liu, H.-l., Sui, J.-f., and Wu, G.-y. Reevaluating the ability of cerebellum in associative motor learning. *Scientific reports*, 9(1):1--15, 2019.

Li, D., Constantinidis, C., and Murray, J. D. Trial-to-trial variability of spiking delay activity in prefrontal cortex constrains burst-coding models of working memory. *Journal of Neuroscience*, 41(43):8928--8945, 2021. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.0167-21.2021. URL `https://www.jneurosci.org/content/41/43/8928`.

Li, N., Chen, T.-W., Guo, Z. V., Gerfen, C. R., and Svoboda, K. A motor cortex circuit for motor planning and movement. *Nature*, 519(7541):51--56, 2015.

Li, W.-K., Hausknecht, M. J., Stone, P., and Mauk, M. D. Using a million cell simulation of the cerebellum: Network scaling and task generality. *Neural networks*, 47:95--102, 2013.

Lincoln, J. S., McCORMICK, D. A., and Thompson, R. F. Ipsilateral cerebellar lesions prevent learning of the classically conditioned nictitating membrane/eyelid response. *Brain research*, 242(1):190--193, 1982.

Loewenstein, Y. and Sompolinsky, H. Temporal integration by calcium dynamics in a model neuron. *Nature neuroscience*, 6(9):961--967, 2003.

Lundqvist, M., Herman, P., and Miller, E. K. Working memory: delay activity, yes! persistent activity? maybe not. *Journal of neuroscience*, 38(32):7013--7019, 2018.

Manoach, D. S. Prefrontal cortex dysfunction during working memory performance in schizophrenia: reconciling discrepant findings. *Schizophrenia research*, 60(2-3):285--298, 2003.

Markow-Rajkowska, G. and Kosmal, A. Organization of cortical afferents to the frontal association cortex in dogs. *Acta Neurobiol. Exp*, 47:137--161, 1987.

Markowitsch, H. J. and Pritzel, M. Comparative analysis of prefrontal learning functions in rats, cats, and monkeys. *Psychological bulletin*, 84(5):817, 1977.

Martin, P. and Albers, M. Cerebellum and schizophrenia: a selective review. *Schizophrenia Bulletin*, 21(2):241--250, 1995.

Mauk, M. D., Li, W., Khilkevich, A., and Halverson, H. Cerebellar mechanisms of learning and plasticity revealed by delay eyelid conditioning. *International review of neurobiology*, 117:21--37, 2014.

McAllister, A. K. and Stevens, C. F. Nonsaturation of ampa and nmda receptors at hippocampal synapses. *Proceedings of the National Academy of Sciences*, 97 (11):6173--6178, 2000.

McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophysiol.*, 54:782--806, 1985.

McCormick, D. A., Clark, G. A., Lavond, D. G., and Thompson, R. F. Initial localization of the memory trace for a basic form of learning. *Proceedings of the National Academy of Sciences*, 79(8):2731--2735, 1982.

McCutcheon, R. A., Krystal, J. H., and Howes, O. D. Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry*, 19(1): 15--33, 2020.

Melchitzky, D. S., Sesack, S. R., Pucak, M. L., and Lewis, D. A. Synaptic targets of pyramidal neurons providing intrinsic horizontal connections in monkey prefrontal cortex. *Journal of Comparative Neurology*, 390(2):211--224, 1998.

Meyer-Lindenberg, A. and Bullmore, E. T. *Functional Brain Imaging in Schizophrenia*, chapter 17, pages 353--371. John Wiley and Sons, Ltd, 2010. ISBN 9781444327298. doi: https://doi.org/10.1002/9781444327298.ch17. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444327298.ch17.

Middleton, F. A. and Strick, P. L. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, 266(5184):458--461, 1994.

Middleton, F. A. and Strick, P. L. Cerebellar projections to the prefrontal cortex of the primate. *Journal of neuroscience*, 21(2):700--712, 2001.

Miller, K. D. and Troyer, W. T. Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.*, 87:653--659, 2002.

Mittleman, G., Goldowitz, D., Heck, D. H., and Blaha, C. D. Cerebellar modulation of frontal cortex dopamine efflux in mice: relevance to autism and schizophrenia. *Synapse*, 62(7):544--550, 2008.

Miyashita, Y. and Chang, H. S. Neuronal correlate of pictorial short-term memory in the primate temporal cortexyasushi miyashita. *Nature*, 331(6151):68--70, 1988.

Monaco, J., Casellato, C., Koch, G., and D'Angelo, E. Cerebellar theta burst stimulation dissociates memory components in eyeblink classical conditioning. *European Journal of Neuroscience*, 40(9):3363--3370, 2014.

Mongillo, G., Barak, O., and Tsodyks, M. Synaptic theory of working memory. *Science*, 319(5869):1543--1546, Mar 2008.

Mongillo, G., Hansel, D., and van Vreeswijk, C. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys. Rev. Lett.*, 108(15):158101, 2012.

Mongillo, G., Amit, D. J., and Brunel, N. Retrospective and prospective persistent activity induced by hebbian learning in a recurrent cortical network. *European Journal of Neuroscience*, 18(7):2011--2024, 2003.

Niki, H. Prefrontal unit activity during delayed alternation in the monkey. 1. relation to direction of response. *Brain Research*, 68:185--196, 1974a.

Niki, H. Prefrontal unit activity during delayed alternation in the monkey. ii. relation to absolute versus relative direction of response. *Brain research*, 68: 197--204, 1974b.

Notaras, M., Hill, R., and Van den Buuse, M. A role for the bdnf gene val66met polymorphism in schizophrenia? a comprehensive review. *Neuroscience & Biobehavioral Reviews*, 51:15--30, 2015.

Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M., and McCormick, D. A. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analysis. *J. Neurophysiol.*, 89:1541--1566, 2003.

Ott, T. and Nieder, A. Dopamine and cognitive control in prefrontal cortex. *Trends in cognitive sciences*, 23(3):213--234, 2019.

Pandya, D., Petrides, M., and Cipolloni, P. B. *Cerebral cortex: architecture, connections, and the dual origin concept*. Oxford University Press, 2015.

Parazzini, M., Rossi, E., Ferrucci, R., Liorni, I., Priori, A., and Ravazzani, P. Modelling the electric field and the current density generated by cerebellar transcranial dc stimulation in humans. *Clinical Neurophysiology*, 125(3):577--584, 2014.

Park, S., Holzman, P. S., et al. Schizophrenics show spatial working memory deficits. *Archives of general psychiatry*, 49:975--975, 1992.

Park, S., Holzman, P. S., and Goldman-Rakic, P. S. Spatial working memory deficits in the relatives of schizophrenic patients. *Archives of general psychiatry*, 52(10):821--828, 1995.

Park, S., Püschel, J., Sauter, B. H., Rentsch, M., and Hell, D. Spatial working memory deficits and clinical symptoms in schizophrenia: a 4-month follow-up study. *Biological Psychiatry*, 46(3):392--400, 1999.

Parker, K., Kim, Y., Kelley, R., Nessler, A., Chen, K., Muller-Ewald, V., Andreasen, N., and Narayanan, N. Delta-frequency stimulation of cerebellar projections can compensate for schizophrenia-related medial frontal dysfunction. *Molecular psychiatry*, 22(5):647--655, 2017.

Parker, K. L., Narayanan, N. S., and Andreasen, N. C. The therapeutic potential of the cerebellum in schizophrenia. *Frontiers in systems neuroscience*, 8:163, 2014.

Passingham, R. E. and Wise, S. P. *The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight.* Number 50. Oxford University Press, 2012.

Patel, K. R., Cherian, J., Gohil, K., and Atkinson, D. Schizophrenia: overview and treatment options. *Pharmacy and Therapeutics*, 39(9):638, 2014.

Perlstein, W. M., Carter, C. S., Noll, D. C., and Cohen, J. D. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *American Journal of Psychiatry*, 158(7):1105--1113, 2001.

Persi, E., Hansel, D., Nowak, L. G., Barone, P., and van Vreeswijk, C. Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex. *PLoS Comput. Biol.*, 7:e1001078, 2011.

Petrides, M. and Pandya, D. N. Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *Journal of Neuroscience*, 27(43): 11573--11586, 2007.

Preuss, T. M. Do rats have prefrontal cortex? the rose-woolsey-akert program reconsidered. *Journal of cognitive neuroscience*, 7(1):1--24, 1995.

Priebe, N. J. Mechanisms of orientation selectivity in the primary visual cortex. *Annual review of vision science*, 2:85--107, 2016.

Rainer, G., Asaad, W. F., and Miller, E. K. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, 393(6685): 577--579, 1998a.

Rainer, G., Asaad, W. F., and Miller, E. K. Memory fields of neurons in the primate prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 95(25):15008--15013, 1998b.

Rao, S. G., Williams, G. V., and Goldman-Rakic, P. S. Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in pfc. *Journal of neurophysiology*, 81(4):1903--1916, 1999.

Rogers, T. D., Dickson, P. E., Heck, D. H., Goldowitz, D., Mittleman, G., and Blaha, C. D. Connecting the dots of the cerebro-cerebellar role in cognitive function: Neuronal pathways for cerebellar modulation of dopamine release in the prefrontal cortex. *Synapse*, 65(11):1204--1212, 2011.

Rogers, T. D., McKimm, E., Dickson, P. E., Goldowitz, D., Blaha, C. D., and Mittleman, G. Is autism a disease of the cerebellum? an integration of clinical and pre-clinical research. *Frontiers in systems neuroscience*, 7:15, 2013.

Rolland, B., Jardri, R., Amad, A., Thomas, P., Cottencin, O., and Bordet, R. Pharmacology of hallucinations: several mechanisms for one single symptom? *BioMed Research International*, 2014, 2014.

Romani, S., Amit, D. J., and Mongillo, G. Mean-field analysis of selective persistent activity in presence of short-term synaptic depression. *Journal of Computational Neuroscience*, 20(2):201--217, 2006.

Romo, R., Brody, C. D., Hernandez, A., and Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399:470--473, 1999.

Rose, S. E., Chen, F., Chalk, J. B., Zelaya, F. O., Strugnell, W. E., Benson, M., Semple, J., and Doddrell, D. M. Loss of connectivity in alzheimer's disease: an evaluation of white matter tract integrity with colour coded mr diffusion tensor imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 69(4):528--530, 2000.

Roudi, Y. and Latham, P. E. A balanced memory network. *PLoS Comput. Biol.*, 3 (9):1679--1700, 2007.

Roxin, A., Brunel, N., and Hansel, D. Role of delays in shaping spatiotemporal dynamics of neuronal activity in large networks. *Phys. Rev. Lett.*, 94(23):238103, 2005.

Rubin, N. and Sompolinsky, H. Neural networks with low local firing rates. *EPL (Europhysics Letters)*, 10(5):465, 1989.

Rubio, M. D., Drummond, J. B., and Meador-Woodruff, J. H. Glutamate receptor abnormalities in schizophrenia: implications for innovative treatments. *Biomolecules & therapeutics*, 20(1):1, 2012.

Sah, P., Hestrin, S., and Nicoll, R. Properties of excitatory postsynaptic currents recorded in vitro from rat hippocampal interneurones. *The Journal of Physiology*, 430(1):605--616, 1990.

Salin, P. A. and Prince, D. A. Electrophysiological mapping of gabaa receptor-mediated inhibition in adult rat somatosensory cortex. *Journal of neurophysiology*, 75(4):1589--1600, 1996.

Schiff, M. L. and Reyes, A. D. Characterization of thalamocortical responses of regular-spiking and fast-spiking neurons of the mouse auditory cortex in vitro and in silico. *Journal of neurophysiology*, 107(5):1476--1488, 2012.

Schmidt, H. and Avitabile, D. Bumps and oscillons in networks of spiking neurons. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(3):033133, 2020.

Schmidt, H., Avitabile, D., Montbrió, E., and Roxin, A. Network mechanisms underlying the role of oscillations in cognitive tasks. *PLoS computational biology*, 14(9):e1006430, 2018.

Schwarz, C. and Schmitz, Y. Projection from the cerebellar lateral nucleus to precerebellar nuclei in the mossy fiber pathway is glutamatergic: a study combining anterograde tracing with immunogold labeling in the rat. *Journal of Comparative Neurology*, 381(3):320--334, 1997.

Seelig, J. D. and Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186--191, May 2015.

Sinclair, L. I., Kumar, A., Darreh-Shori, T., and Love, S. Visual hallucinations in alzheimer's disease do not seem to be associated with chronic hypoperfusion of to visual processing areas v2 and v3 but may be associated with reduced cholinergic input to these areas. *Alzheimer's research & therapy*, 11(1):1--15, 2019.

Sobczak-Edmans, M., Lo, Y.-C., Hsu, Y.-C., Chen, Y.-J., Kwok, F. Y., Chuang, K.-H., Tseng, W.-Y. I., and Chen, S. Cerebro-cerebellar pathways for verbal working memory. *Frontiers in human neuroscience*, 12:530, 2019.

Spencer, J. L., Collaborators, G., et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. 2018.

Spruston, N., Jonas, P., and Sakmann, B. Dendritic glutamate receptor channels in rat hippocampal ca3 and ca1 pyramidal neurons. *The Journal of physiology*, 482(2):325--352, 1995.

Srinivasan, R. and Chiel, H. J. Fast calculation of synaptic conductances. *Neural Computation*, 5:200--204, 1993.

Stahl, S. M. *Essential psychopharmacology: Neuroscientific basis and practical applications*. Cambridge university press, 2000.

Sterratt, D., Graham, B., Gillies, A., and Willshaw, D. *Principles of computational modelling in neuroscience*. Cambridge University Press, 2011.

Stoodley, C. J., Valera, E. M., and Schmahmann, J. D. Functional topography of the cerebellum for motor and cognitive tasks: an fmri study. *Neuroimage*, 59 (2):1560--1570, 2012.

Stoodley, C. J., D'Mello, A. M., Ellegood, J., Jakkamsetti, V., Liu, P., Nebel, M. B., Gibson, J. M., Kelly, E., Meng, F., Cano, C. A., et al. Altered cerebellar connectivity in autism and cerebellar-mediated rescue of autism-related behaviors in mice. *Nature neuroscience*, 20(12):1744--1751, 2017.

Stoykova, R., Matharan, F., Raoux, N., and Amieva, H. An alternative word-list for the free and cued selective reminding test (fcsrt): list presentation and reliability study. *Geriatrie et psychologie neuropsychiatrie du vieillissement*, 11 (3):317--322, 2013.

Takeda, K. and Funahashi, S. Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *J. Neurophysiol.*, 87(1):567--588, 2002.

Takeda, K. and Funahashi, S. Relationship between prefrontal task-related activity and information flow during spatial working memory performance. *Cortex*, 43 (1):38--52, 2007.

Takehara-Nishiuchi, K. The anatomy and physiology of eyeblink classical conditioning. *Behavioral neuroscience of learning and memory*, pages 297--323, 2016.

Tarullo, A. R., Isler, J. R., Condon, C., Violaris, K., Balsam, P. D., and Fifer, W. P. Neonatal eyelid conditioning during sleep. *Developmental psychobiology*, 58(7): 875--882, 2016.

Thompson, R. and Steinmetz, J. The role of the cerebellum in classical conditioning of discrete behavioral responses. *Neuroscience*, 162(3):732--755, 2009.

Thomson, A. M. Activity-dependent properties of synaptic transmission at two classes of connections made by rat neocortical pyramidal axons in vitro. *The Journal of physiology*, 502(1):131--147, 1997.

Timmann, D., Drepper, J., Frings, M., Maschke, M., Richter, S., Gerwig, M., and Kolb, F. P. The human cerebellum contributes to motor, emotional and cognitive associative learning. a review. *Cortex*, 46(7):845--857, 2010.

Tripathy, R. S. and Gerkin, R. Neuroelectro: organizing information on cellular neurophysiology. *URL: https://neuroelectro. org/*, 101, 2012.

Uylings, H. B., Groenewegen, H. J., and Kolb, B. Do rats have a prefrontal cortex? *Behavioural brain research*, 146(1-2):3--17, 2003.

van der Vliet, R., Jonker, Z. D., Louwen, S. C., Heuvelman, M., de Vreede, L., Ribbers, G. M., De Zeeuw, C. I., Donchin, O., Selles, R. W., van der Geest, J. N., et al. Cerebellar transcranial direct current stimulation interacts with bdnf val66met in motor learning. *Brain stimulation*, 11(4):759--771, 2018.

van Vreeswijk, C. and Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724--1726, 1996.

Van Vreeswijk, C., Abbott, L., and Ermentrout, G. B. When inhibition not excitation synchronizes neural firing. *Journal of computational neuroscience*, 1 (4):313--321, 1994.

van Vugt, B., van Kerkoerle, T., Vartak, D., and Roelfsema, P. R. The contribution of ampa and nmda receptors to persistent firing in the dorsolateral prefrontal cortex in working memory. *Journal of Neuroscience*, 40(12):2458--2470, 2020.

Waltz, J. A. and Gold, J. M. Motivational deficits in schizophrenia and the representation of expected value. *Behavioral Neuroscience of Motivation*, pages 375--410, 2015.

Wang, H.-X. and Gao, W.-J. Cell type-specific development of nmda receptors in the interneurons of rat prefrontal cortex. *Neuropsychopharmacology*, 34(8): 2028--2040, 2009.

Wang, K. H., Majewska, A., Schummers, J., Farley, B., Hu, C., Sur, M., and Tonegawa, S. In vivo two-photon imaging reveals a role of arc in enhancing orientation specificity in visual cortex. *Cell*, 126:389–402, 2006.

Wang, M., Yang, Y., Wang, C. J., Gamo, N. J., Jin, L. E., Mazer, J. A., Morrison, J. H., Wang, X. J., and Arnsten, A. F. NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron*, 77(4): 736--749, 2013.

Wang, X.-J. Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *Journal of Neuroscience*, 19:9587--9603, 1999.

Wang, X.-J. Synaptic reverberation underlying mnemonic persistent activity. *Trends in neurosciences*, 24(8):455--463, 2001.

Wang, X.-J. and Buzsáki, G. Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *Journal of neuroscience*, 16(20): 6402--6413, 1996.

Watson, T. C., Jones, M. W., and Apps, R. Electrophysiological mapping of novel prefrontal-cerebellar pathways. *Frontiers in integrative neuroscience*, 3:18, 2009.

Watson, T. C., Becker, N., Apps, R., and Jones, M. W. Back to front: cerebellar connections and interactions with the prefrontal cortex. *Frontiers in systems neuroscience*, 8:4, 2014.

Weinberger, D. R., Berman, K. F., and Zec, R. F. Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: I. regional cerebral blood flow evidence. *Archives of general psychiatry*, 43(2):114--124, 1986.

Weiss, C. and Disterhoft, J. F. Eyeblink conditioning, motor control, and the analysis of limbic-cerebellar interactions. *Behavioral and Brain Sciences*, 19(3): 479--481, 1996.

Wilson, H. R. and Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12(1):1--24, 1972.

Wilson, H. R. and Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55--80, 1973.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.*, 17(3):431--439, Mar 2014.

Xiang, Z., Huguenard, J. R., and Prince, D. A. Gabaa receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex. *the Journal of Physiology*, 506(3):715--730, 1998.

Zaitsev, A. V., Povysheva, N. V., Gonzalez-Burgos, G., and Lewis, D. A. Electrophysiological classes of layer 2/3 pyramidal cells in monkey prefrontal cortex. *Journal of neurophysiology*, 108(2):595--609, 2012.

Zaksas, D. and Pasternak, T. Directional signals in the prefrontal cortex and in area mt during a working memory for visual motion task. *Journal of Neuroscience*, 26(45):11726--11742, 2006.

Zhou, Y., Fan, L., Qiu, C., and Jiang, T. Prefrontal cortex and the dysconnectivity hypothesis of schizophrenia. *Neuroscience bulletin*, 31(2):207--219, 2015.

Zuchowski, M. L., Timmann, D., and Gerwig, M. Acquisition of conditioned eyeblink responses is modulated by cerebellar tdcs. *Brain stimulation*, 7(4): 525--531, 2014.

PART IV

TECHNICAL APPENDIX

A Mathematical Framework

Appendix A includes the mathematical description of the rate model for visuospatial working memory as well as the derivation of the equations describing the steady state dynamics and its stability.

A.1 General rate model

In the general model we define a population of N_E excitatory neurons and a population of N_I inhibitory neurons. In each population, each neuron is assigned with a preferred direction (PD) angle which is evenly distributed between $-\pi$ and π. In this case the PD angle of a neuron is defined as the angular direction of the external stimulus at which the response of the neuron is maximal. Each neuron receives three types of inputs: external stimulation I_{stim} relative to the visual cues, a background input C considered as a global input from adjacent cortical columns or other areas and a recurrent input I_{rec}^X that describes the synaptic interaction between the neurons in the network (where X is the postsynaptic population, $X \in \{E, I\}$). The output of a neuron from population X with PD θ is its firing rate $r_X(\theta, t)$, which, under the assumption of very slow synaptic time constants, can be computed instantaneously from the neuron's input-output transfer function (TF), $g_X(I)$:

$$r_X(\theta, t) = g_X\left(C + I_{rec}^X(\theta, t) + I_{stim}(\theta, t)\right) \tag{A.1}$$

In order to simplify the analytical calculations in the study of the effect of neuronal nonlinearities on the network, we chose to describe the TF as a piecewise linear

function. We define the general form of $g_X(I)$ as follows:

$$g_X(I) = \begin{cases} 0, & I < 0 \\ \alpha_X I, & 0 \leq I < T_X \\ \beta_X(I - T_X) + \alpha_X T_X, & T_X \leq I \end{cases} \tag{A.2}$$

The activity of the outgoing synapses of neuron with PD θ in population X is represented by the variable $m_X(\theta, t)$ which has linear dynamics, with the neuron's firing rate as input:

$$\tau_X \dot{m}_X(\theta, t) = -m_X(\theta, t) + r_X(\theta, t) \tag{A.3}$$

The recurrent input $I^X_{rec}(\theta, t)$ is described as follows:

$$I^X_{rec}(\theta, t) = \tfrac{1}{2\pi} \textstyle\int_{2\pi} J_{XE}(\theta - \xi) m_E(\xi, t) \mathrm{d}\xi - \tfrac{1}{2\pi} \int_{2\pi} J_{XI}(\theta - \xi) m_I(\xi, t) \mathrm{d}\xi \tag{A.4}$$

where $J_{XY}(\theta - \xi)$ is the weight of the interaction between presynaptic neuron with PD ξ in population Y and the postsynaptic neuron with PD θ in population X. From now on we consider that the connectivity is independent of the postsynaptic population i.e. the nature of the interaction is determined by the presynaptic population, $J_{EY} = J_{IY} \triangleq J_Y$.

Neurons with similar PDs are strongly connected, this feature is represented by the connectivity functions chosen as in Ben-Yishai et al. (1995):

$$J_Y(\theta - \xi) = J^Y_0 + J^Y_1 \cos(\theta - \xi) \tag{A.5}$$

where J^Y_0 represents the mean connectivity and J^Y_1 the spatial modulation. Since the connectivity is independent on X, by using the trigonometrical identity $\cos(\theta - \xi) = \cos(\theta)\cos(\xi) + \sin(\theta)\sin(\xi)$, Equation (A.4) can be rewritten as:

$$\begin{aligned} I_{rec}(\theta, t) = J^E_0 m^E_0(t) - J^I_0 m^I_0(t) + \left(J^E_1 m^E_1(t) - J^I_1 m^I_1(t)\right) \cos(\theta) \\ + \left(J^E_1 m^E_2(t) - J^I_1 m^I_2(t)\right) \sin(\theta) \end{aligned} \tag{A.6}$$

where $m^Y_i(t)$ are the Fourier components of the activity:

$$m^Y_0(t) = \frac{1}{2\pi} \int_{2\pi} m_Y(\theta, t) \mathrm{d}\theta \tag{A.7}$$

$$m_1^Y(t) = \frac{1}{2\pi}\int_{2\pi} m_Y(\theta, t)\cos(\theta)\mathrm{d}\theta \tag{A.8}$$

and

$$m_2^Y(t) = \frac{1}{2\pi}\int_{2\pi} m_Y(\theta, t)\sin(\theta)\mathrm{d}\theta \tag{A.9}$$

A.2 Steady state of the general model

In the absence of I_{stim}, the network presents two different types of steady states or fixed points (FPs). These equilibrium points are a homogeneous PF in which all the neurons display the same firing rate, and a “bump" FP in which a specific group of neurons with similar PD fire at higher rates than the rest of the neurons. Due to the rotational symmetry of the network, we can describe bump steady states profiles symmetric around $\theta = 0$ without loss of generality. The general equation for steady states can be found by combining Equations (A.1), (A.3) and (A.6) and setting to zero the left-hand side (LHS) in Equation (A.3):

$$m_X(\theta) = g_X\left(C + I_0 + I_1\cos(\theta)\right) \tag{A.10}$$

where $X \in \{E, I\}$ and $I_k \triangleq J_k^E m_k^E - J_k^I m_k^I$ with $k \in \{0, 1\}$. Notice that, since we are describing the bump state profiles centered at$\theta = 0$, the integral of the sine for the calculation of m_2^Y in Equation (A.9) is equal to zero. Thus, the total input to a neuron in the steady state is defined by:

$$I^*(\theta) \triangleq C + I_0 + I_1\cos(\theta) \tag{A.11}$$

In the homogeneous steady state, since all neurons have the same firing rate, the integrals of the cosine for the calculation of m_1^Y in Equation (A.8) is equal to zero leading to $I_1 = 0$ and thus Equation (A.11) for the homogeneous steady state becomes:

$$m_X(\theta) = g_X\left(C + I_0\right) \tag{A.12}$$

Using the definition of I_0 together with Equation (A.12) we obtain:

$$I_0 = F_0(C + I_0) \tag{A.13}$$

where $F_0(x) \triangleq J_0^E g_E(x) - J_0^I g_I(x)$. Thus, the homogeneous steady state firing rate can be obtained by solving Equation (A.13) and then substituting I_0 into equation A.12.

The solution of the bump steady state can be characterized by three parameters: θ_0, as the maximal value of θ at which the input to neuron with PD θ (i.e $I^*(\theta)$), is above zero; θ_1^E and θ_1^I below which $I^*(\theta)$ is larger than T_E and T_I, respectively. Substituting Equation (A.10) in the definition of the Fourier components (Equations (A.7) to (A.9)) and taking into account the definition of the TF in Equation (A.2) it yields:

$$\begin{aligned} m_i^X(t) = (\beta_X - \alpha_X)\frac{1}{\pi}\int_0^{\theta_1^X} [I^*(\theta) - T_X]\cos(i\cdot\theta)\mathrm{d}\theta \\ + \alpha_X \frac{1}{\pi}\int_0^{\theta_0} I^*(\theta)\cos(i\cdot\theta)\mathrm{d}\theta \end{aligned} \tag{A.14}$$

where $i \in \{0, 1\}$. Since we consider profiles symmetric around $\theta = 0$ the m_2^Y component will be zero. This also allows to define the integration limits from zero to θ_i^X. In the bump steady state, there is a the case in which θ_0, θ_1^E and θ_1^I are all smaller than π, meaning that certain group of neurons is receiving a total input smaller than zero, other group receives an input between zero and T_X and other group receives an input above T_X, the equations defining these parameters are:

$$C + I_0 + I_1\cos(\theta_0) = 0 \tag{A.15}$$

and

$$C + I_0 + I_1\cos(\theta_1^X) = T_X \tag{A.16}$$

Combining Equation (A.16) for the excitatory and inhibitory populations and substituting I_1 from Equation (A.15) we obtain the following relation:

$$(T_E - T_I)\cos(\theta_0) = T_E\cos(\theta_1^I) - T_I\cos(\theta_1^E) \tag{A.17}$$

Now, using Equations (A.11), (A.15) and (A.16) we can rewrite Equation (A.14) as:

$$m_i^X = I_1\left((\beta_X - \alpha_X)f_i(\theta_1^X) + \alpha_X f_i(\theta_0)\right) \tag{A.18}$$

where

$$f_i(\theta) = \frac{1}{\pi}\int_0^{\theta} (\cos(\xi) - \cos(\theta))\cos(i \cdot \xi)\mathrm{d}\xi \tag{A.19}$$

with $\theta \in \{\theta_0, \theta_1^X\}$ and $i \in \{0, 1\}$.

Substituting Equation (A.18) into the definition of I_1 we derive a second equation for θ_0, θ_1^E and θ_1^I:

$$J_1^E(\beta_E - \alpha_E)f_1(\theta_1^E) - J_1^I(\beta_I - \alpha_I)f_1(\theta_1^I) + (J_1^E\alpha_E - J_1^I\alpha_I)f_1(\theta_0) = 1 \tag{A.20}$$

Then, taking Equation (A.18) together with the definition of I_0 we can obtain:

$$I_0 = I_1\left[J_0^E(\beta_E - \alpha_E)f_0(\theta_1^E) - J_0^I(\beta_I - \alpha_I)f_0(\theta_1^I) + (J_0^E\alpha_E - J_0^I\alpha_I)f_0(\theta_0)\right] \tag{A.21}$$

Now, substituting Equation (A.21) in Equations (A.15) and (A.16)) we can write a third equation for θ_0, θ_1^E and θ_1^I:

$$\begin{aligned} &J_0^E(\beta_E - \alpha_E)f_0(\theta_1^E) - J_0^I(\beta_I - \alpha_I)f_0(\theta_1^I) + (J_0^E\alpha_E - J_0^I\alpha_I)f_0(\theta_0) = \\ &-\frac{C}{T_E}\left[\cos(\theta_1^E) - \cos(\theta_0\right] - \cos(\theta_0) \end{aligned} \tag{A.22}$$

At this point, from Equations (A.17), (A.20) and (A.22) we can solve the system numerically in order to find θ_0, θ_1^E and θ_1^I for the case in which these angles are smaller than π. Once we find the solution of these angles we can substitute them in Equations (A.15) and (A.16) to find the values of I_k, and finally calculate the steady state rates from Equation (A.10).

In the case in which the input is above zero for all the neurons, i.e $I^*(\theta) > 0;\ \forall\ \theta \in \{-\pi, \pi\}$, Equation (A.15) is no longer valid. Instead, we substitute π for θ_0 in Equation (A.14) to obtain:

$$m_0^X = I_1(\beta_X - \alpha_X)f_0(\theta_1^X) + \alpha_X\,(C + I_0) \tag{A.23}$$

and

$$m_1^X = I_1\left[(\beta_X - \alpha_X)f_1(\theta_1^X) + \frac{\alpha_X}{2}\right] \tag{A.24}$$

Substituting Equation (A.24) into the definition of I_1 we can write and equation for θ_1^E and θ_1^I:

$$J_1^E(\beta_E - \alpha_E)f_1(\theta_1^E) - J_1^I(\beta_I - \alpha_I)f_1(\theta_1^I) + \frac{1}{2}(J_1^E\alpha_E - J_1^I\alpha_I) = 1 \tag{A.25}$$

By substituting Equation (A.23) into the definition of I_0 and using Equation (A.16) we can get a second equation for θ_1^E and θ_1^I:

$$\frac{J_0^E(\beta_E - \alpha_E) f_0(\theta_1^E) - J_0^I(\beta_I - \alpha_I) f_0(\theta_1^I) + \frac{C\left[\cos(\theta_1^E) - \cos(\theta_1^I)\right]}{T_E - T_I}}{1 - (J_0^E \alpha_E - J_0^I \alpha_I)} = \frac{T_I \cos(\theta_1^I) - T_E \cos(\theta_1^E)}{T_E - T_I} \tag{A.26}$$

By solving numerically Equations (A.26) and (A.31) we will found the solution for θ_1^E and θ_1^I in the steady states. Then, with Equation (A.16) we can compute I_0 and I_1 and finally evaluate the steady state rate with Equations (A.23) and (A.24).

There is a third case in which the input to all the neurons is above zero and T_E ($I^*(\theta) > T_E\,;\ \forall\ \theta \in \{-\pi, \pi\}$). In this case, in which we consider $T_E \leq T_I$, a group of neurons receive inputs between T_E and T_I (neurons with PD between θ_1^E and θ_1^I), and another group receive inputs larger than T_I (neurons with PD larger than θ_1^I). In this case we substitute π for θ_0 and θ_1^E in Equation (A.14) to obtain the steady state rate equations:

$$m_0^E = \beta_E(C + I_0 - T_E) + \alpha_E T_E \tag{A.27}$$

$$m_1^E = \beta_E I_1/2 \tag{A.28}$$

$$m_0^I = I_1(\beta_I - \alpha_I) f_0(\theta_1^I) + \alpha_I\,(C + I_0) \tag{A.29}$$

and

$$m_1^I = I_1\left[(\beta_I - \alpha_I) f_1(\theta_1^I) + \frac{\alpha_I}{2}\right] \tag{A.30}$$

Substituting Equations (A.28) and (A.30) into the definition of I_1 we obtain a equation for θ_1^I :

$$J_1^E \beta_E/2 - J_1^I\left[(\beta_I - \alpha_I) f_1(\theta_1^I) + \alpha_I/2\right] = 1 \tag{A.31}$$

From this equation we can find numerically θ_1^I, calculate I_0 and I_1 using Equations (A.16), (A.27) and (A.29) and the definition of I_0, and finally calculate the steady state rate using equations **??--??**

A.3 Reduction to one population

In order to study explicitly the effect of nonlinearities in the neurons TF, we can reduce the model to one effective population by considering that the TFs (Equation (A.2)) and the time constants for excitatory and inhibitory neurons are identical. Thus, the rate dynamics (Equation (A.3)) of both populations is exactly the same and it can be described by a single variable $m(\theta, t)$ with dynamics

$$\tau \dot{m}(\theta,t) = -m(\theta,t) + g\left(C + J_0 m_0(t) + J_1 m_1(t)\cos(\theta) + J_1 m_2(t)\sin(\theta) + I_{stim}(\theta,t)\right). \quad \text{(A.32)}$$

where $J_i = J_i^E - J_i^I$ and $m_i(t) = m_i^E(t) = m_i^I(t)$. The same network was previously described in Ben-Yishai et al. (1995) with a threshold-linear TF. In our work, in order to investigate the effects of the neural nonlinearities, we choose the TF to be a piecewise-linear function as described in Equation (A.2) taking $T_E = T_I = 1$, $\alpha_E = \alpha_I = 1$ and $\beta_E = \beta_I = \beta$:

$$g(I) = \begin{cases} 0, & I < 0 \\ I, & 0 \leq I < 1 \\ \beta(I-1)+1, & 1 \leq I \end{cases} \quad \text{(A.33)}$$

As in the previous section we only address the steady state profiles that are symmetric around $\theta = 0$. In this case the general equation for the steady states is:

$$m(\theta) = g\left[C + J_0 m_0 + J_1 m_1 \cos(\theta)\right] \quad \text{(A.34)}$$

In the homogeneous state all neurons share the same rate, thus $m_1 = 0$, and it can be determined by solving the following equation:

$$I^* - C = J_0 g(I^*) \quad \text{(A.35)}$$

where $I^* = C + J_0 m_0$. The stability analysis given by the linearisation of Equation (A.35) about the FP yields the stability condition $J_0 g'(I^*) < 1$. And using Equation (A.34) into **??** and linearising around I^* yields a second stability condition for the homogeneous FP $J_1 g'(I^*) < 2$ (where $g'(I^*)$ is the derivative of

$g(I)$ at the point I^*). These two stability conditions are the same as described in Ben-Yishai et al. (1995). The two different cases $\beta > 1$ and $\beta < 1$ are shown in C *FigureXX*.

Since $T_E = T_I$, the bump solution in the reduced model is characterised by only two angles: θ_0 and θ_1 ($\theta_1^E = \theta_1^I \triangleq \theta_1$). From Equation (A.14) we can define the integrals of the form $G_i(\theta_0, \theta_1) = \frac{1}{\pi}\left(\int_0^{\theta_0} \cos^i(\theta)\mathrm{d}\theta + (\beta - 1)\int_0^{\theta_1} \cos^i(\theta)\mathrm{d}\theta\right)$, and a matrix

$$\mathbf{A}\left(\theta_0, \theta_1\right) = \begin{bmatrix} G_0\left(\theta_0, \theta_1\right) J_0 - 1 & G_1\left(\theta_0, \theta_1\right) J_1 \\ G_1\left(\theta_0, \theta_1\right) J_0 & G_2\left(\theta_0, \theta_1\right) J_1 - 1 \end{bmatrix} \tag{A.36}$$

then the steady state equations of the order parameters m_0 and m_1 for the bump state can be defined in matrix form in which m_0 and m_1 obey the linear relation:

$$\mathbf{A}\left(\theta_0, \theta_1\right) \cdot \begin{bmatrix} m_0 \\ m_1 \end{bmatrix} = \frac{(\beta - 1)}{\pi} \begin{bmatrix} \theta_1 \\ \sin(\theta_1) \end{bmatrix} - C \cdot \begin{bmatrix} G_0\left(\theta_0, \theta_1\right) \\ G_1\left(\theta_0, \theta_1\right) \end{bmatrix} \tag{A.37}$$

For the case in which θ_0 and θ_1 are smaller than π, we can find their values by numerically solving the following two equations:

$$J_0\left(f_0(\theta_0) + (\beta - 1) f_0(\theta_1)\right) = C \cdot (\cos(\theta_0) - \cos(\theta_1)) - \cos(\theta_0) \tag{A.38}$$

and

$$J_1\left(f_1(\theta_0) + (\beta - 1) f_1(\theta_1)\right) = 1 \tag{A.39}$$

where $f_i(\theta)$ are as defined in Equation (A.19). These equations are obtained from the definition of m_0 and m_1 in Equation (A.37) together with Equations (A.15) and (A.16). In the case where all the neurons receive an input larger than zero, the value of θ_1 can be found from Equation (A.39) by substituting θ_0 by π and solving numerically. The existence of solution of the bump state is restricted to certain values of the background input C. In order to find the critical values of C we use Equation (A.38) to have an expression of C as a function of the two angles (θ_0 and θ_1), and then search for the extreme of this function under the constraint given by Equation (A.39) by using Lagrange multipliers method. We find that the values of θ_i at the extrema can be found by numerically solving Equation (A.39) with the equation:

$$\Psi_0 + (\beta - 1)\,\Psi_1 = 0 \tag{A.40}$$

where

$$\Psi_0 = \left[J_0 \left(f_0(\theta_0) + (\beta - 1) \frac{\sin(\theta_1) - \theta_1 \cos(\theta_0)}{\pi} \right) + \cos(\theta_0) \right] \cdot \sin(\theta_0) \quad \text{(A.41)}$$

and

$$\Psi_1 = \left[J_0 \left(\frac{\sin(\theta_0) - \theta_0 \cos(\theta_1)}{\pi} + (\beta - 1) f_0(\theta_1) \right) + \cos(\theta_1) \right] \cdot \sin(\theta_1) \quad \text{(A.42)}$$

Once we find θ_i, we substitute the values in Equation (A.38) to find the critical value of C.

There is another situation in which the bump state does not exist and $m_i \to \infty$. In order to find the critical value of C in this case we take the limit $\theta_1 \to \theta_0$. By substituting $\theta_0 = \theta_1 \triangleq \theta_c$ in Equation (A.38) we obtain the following equation:

I

$$\beta J_0 \frac{1}{\pi} \left(\theta_c - \tan(\theta_c) \right) = 1 \quad \text{(A.43)}$$

From Equation (A.43) we can numerically find the value of θ_c and then substituting θ_0 and θ_1 by θ_c in Equation (A.39) we can find the maximal value of for which the bump state exists. Above this critical value of the system diverges (note that this value is independent on C and only depends on the connectivity parameter J_0). In order to analyze the stability of the bump steady states we consider the response of the network to a small perturbation away from the FP described as $m(\theta, t) = m^*(\theta) + \delta m(\theta, t)$, where $m^*(\theta)$ is a FP that satisfies Equation (A.34) and $\delta m(\theta, t)$ is the perturbation. To see wether the perturbation grows or decays, we derive a differential equation for the perturbation $\delta \dot{m}(\theta, t) = \frac{d}{dt}(m(\theta, t) - m^*(\theta)) = \dot{m}(\theta, t)$ (since $m^*(\theta)$ is constant). Thus, $\delta \dot{m}(\theta, t) = \dot{m}(\theta, t) = f(m(\theta, t)) = f(m^*(\theta) + \delta m(\theta, t))$, where $f(x)$ corresponds to the RHS in Equation (A.32). By linearizing this function around the FP using Taylor's expansion we obtain that the dynamics of the perturbation is governed by:

$$\tau \dot{\delta m}(\theta, t) = -\delta m(\theta, t) + g'\left(I^*(\theta)\right) \cdot [J_0 \delta m_0(t) + J_1 \delta m_1(t) \cos(\theta) + J_1 \delta m_2(t) \sin(\theta)] \quad \text{(A.44)}$$

where

$$\delta m_0(t) = \frac{1}{2\pi} \int_{2\pi} \delta m(\theta, t) \mathrm{d}\theta \quad \text{(A.45)}$$

$$\delta m_1(t) = \frac{1}{2\pi}\int_{2\pi} \delta m(\theta,t)\cos(\theta)\mathrm{d}\theta \tag{A.46}$$

$$\delta m_2(t) = \frac{1}{2\pi}\int_{2\pi} \delta m(\theta,t)\sin(\theta)\mathrm{d}\theta \tag{A.47}$$

I^* is the total input to neuron with PD θ at the FP (see Equation (A.35)) and $g\prime(I)$ is the derivative of the TF $g(I)$:

$$g'(I) = \begin{cases} 0, & I < 0 \\ 1, & 0 \leq I \leq 1 \\ \beta, & 1 < I \end{cases} \tag{A.48}$$

By differentiating Equations (A.45) and (A.46) with respect to time and using Equation (A.44) we obtain the equations for the dynamics of δm_0 and δm_1:

$$\tau\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\delta m_0 \\ \delta m_1\end{bmatrix} = \mathbf{A}\left(\theta_0,\theta_1\right)\cdot\begin{bmatrix}\delta m_0 \\ \delta m_1\end{bmatrix} \tag{A.49}$$

where $\mathbf{A}$ is the matrix defined in Equation (A.36). Due to the symmetry of the network, the phase of the bumps state at the steady state is arbitrary and thus $\delta\dot{m}_2 = 0$. The stability requirement is that the eigenvalues of the matrix $\mathbf{A}$ at the steady state $m(\theta)$ must be negative. The two stability conditions are:

$$G_0\left(\theta_0,\theta_1\right)\cdot\left(J_0+J_1\right) < 3 \tag{A.50}$$

and

$$\left(J_0G_0\left(\theta_0,\theta_1\right)-1\right)\cdot\left(J_1G_2\left(\theta_0,\theta_1\right)-1\right) - J_0J_1G_1\left(\theta_0,\theta_1\right)^2 > 0 \tag{A.51}$$

To obtain the flow diagrams in the activity space in Figure 3.2 B and D, we write the dynamics of m_k in the absence of an external stimulus by Fourier transforming Equation (A.32):

$$\begin{aligned}&\tau\dot{m}_0 = \\ &-m_0 + \frac{1}{2\pi}\int_{2\pi} g\left(C + J_0m_0 + J_1m_1\cos(\theta) + J_1m_2\sin(\theta)\right)\mathrm{d}\theta\end{aligned} \tag{A.52}$$

$$\begin{aligned}&\tau\dot{m}_0 = \\ &-m_0 + \frac{1}{2\pi}\int_{2\pi} g\left(C + J_0m_0 + J_1m_1\cos(\theta) + J_1m_2\sin(\theta)\right)\cos(\theta)\mathrm{d}\theta\end{aligned} \tag{A.53}$$

$$\tau \dot{m_0} = \\ -m_0 + \frac{1}{2\pi} \int_{2\pi} g\left(C + J_0 m_0 + J_1 m_1 \cos(\theta) + J_1 m_2 \sin(\theta)\right) \sin(\theta) \mathrm{d}\theta \tag{A.54}$$

The flow on the $m_0 - m_1$ plane reflects the case where the activity profile is completely symmetric around $\theta = 0$, thus $m_2 = 0$. Given the symmetry of the system we can tell that the flow is symmetric about the m_0 axis and therefore the lower half plane is a mirror image of the upper half plane ($+m_1$). In the upper half plane the flow is given by:

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} m_0 \\ m_1 \end{bmatrix} = \mathbf{A}\left(\theta_0, \theta_1\right) \cdot \begin{bmatrix} m_0 \\ m_1 \end{bmatrix} - \frac{(\beta - 1)}{\pi} \begin{bmatrix} \theta_1 \\ \sin(\theta_1) \end{bmatrix} + C \cdot \begin{bmatrix} G_0\left(\theta_0, \theta_1\right) \\ G_1\left(\theta_0, \theta_1\right) \end{bmatrix} \tag{A.55}$$

where the angles θ_k can be computed by defining $u_k = \frac{k - C - J_0 m_0}{J_1 m_1}$ from Equations (A.15) and (A.16) for one population and then:

$$\theta_k = \begin{cases} \pi, & u_k < -1 \\ \cos^{-1}(u_k), & -1 \leq u_k \leq 1 \\ 0, & 1 < u_k \end{cases} \tag{A.56}$$

In order to find the stable branch of the unstable FP defining the separatrix, we first calculate the normalized eigenvectors of $\mathbf{A}$ at this FP. Then, by using Equation (A.55), we calculate the trajectory in time starting near the FP in the stable eigendirection (direction of the eigenvector with negative eigenvalue) backwards in time.

A.4 Homogeneous state stability in the general model

As done in the reduced model, we study the dynamics of the network of the general model following a small perturbation from the homogeneous FP expressed as $m_X(\theta, t) = m_X^* + \delta m_X(\theta, t)$ where m_X^* is the activity at the homogeneous FP of population$X$, $X \in \{E, I\}$. The dynamics of the perturbation is described as:

$$\tau_X \delta \dot{m}_X(\theta, t) = -\delta m_X(\theta, t) + g_X'\left(I^*\right) \cdot \delta I(\theta, t) \tag{A.57}$$

where

$$\begin{aligned}\delta I(\theta,t) = J_0^E \delta m_0^E(t) - J_0^I \delta m_0^I(t) + \left(J_1^E \delta m_1^E(t) - J_1^I \delta m_1^I(t)\right)\cos(\theta) \\ + \left(J_1^E \delta m_2^E(t) - J_1^I \delta m_2^I(t)\right)\sin(\theta)\end{aligned} \tag{A.58}$$

the derivative of the TF $g_X(I)$ is

$$g_X'(I) = \begin{cases} 0, & I < 0 \\ \alpha_X, & 0 \leq I < T_X \\ \beta_X, & T_X \leq I \end{cases} \tag{A.59}$$

and the order parameters δm_i^X are

$$\delta m_0^X(t) = \frac{1}{2\pi}\int_{2\pi} \delta m_X(\theta,t)\mathrm{d}\theta \tag{A.60}$$

$$\delta m_1^X(t) = \frac{1}{2\pi}\int_{2\pi} \delta m_X(\theta,t)\cos(\theta)\mathrm{d}\theta \tag{A.61}$$

$$\delta m_2^X(t) = \frac{1}{2\pi}\int_{2\pi} \delta m_X(\theta,t)\sin(\theta)\mathrm{d}\theta \tag{A.62}$$

By substituting the definition of $\delta m_X(\theta,t)$ from Equation (A.57) into the order parameters equations, we obtain:

$$\begin{bmatrix} \delta m_i^E \\ \delta m_i^I \end{bmatrix} = \begin{bmatrix} \left(\widetilde{J}_i^{EE} - 1\right)/\tau_E & -\widetilde{J}_i^{EI}/\tau_E \\ \widetilde{J}_i^{IE}/\tau_I & -\left(\widetilde{J}_i^{II} + 1\right)/\tau_I \end{bmatrix} \cdot \begin{bmatrix} \delta m_i^E \\ \delta m_i^I \end{bmatrix} \tag{A.63}$$

where $\widetilde{J}_i^{XY}$ is defined as $g_X'(I^*)J_0^Y$ if $i = 0$ and $g_X'(I^*)J_1^Y/2$ otherwise.

The stability condition is that the eigenvalues of the matrix must be negative, in order to accomplish this, the determinant has to be positive and the trace negative. The conditions for the determinant to be positive are:

$$g_E'(I^*)J_0^E - g_I'(I^*)J_0^I < 1 \tag{A.64}$$

and

$$g_E'(I^*)J_1^E - g_I'(I^*)J_1^I < 2 \tag{A.65}$$

Notice that this two conditions are independent on the time constants τ_E and τ_I. Then, the two conditions for the negative trace are:

$$\tau_E/\tau_I > \frac{g_E'(I^*)J_0^E - 1}{g_I'(I^*)J_0^I + 1} \tag{A.66}$$

and

$$\tau_E/\tau_I > \frac{g'_E(I^*)J_1^E/2 - 1}{g'_I(I^*)J_1^I/2 + 1} \tag{A.67}$$

In the case $\tau_E = \tau_I$, the conditions for the trace will be automatically fulfilled provided that the conditions for the determinant are also satisfied.

A.5 Bump state stability in the general model

From the Equation (A.57) but in this case taking $I^* = C + I_0 + I_1\cos(\theta)$ (as given in Equation (A.11)), we define the steady state profile values:

$$K_i^X = \frac{1}{\pi}\left(\alpha_X \int_0^{\theta_0} \cos^i(\theta)\mathrm{d}\theta + (\beta_X - \alpha_X)\int_0^{\theta_1^X} \cos^i(\theta)\mathrm{d}\theta\right) \tag{A.68}$$

Then, using Equation (A.57) and the definitions of δm_0^X and δm_1^X we can derive the linearized dynamics of the perturbation:

$$\tau_E \frac{d}{dt}\begin{bmatrix} \delta m_0^E \\ \delta m_0^I \\ \delta m_1^E \\ \delta m_1^I \end{bmatrix} = \mathbf{K}\cdot\begin{bmatrix} \delta m_0^E \\ \delta m_0^I \\ \delta m_1^E \\ \delta m_1^I \end{bmatrix} \tag{A.69}$$

where

$$\mathbf{K} \triangleq \begin{bmatrix} J_0^E K_0^E - 1 & -J_0^I K_0^E & J_1^E K_1^E & -J_1^I K_1^E \\ \frac{\tau_E}{\tau_I} J_0^E K_0^I & -\frac{\tau_E}{\tau_I}(J_0^I K_0^I + 1) & \frac{\tau_E}{\tau_I} J_1^E K_1^I & -\frac{\tau_E}{\tau_I} J_1^I K_1^I \\ J_0^E K_1^E & -J_0^I K_1^E & J_1^E K_2^E - 1 & -J_1^I K_2^E \\ \frac{\tau_E}{\tau_I} J_0^E K_1^I & -\frac{\tau_E}{\tau_I} J_0^I K_1^I & \frac{\tau_E}{\tau_I} J_1^E K_2^I & -\frac{\tau_E}{\tau_I}(J_1^I K_2^I + 1) \end{bmatrix} \tag{A.70}$$

The bump state is only stable if all the eigenvalues of $\mathbf{K}$ have negative real parts. These eigenvalues can be found numerically. The definition of δm_2^X together with Equation (A.57) yields linearized dynamics for δm_2^X:

$$\tau_E \frac{d}{dt}\begin{bmatrix} \delta m_2^E \\ \delta m_2^I \end{bmatrix} = \mathbf{B}\cdot\begin{bmatrix} \delta m_2^E \\ \delta m_2^I \end{bmatrix} \tag{A.71}$$

where

$$\mathbf{B} \triangleq \begin{bmatrix} J_1^E\left[(\beta_E-\alpha_E)f_1(\theta_1^E)+\alpha_E f_1(\theta_0)\right]-1 & -J_1^I\left[(\beta_E-\alpha_E)f_1(\theta_1^E)+\alpha_E f_1(\theta_0)\right] \\ \frac{\tau_E}{\tau_I}J_1^E\left[(\beta_I-\alpha_I)f_1(\theta_1^I)+\alpha_I f_1(\theta_0)\right] & -\frac{\tau_E}{\tau_I}\left(J_1^I\left[(\beta_I-\alpha_I)f_1(\theta_1^I)+\alpha_I f_1(\theta_0)\right]-1\right) \end{bmatrix} \tag{A.72}$$

In order for the eigenvalues of $\mathbf{B}$ to be negative, the determinant of $\mathbf{B}$ must be non-negative and the trace must be negative. In this case, because of the steady state Equation (A.20) we find that the determinant is always zero. The demand for the negative trace leads to the condition

$$\tau_E/\tau_I > \frac{J_1^E\left[(\beta_E-\alpha_E)f_1(\theta_1^E)+\alpha_E f_1(\theta_0)\right]-1}{J_1^I\left[(\beta_I-\alpha_I)f_1(\theta_1^I)+\alpha_I f_1(\theta_0)\right]+1} \tag{A.73}$$

B Spiking Network Model

B.1 Two Population Network of Leaky integrate-and-fire neurons

The network is composed of N_E excitatory and N_I inhibitory leaky integrate-and-fire neurons connected all-to-all. As in the rate model, the neurons are labeled with their PDs. In this case, the model describes the dynamics of the voltage membrane of neuron with PD θ in the postsynaptic population $X \in \{E, I\}$. The voltage membrane $V_X(\theta, t)$ in the subthreshold regime evolves in time as follows:

$$C_X \dot{V}_X(\theta, t) = g_l^X (V_X(\theta, t) - v_l^X) + I_{rec}^{EX}(\theta, t) - I_{rec}^{IX}(\theta, t) + I_{bg}(\theta, t) + I_{stim}(\theta, t) \tag{B.1}$$

where C_X is the membrane capacitance, g_l^X and v_l^X are the leak conductance and the membrane reversal potential, respectively. I_{rec}^{YX} is the recurrent input from presynaptic neurons in population Y to a postsynaptic neuron in population X where $X, Y \in \{E, I\}$. I_{bg} is a background input representing synaptic noise and I_{stim} is a stimulus dependent input. Whenever the voltage membrane V_X reaches the firing threshold v_t^X a spike is emitted and the membrane potential is reset to v_r^X, without a refractory period. The parameters in the model are modified depending on the TF model chosen, all these parameters are described in *section B.6.*

B.2 Intrinsic Neuron Properties

For the model in Section 3.2.4 we choose the neurons properties to obtain the desired TF form and thus, without any biological criteria. For model in Section 3.2.6 and Section 4.2.2 membrane parameters are similar as in McCormick et al. (1985), additionally we make sure all the values we use lie on the biological range by checking electrophysiology databases such as NeuroElectro.org (Tripathy and Gerkin, 2012). Reset potential after the spike for the excitatory is often taken to be more negative ($\approx -90\,\text{mV}$) than the inhibitory ($\approx -60\,\text{mV}$). Since the excitatory membrane time constant ($\tau_E = 20\,\text{ms}$) is typically twice the inhibitory membrane time constant ($\tau_I = 10\,\text{ms}$), the more negative reset potential will introduce an effective refractory period accounting for the fact that excitatory neurons usually display spike-frequency adaptation.

B.3 Recurrent Synaptic Inputs

B.3.1 Recurrent Synaptic Input Equations

When a spike occurs at time t_{sp} in a presynaptic neuron with PD ξ in population Y, the current evoked in the postsynaptic neuron with PD θ in population X, also known as postsynaptic current (PSC) is:

$$I_{sp}^{Y}(\theta - \xi, t - t_{sp}) = J_Y(\theta - \xi)\, s_Y(t - t_{sp}) \tag{B.2}$$

where $J_Y(\theta - \xi)$ is the total charge transferred in the synapse due to a single presynaptic spike, and it represents the synaptic strength (units of charge). This strength is scaled inversely with the size of the presynaptic population. As in the rate model, the connectivity is taken to have a cosine shape and to be independent on the nature of the postsynaptic population X (i.e. $J_{EE} = J_{EI} = J_E$ and $J_{II} = J_{IE} = J_I$):

$$J_Y(\theta - \xi) = \frac{1}{N_Y}\left(J_0^Y + J_1^Y\, \cos(\theta - \xi)\right) \tag{B.3}$$

On the other hand, the quantity $s_Y(t-t_{sp})$ describes the time course of the PSC in the postsynaptic neuron elicited by a spike at time t_{sp} from a presynaptic neuron in population Y. It has units of $[1/s]$ and it is described by a dual exponential waveform describing the fast dynamics at the opening and the slow dynamics at the closing of the synaptic receptors (Sterratt et al., 2011):

$$s_Y(t-t_{sp}) = \frac{1}{\tau_d^Y - \tau_r^Y}\left[\exp\left(\frac{-(t-t_{sp})}{\tau_d^Y}\right) - \exp\left(\frac{-(t-t_{sp})}{\tau_r^Y}\right)\right] \tag{B.4}$$

Where τ_r^Y and τ_d^Y are the synaptic rise and decay time constants, respectively. The normalising prefactor ensures that the total area under the PSC curve s_Y generated by a single spike is equal to 1. The recurrent input from the presynaptic population Y to the neuron with PD θ in the postsynaptic population X is defined as the sum over all the currents evoked by all the presynaptic neurons:

$$I_{rec}^{YX}(\theta,t) = \sum_i \sum_j I_{sp}^Y(\theta-\xi_i, t-t_j(\xi_i)) \tag{B.5}$$

where $t_j(\xi_i)$ is the time of the j-th spike fired by the presynaptic neuron with PD ξ_i.

B.3.2 Fast Synaptic Input Computation

As we can see, to calculate the synaptic current with Equation (B.4) we need to store all previous spike times and then recalculate the exponentials in the summation in Equation (B.5) each time step. This drawback can be overcome by using the method proposed in Srinivasan and Chiel (1993). This method for computing synaptic conductances separates Equation (B.4) in two components: one that is function of the current time of the simulation and one that accumulates the contributions of previous spike events to the synaptic conductance or current (in our case). This methods allows to only store two running sums and the time constants for each synapse.

Hereunder we detail the implementation in our specific case. Taken Equation (B.2) we can rewrite Equation (B.5) as follows:

$$I_{rec}^{YX}(\theta,t) = \sum_{i}^{N_Y}\left[J_Y(\theta-\xi_i)\sum_j s_Y(t-t_j(\xi_i))\right] \tag{B.6}$$

On one hand the connectivity J_Y can be rewritten as:

$$J_Y(\theta-\xi) = \frac{1}{N_Y}\left[J_0^Y + J_1^Y\left(\cos(\theta)\cos(\xi)+\sin(\theta)\sin(\xi)\right)\right] \tag{B.7}$$

And on the other hand, from Equation (B.6) we can define a new variable $S_{Y,i}(t)$ defined as:

$$S_{Y,i}(t) = \sum_{j}^{k} s_Y(t-t_j(\xi_i)) \tag{B.8}$$

Where k is the total number of spikes emitted by presynaptic neuron i until time t. For language economy, since now on $t_j(\xi_i) \equiv t_j$ since the variable $S_{Y,i}(t)$ already keeps track of the i label related to the neurons in the presynaptic population Y. Using Equation (B.4) we rewrite Equation (B.8) as:

$$S_{Y,i}(t) = \frac{1}{\tau_d^Y - \tau_r^Y}\left[\sum_j^k e^{-(t-t_j)/\tau_d^Y} - \sum_j^k e^{-(t-t_j)/\tau_r^Y}\right] \tag{B.9}$$

The two terms inside the brackets can be treated equally, here we take a general form to rewrite them as:

$$e^{-t/\tau}\sum_j^k e^{t_j/\tau} = e^{-(t-t_k)/\tau}\sum_j^k e^{-(t_k-t_j)/\tau} \tag{B.10}$$

We refer to the terms within the summation as $\text{Sum}(t_k)$. When the spike $k+1$ occurs at time t_{k+1} Equation (B.10) is updated as follows:

$$e^{-(t-t_{k+1})/\tau}\sum_j^{k+1} e^{-(t_{k+1}-t_j)/\tau} = e^{-(t-t_{k+1})/\tau}\left[e^{-(t_{k+1}-t_k)/\tau}\sum_j^k e^{-(t_k-t_j)/\tau} + 1\right] \tag{B.11}$$

Notice that we force to appear the $\text{Sum}(t_k)$ by multiplying the terms within the summation by $e^{(t_k - t_k)/\tau}$ $(= 1)$. Thus, by combining Equations (B.10) and (B.11) we can write an updating rule for the spike $k+1$ without the need of storing previous spikes:

$$\text{Sum}\,(t_{k+1}) = e^{-(t_{k+1}-t_k)/\tau} \cdot \text{Sum}\,(t_k) + 1 \tag{B.12}$$

At time $t > t_{k+1}$ the Equation (B.9) can be expressed as:

$$S_{Y,i}(t > t_{k+1}) = \frac{1}{\tau_d^Y - \tau_r^Y}\left[e^{-(t-t_{k+1})/\tau_d^Y}\,\text{Sum}_d(t_{k+1}) - e^{-(t-t_{k+1})/\tau_r^Y}\,\text{Sum}_r(t_{k+1})\right] \tag{B.13}$$

where $\text{Sum}_d(t_{k+1})$ and $\text{Sum}_r(t_{k+1})$ are the sums at time t_{k+1} for the decay and raise components, respectively. The quantity $S_{Y,i}(t)$ needs to be computed each time step of the simulation (i.e. from time t to time $t + \Delta t$). Thus, $S_{Y,i}(t + \Delta t)$ can be obtained by multiplying the terms $e^{-(t-t_{k+1})/\tau}$, in Equation (B.13) at time t, by $e^{-\Delta t/\tau}$. This yields the terms $e^{-[(t+\Delta t)-t_{k+1}]/\tau}$ at time $t + \Delta t$ and thus we can update the sums as follows:

$$\text{Sum}\,(t + \Delta t) = e^{-\Delta t/\tau} \cdot \text{Sum}\,(t) + \text{spk}(\xi_i, t) \tag{B.14}$$

where spk=1 if a spike is emitted by presynaptic neuron i at time t and spk=0 otherwise. The quantity $S_{Y,i}$ for each presynaptic neuron i can be calculated at time $t + \Delta t$ as follows:

$$S_{Y,i}(t + \Delta t) = \frac{1}{\tau_d^Y - \tau_r^Y}\left[\text{Sum}_d(t + \Delta t) - \text{Sum}_r(t + \Delta t)\right] \tag{B.15}$$

At this point we can rewrite Equation (B.6) at time $t + \Delta t$ using Equations (B.7) and (B.15):

$$I_{rec}^{YX}\,(\theta, t{+}\Delta t) = J_0^Y\;\tilde{S}_Y^0(t{+}\Delta t){+}(J_1^Y\;\tilde{S}_Y^1(t{+}\Delta t))\;\cos(\theta){+}(J_1^Y\;\tilde{S}_Y^2(t{+}\Delta t))\;\sin(\theta)) \tag{B.16}$$

where $\tilde{S}_Y^l$ are the Fourier components or order parameters of the synaptic recurrent input:

$$\tilde{S}_Y^0(t+\Delta t) = \frac{1}{N_Y}\sum_i^{N_Y} S_{Y,i}(t+\Delta t) \tag{B.17}$$

$$\tilde{S}_Y^1(t+\Delta t) = \frac{1}{N_Y}\sum_i^{N_Y} S_{Y,i}(t+\Delta t)\ \cos(\xi_i) \tag{B.18}$$

$$\tilde{S}_Y^2(t+\Delta t) = \frac{1}{N_Y}\sum_i^{N_Y} S_{Y,i}(t+\Delta t)\ \sin(\xi_i) \tag{B.19}$$

Now, substituting Equation (B.15) in the order parameters and taking into account Equation (B.14) we obtain:

$$\tilde{S}_Y^0(t+\Delta t) = \frac{1}{N_Y}\cdot\frac{1}{\tau_d^Y-\tau_r^Y}\left[\left(e^{-\Delta t/\tau_d^Y}\,\mathrm{Sum}_d(t) + \sum_i^{N_Y}\mathrm{spk}(\xi_i,t)\right) - \left(e^{-\Delta t/\tau_r^Y}\,\mathrm{Sum}_r(t) + \sum_i^{N_Y}\mathrm{spk}(\xi_i,t)\right)\right] \tag{B.20}$$

$$\tilde{S}_Y^1(t+\Delta t) = \frac{1}{N_Y}\cdot\frac{1}{\tau_d^Y-\tau_r^Y}\left[\left(e^{-\Delta t/\tau_d^Y}\,\mathrm{Sum}_d(t) + \sum_i^{N_Y}\mathrm{spk}(\xi_i,t)\ \cos(\xi_i)\right) - \left(e^{-\Delta t/\tau_r^Y}\,\mathrm{Sum}_r(t) + \sum_i^{N_Y}\mathrm{spk}(\xi_i,t)\ \cos(\xi_i)\right)\right] \tag{B.21}$$

$$\tilde{S}_Y^2(t+\Delta t) = \frac{1}{N_Y}\cdot\frac{1}{\tau_d^Y-\tau_r^Y}\left[\left(e^{-\Delta t/\tau_d^Y}\,\mathrm{Sum}_d(t) + \sum_i^{N_Y}\mathrm{spk}(\xi_i,t)\ \sin(\xi_i)\right) - \left(e^{-\Delta t/\tau_r^Y}\,\mathrm{Sum}_r(t) + \sum_i^{N_Y}\mathrm{spk}(\xi_i,t)\ \sin(\xi_i)\right)\right] \tag{B.22}$$

With Equation (B.16) and Equations (B.20) to (B.22) we can compute the recurrent synaptic input to a given postsynaptic neuron with PD θ at time $t+\Delta t$ by only storing the label of the neurons which have spiked at the current time t and the value of the expression inside the parenthesis in Equations (B.20) to (B.22) at time t.

B.3.3 NMDA and AMPA currents

In a more biologically realistic implementation of the synaptic currents we consider that excitatory synaptic inputs are a combination of fast AMPA and slow NMDA currents (Section 3.2.6 and Section 4.2.2). By using biological parameters for the rise and decay time constants in Equation (B.4) we describe the time course of the AMPA and NMDA EPSC separately. These dynamics represents the ion flow occurring in the synapse during the opening and closing of AMPAR and NMDAR. For the AMPA EPSC we chose a rise time constant (τ_r) of 0.5 ms and a decay time constant (τ_d) of 5 ms (Kleppe and Robinson, 1999; Gonzalez-Burgos et al., 2008). For the NMDA EPSC we chose $\tau_r = 1$ ms and $\tau_d = 50$ ms (Jonas et al., 1993; Gonzalez-Burgos et al., 2008). We found in the literature that in the excitatory to excitatory synapse ($E - E$) the mean NMDA/AMPA ratio calculated using the area under the EPSC in dlPFC neurons in monkey is 2.7 (Gonzalez-Burgos et al., 2008). On the other hand, the NMDA/AMPA ratio in the excitatory to inhibitory synapse ($E - I$) is 0.5 (Wang and Gao, 2009). This value is computed by taking the NMDA and AMPA charge (EPSC area) of the fast spiking interneurons and taking into account that, in the adult rat, only the 26 $\%$ of this interneurons have NMDA currents. To calculate the resultant synaptic current we proceed as follows:

$$I_{rec}^{EX}(\theta, t) = \frac{r^{EX}}{r^{EX}+1} \cdot I_{rec}^{NMDA}(\theta, t) + \frac{1}{r^{EX}+1} \cdot I_{rec}^{AMPA}(\theta, t) \tag{B.23}$$

where r^{EX} is the NMDA/AMPA ratio for the $E - X$ synapses.

B.4 Background Noise Input

Each neuron receives a noisy background input of the form:

$$I_{bg}^{X}(\theta, t) = \mu_{bg}^{X} + \sigma_{bg}^{X} \cdot \xi(t) \tag{B.24}$$

where $\xi(t)$ is a Gaussian white noise (GWM) with zero correlation in time and between neurons. This input is mean to describe the continuous bombardment of synaptic activity. The addition of noise in the model originates an expansive

nonlinearity in the neuron TF at the firing threshold as described in Hansel and van Vreeswijk (2002); Miller and Troyer (2002). To implement the difference in threshold mechanism in the spiking model, we chose the mean background input of the excitatory to be slightly larger than the inhibitory so the rheobase current will be lower for the excitatory population creating a difference in the spiking threshold of the excitatory respect to inhibitory neurons.

B.5 External Stimulus

The tuned external external stimulus is described by a cosine function which maximum is centered at the direction of the cue θ_{cue}:

$$I_{stim}(\theta, t) = c[1 - \epsilon + \epsilon \cos(\theta - \theta_{cue})] \tag{B.25}$$

Where c is the contrast of the stimulus and ϵ controls the angular anisotropy of the input (as in Ben-Yishai et al. (1995)). Notice that when $\theta = \theta_{cue}$ the input to neuron with PD θ is maximum.

B.6 Spiking Model Parameters

In this section we detail the parameters used in the different spiking network models of leaky integrate-and-fire neurons that have appear throughout the thesis in Chapters 3 and 4.

B.6.1 Acceleration Difference Model

For the spiking model presented in Section 3.2.4 we used the parameters detailed in Tables B.1 and B.2.

Parameters	Excitatory	Inhibitory	Units
N_X	12000	12000	-
v_l^X	-70	-70	mV
g_l^X	50	50	nS
C_X	2.0	0.5	nF
v_t^X	-30	-30	mV
v_r^X	-60	-55	mV
μ_{bg}^X	250	0	pA
σ_{bg}^X	20	20	pA

Table B.1 – Acceleration Difference Integrate-and-Fire model: neuron parameters

Parameters	E Presynaptic	I Presynaptic	Units
τ_r^E	1	-	ms
τ_d^E	40	-	ms
$g\tau_r^I$	-	1	ms
τ_d^I	-	5	ms
J_0^{YX}	40	36	nC
J_1^{YX}	120	20	nC

Table B.2 – Acceleration Difference Integrate-and-Fire model: synaptic parameters

For the model presented in Section 3.2.6 the parameters are detailed in Tables B.3 and B.4:

Parameters	Excitatory	Inhibitory	Units
N_X	12000	12000	-
v_l^X	-70	-70	mV
g_l^X	30	20	nS
C_X	0.6	0.2	nF
v_t^X	-50	-50	mV
v_r^X	-90	-60	mV
μ_{bg}^X	320	50	pA
σ_{bg}^X	9	9	pA

Table B.3 – Acceleration Difference Integrate-and-Fire model: biological neuron parameters.

Parameters	E Presynaptic	I Presynaptic	Units
τ_r^{AMPA}	0.5	-	ms
τ_d^{AMPA}	5	-	ms
$g\tau_r^{NMDA}$	1	-	ms
τ_d^{NMDA}	50	-	ms
$\tau_r^{GABA_A}$	-	1	ms
$\tau_d^{GABA_A}$	-	5	ms
J_0^{YX}	20	18	nC
J_1^{YX}	60	12.5	nC

Table B.4 – Acceleration Difference Integrate-and-Fire model: biological synaptic parameters.

B.6.2 Threshold Difference Model

In Section 4.2.1 we used a threshold difference model with parameters defined in the following tables:

Parameters	Excitatory	Inhibitory	Units
N_X	12000	12000	-
v_l^X	-70	-70	mV
g_l^X	30	20	nS
C_X	1.6	1.5	nF
v_t^X	-30	-30	mV
v_r^X	-60	-55	mV
μ_{bg}^X	200	0	pA
σ_{bg}^X	15	15	pA

Table B.5 – Threshold Difference Integrate-and-Fire model: neuron parameters.]

Parameters	E Presynaptic	I Presynaptic	Units
τ_r^E	1	-	ms
τ_d^E	40	-	ms
$g\tau_r^I$	-	1	ms
τ_d^I	-	5	ms
J_0^{YX}	40	36	nC
J_1^{YX}	100	25	nC

Table B.6 – Threshold Difference Integrate-and-Fire model: synaptic parameters.

C Supplementary Figures

In this appendix we show supplementary figures related to the works presented in Chapters 3 to 5.

C.1 Supplementary Figures from Chapter 3

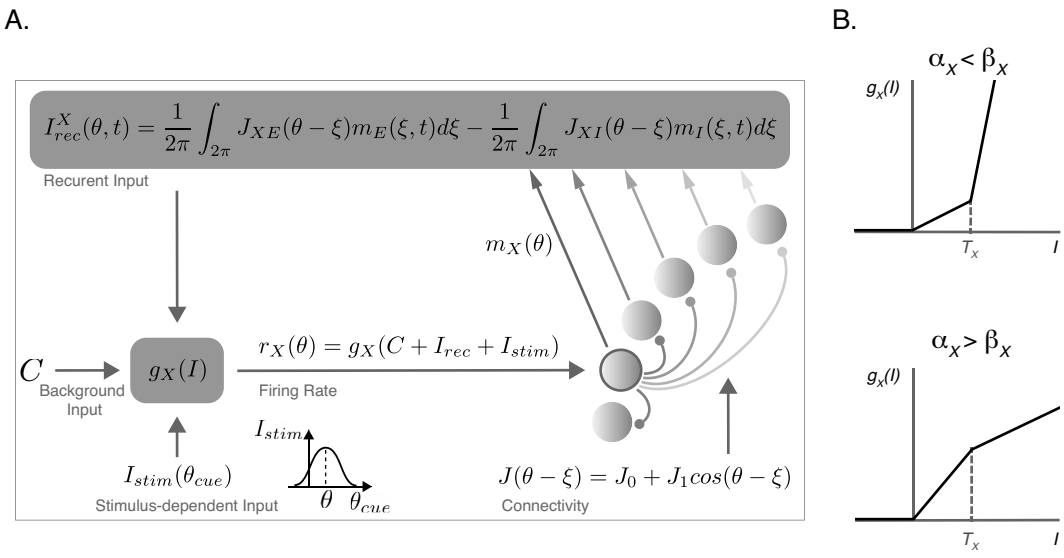

Figure C.1. Rate model diagram A. A diagram of the rate model. **B.** Transfer functions that are considered in this work. For $0 < I < T_X$ the slope of $g_X(I)$ is α_X, and for $T_X < I$ the slope is β_X. The $\alpha_X < \beta_X$ case is termed *expansive* nonlinearity (top) and the $\alpha_X > \beta_X$ is termed *compressive* nonlinearity (bottom).

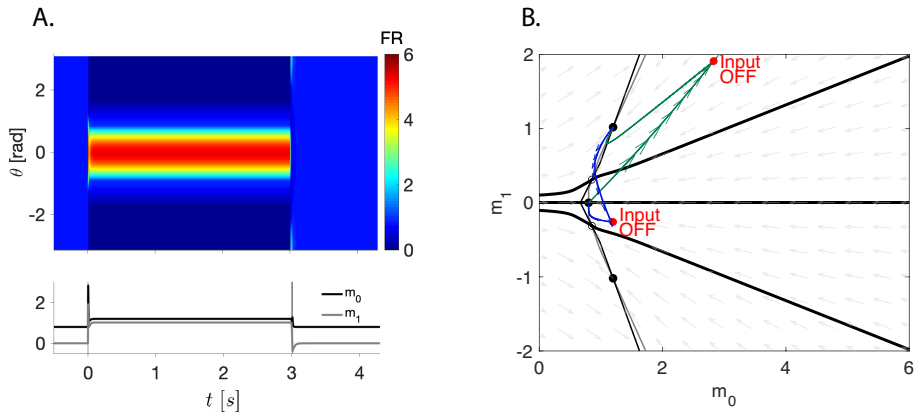

Figure C.2. Expansive nonlinearity, switch of with anti-phase excitation. A and B. Same as in figure 3.2**A** and **B**, except that the switch off input is a 12 ms tuned excitatory pulse, centered at $\theta = \pi$.

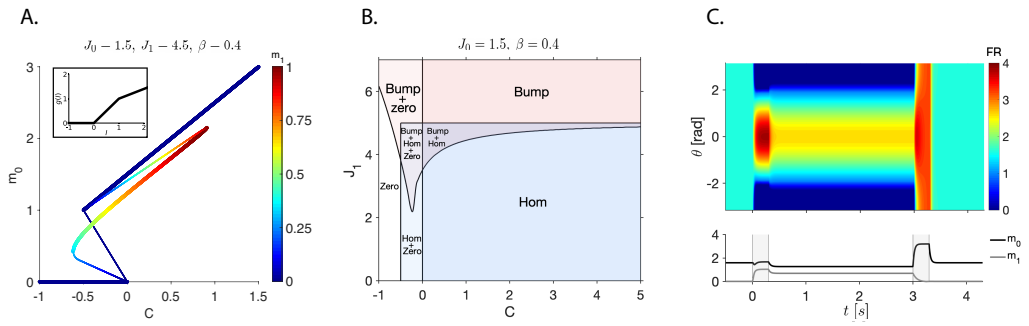

Figure C.3. Compressive nonlinearity model ($\beta < 1$) for $J_0 > 1$. A-C: $\beta = 0.4$. **A**. graphics are as in figure 3.1 with and $J_0 = 1.5$, $J_1 = 4.5$. In the case $J_0 < 1$ (figure 3.1 **D**) the unstable homogeneous state is only *spatially* unstable while in the case $J_0 > 1$ the unstable homogeneous state is both spatially- and rate-unstable. **B**. Phase diagram in the $J_1 - C$ plane for $J_0 = 1.5$. Note the existence of a regime where two stable homogeneous states exist. In the low homogeneous state m_0 is zero. **C**. Dynamics of the network and of the order parameters in response to a 0.3 s tuned input at $t = 0$ s centered at $\theta = 0$, and a 0.3 s homogeneous excitatory pulse at $t = 3$ s. $C = 0.1$. Here, the switch-off is possible with a prolonged homogeneous pulse.

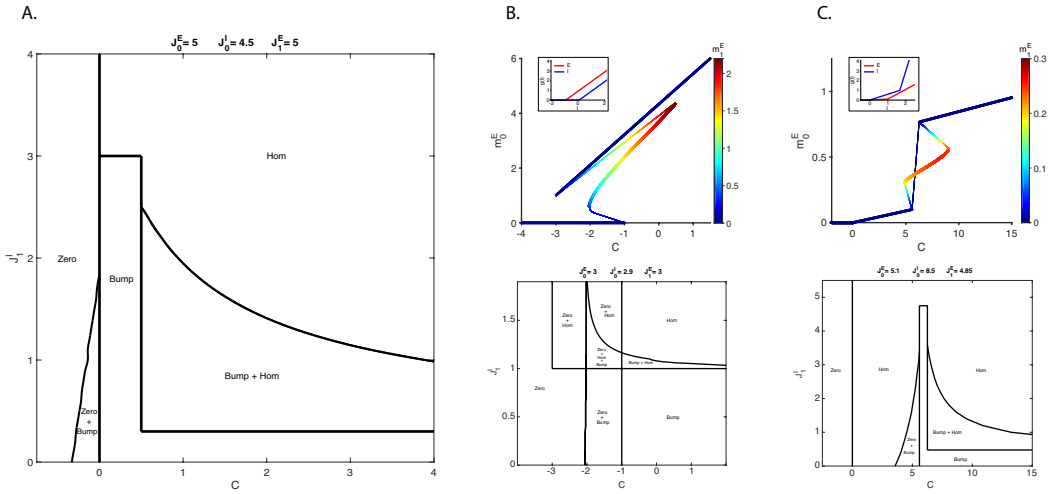

Figure C.4. Two-population mechanisms. **A.** Phase diagram in the $J_1 - C$ plane for the acceleration difference model. **B** and **C**. Bifurcation diagrams of $m_0^E - C$ (top) and phase diagrams in the $J_1 - C$ plane (bottom) in the 2-population network for the threshold difference and a combination of acceleration and threshold difference model, respectively. Color code: m_1^E. Insets: the TFs of the excitatory (red) and the inhibitory (blue) neurons. Parameters in **B**: $J_E^0 = 3$, $J_I^0 = 2.9$, $J_E^1 = 3$, $J_I^1 = 1.07$. Parameters in **C**: $J_E^0 = 5.1$, $J_I^0 = 8.5$, $J_E^1 = 4.85$, $J_I^1 = 1.5$, $\alpha_E = 0.1$, $\beta_E = 1$, $\alpha_I = 0.6$, $\beta_I = 6$, $T_E = 1$, $T_I = 5/3$.

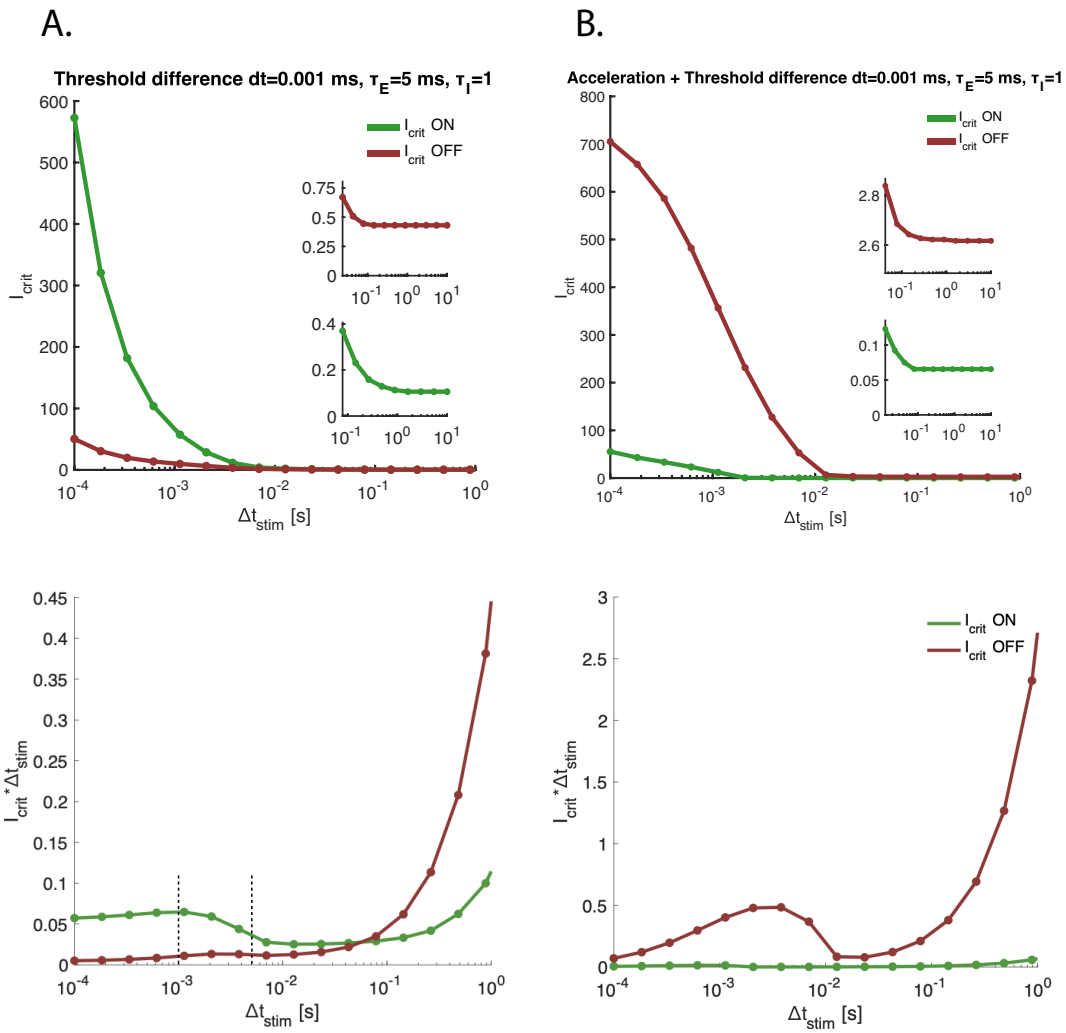

Figure C.5. Switch-off critial input in the two-population. A. Minimal input current (top) and minimal input charge ($I_{crit} \cdot \Delta t_{stim}$, bottom) needed to perform a state transition plotted against the duration of that input, Δt for the threshold difference model. Homogeneous to bump state transition (H-B) in green and bump to homogeneous state transition (B-H) in red. Inset: zoom for durations above $0.1\ s$. **B.** Same for the acceleration and threshold difference model.

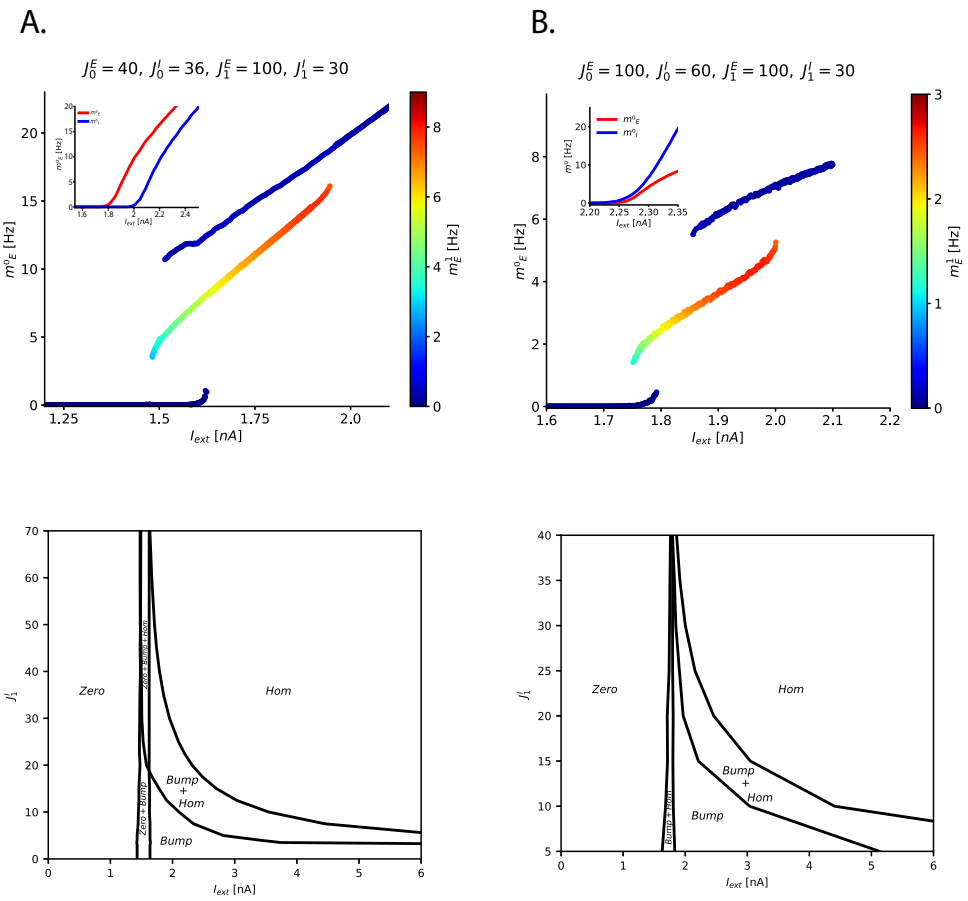

Figure C.6. Two-population mechanisms in the LIF model. Bifurcation diagram of m_0^E-C (top) and $J_1^I - C$ phase diagrams (bottom) in the two population spiking network in the threshold difference (**A**) and acceleration and threshold difference case (**B**), respectively. Color code: first Fourier component of the activity, m_1^E. Inset: the TFs of the excitatory (red) and the inhibitory (blue) neurons.

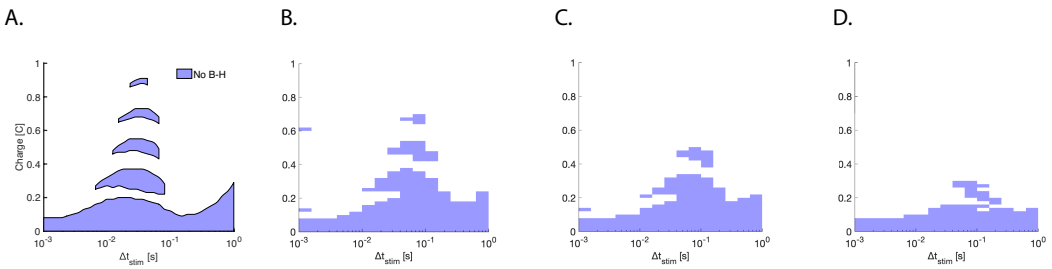

Figure C.7. Minimal input charge for the switch-off transition in the LIF model at different synaptic time constants. Input charge for the bump to homogeneous transition as a function of stimulus duration (Δt_{stim}). Colored regions depicts the values at which the transition cannot be performed. In **A** $\tau_d^E = 40\,ms$, $\tau_d^I = 5\,ms$; **B** $\tau_d^E = 80\,ms$, $\tau_d^I = 10\,ms$; **C** $\tau_d^E = 120\,ms$, $\tau_d^I = 15\,ms$; **D** $\tau_d^E = 160\,ms$, $\tau_d^I = 20\,ms$.

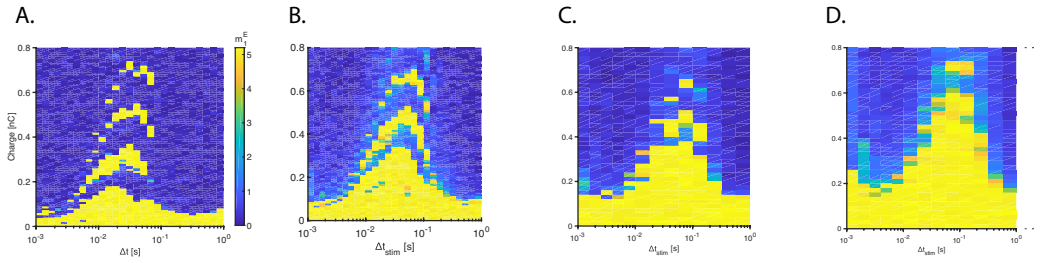

Figure C.8. Minimal input charge for the switch-off transition in the LIF model with biological parameters at different synaptic time constants. Input charge for the bump to homogeneous transition as a function of stimulus duration (Δt_{stim}). Color coded the value of m_1^E at the end of the switch-off transition. In **A** $\tau_d^{NMDA} = 50\,ms$, $\tau_d^{GABA} = 5\,ms$; **B** $\tau_d^{NMDA} = 100\,ms$, $\tau_d^{GABA} = 10\,ms$; **C** $\tau_d^{NMDA} = 200\,ms$, $\tau_d^{GABA} = 20\,ms$; **D** $\tau_d^{NMDA} = 300\,ms$, $\tau_d^{GABA} = 30\,ms$.