



**Etude de la balance Excitation/Inhibition de régions cérébrales  
impliquées dans une tâche de contrôle inhibiteur :  
modélisation de données obtenues en Imagerie par Résonance  
Magnétique fonctionnelle et inversion**

Anais Vallet

► **To cite this version:**

Anais Vallet. Etude de la balance Excitation/Inhibition de régions cérébrales impliquées dans une tâche de contrôle inhibiteur : modélisation de données obtenues en Imagerie par Résonance Magnétique fonctionnelle et inversion. Psychologie. Normandie Université, 2024. Français. ⟨NNT : 2024NORMC014⟩. ⟨tel-05051939⟩

**HAL Id: tel-05051939**

**<https://theses.hal.science/tel-05051939v1>**

Submitted on 30 Apr 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# THÈSE

Pour obtenir le diplôme de doctorat

Spécialité **PSYCHOLOGIE**

Préparée au sein de l'**Université de Caen Normandie**

**Etude de la balance Excitation/Inhibition de régions cérébrales impliquées dans une tâche de contrôle inhibiteur : modélisation de données obtenues en Imagerie par Résonance Magnétique fonctionnelle et inversion**

Présentée et soutenue par  
**ANAIS VALLET**

**Thèse soutenue le 17/10/2024**  
devant le jury composé de :

|                        |                                                                        |                       |
|------------------------|------------------------------------------------------------------------|-----------------------|
| M. FRANCIS EUSTACHE    | Directeur d'Etudes EPHE - Université de Caen Normandie (UCN)           | Directeur de thèse    |
| M. PIERRE GAGNEPAIN    | Chargé de recherche Inserm HDR - GIP CYCERON Caen                      | Co-directeur de thèse |
| MME BERENGERE GUILLERY | Directeur d'Etudes EPHE - Université de Caen Normandie (UCN)           | Président du jury     |
| M. JACQUES GAUTRAIS    | Chargé de recherche au CNRS - Université de Toulouse 3 - Paul Sabatier | Membre du jury        |
| MME SOPHIE ACHARD      | Directeur de recherche au CNRS - Université Grenoble Alpes             | Rapporteur du jury    |
| M. JEAN DAUNIZEAU      | Chargé de recherche Inserm HDR - Institut du Cerveau                   | Rapporteur du jury    |

Thèse dirigée par **FRANCIS EUSTACHE** (Neuropsychologie et Imagerie de la Mémoire Humaine (NIMH)) et **PIERRE GAGNEPAIN** (Neuropsychologie et Imagerie de la Mémoire Humaine (NIMH))



"Ce que l'on conçoit bien s'énonce clairement", Nicolas Boileau





# Remerciements

**Mme Sophie Achard** et **M Jean Daunizeau**, merci d’avoir accepté de lire et de rapporter ce travail. Je suis reconnaissante du temps et de l’énergie que vous avez accordés à l’étude du manuscrit de ma thèse. **M Jacques Gautrais** et **Mme Bérengère Guillery**, merci d’avoir accepté de l’examiner le jour de ma soutenance. Je vous remercie tous les quatre de me donner l’opportunité de présenter et de discuter de ma thèse. La diversité de vos disciplines d’origine permet de donner une couleur interdisciplinaire au jury (entre mathématiques, physique, biologie et neuropsychologie), qui a beaucoup de sens pour moi étant donné que j’ai eu à cœur de maintenir présente cette dimension de mon travail tout au long de ces dernières années, et que j’ai essayé d’y accorder un soin particulier pendant la rédaction de ma thèse. J’espère que ma soutenance sera l’occasion d’un moment de dialogue interdisciplinaire et je suis reconnaissante que nous ayons collectivement permis que de tels échanges puissent se produire. Ils seront pour moi d’une grande richesse.

Merci à mes directeurs de thèse, **M Francis Eustache** et **M Pierre Gagnepain**, d’avoir accepté la prise de risque que représente d’engager une physicienne dans un laboratoire de neuropsychologie et de neuroimagerie. Cette immersion a été d’une extrême richesse pour ma formation au travail interdisciplinaire, notamment en ce qu’elle m’a permis d’apprendre à communiquer avec des collègues d’une autre discipline que la mienne. J’ai adoré travailler sur une problématique de modélisation biophysique appliquée à une question de neurosciences et cela a constitué pour moi un levier de motivation intrinsèque important qui m’a aidée durant tout mon parcours doctoral. Merci à tous les deux d’avoir accepté une collaboration pendant la seconde partie de mon doctorat, de sorte que notre projet commun ait pu bénéficier d’expertises en physique du non linéaire et en inversion, en plus de vos expertises en neuropsychologie, neurosciences et neuroimagerie. Ma thèse a pu grandir de ses apports multiples et j’en suis très reconnaissante.

**Francis**, ça a été un plaisir de pouvoir vous rencontrer dans le travail pendant cette dernière année. Je suis contente d’avoir eu la chance de quelques échanges formateurs qui m’ont permis de préciser ma pensée sur des notions de neuropsychologie, notamment en lien avec les intrusions mnésiques. Merci de vos encouragements répétés durant toute la dernière ligne droite de rédaction, de vos annotations manuscrites sur les pages de ma thèse et de vos lettres pour m’adresser votre travail. J’ai beaucoup apprécié ce retour au papier, colorant ces échanges d’une dimension plus personnelle et chaleureuse.

**Pierre**, merci d’avoir orienté le travail vers le choix d’un modèle non linéaire, un corpus de la physique qui m’a passionnée pendant mes études et auquel je pensais avoir dû dire au revoir en m’orientant en master 2 de neuropsychologie et neurosciences cliniques. Quelle joie que ce n’est finalement pas été le cas. Merci également des discussions qui ont aidé à préciser le plan de ce manuscrit.

**Mme Jessica Mange** et **M Stéphane Blanco**, je tiens à vous remercier pour votre accompagnement et vos conseils durant ces cinq années en tant que comité de suivi.

**M Madeline**, je suis également reconnaissante de votre compréhension et votre réactivité.

**Bérengère**, merci de ton implication dans les derniers mois de mon doctorat, de ta bien-

veillance et tes conseils. Ta présence m'a aidée à être davantage disponible au travail de rédaction du manuscrit. **Annick**, merci d'avoir été si à l'écoute, empathique et disponible quand j'en avais besoin.

La construction de cette thèse a été avant toute chose une aventure humaine, pleine de partage. Elle n'aurait pu être menée sans l'entourage de toutes les personnes qui m'ont accompagnée, soutenue et formée. Un énorme merci !

**Jacques, Jean-Louis, Jean-Yves et Stéphane**, merci infiniment du cadeau de la formation, tellement multidimensionnelle, aussi bien sur le plan de l'analyse non linéaire que sur celui des pratiques de recherche. C'est un sac-à-dos avec pleins d'outils que j'embarque avec moi et que je peux partager. Je suis très reconnaissante d'avoir eu la chance de bénéficier de cet héritage. Merci également de vos retours positifs, qui m'ont encouragée et motivée. Cela me semble tellement important quand on apprend à avoir confiance en son travail en début de carrière.

**Jacques**, nous nous sommes rencontrées en 2018 quand j'ai fait mon stage de premier master 2 (physique pour le vivant), avec vous et en partie avec Manon Azaïs. J'ai pu être témoin d'une pratique de l'interdisciplinarité qui m'a guidée tout au long de ma thèse. Je ne peux que vous être reconnaissante de cet exemple, ainsi qu'à Manon. En 2019, à la fin de mon stage de second master 2 avec Pierre, merci de m'avoir rappelé à quel point ce sujet de thèse semblait concorder avec mes aspirations à travailler à l'interface entre la physique et les neurosciences, tout particulièrement dans un programme de recherche sur le trouble de stress post-traumatique. Merci aussi de m'avoir rassurée sur le fait que je pourrais compter sur votre collaboration et ne pas être la seule physicienne impliquée dans ce projet. Nous étions sous un arbre devant le bâtiment du Laplace et je me souviens encore du moment où mes doutes se sont dissipés et où je savais que l'aventure de ce doctorat allait commencer. Je tiens à vous remercier de m'avoir aidée à prendre la décision qui m'a permis d'en apprendre tellement en cinq ans. Votre calme communicatif et vos capacités d'analyse m'ont été d'une aide précieuse durant ces dernières années et je vous en suis très reconnaissante. Merci de comprendre ce que je ne comprends pas depuis là où j'ai été formée, vos traductions entre Pierre et moi nous ont souvent été très utiles. Merci de savoir saisir si rapidement mon méta-discours, c'est à la fois très confortable, libérateur et rassurant, et ça l'a été tout particulièrement au moment où j'ai proposé mon projet de rédaction du chapitre 13. Votre proposition de chapeaux en début de chapitres a permis de faciliter l'écriture et j'espère la lecture du manuscrit. Merci d'avoir travaillé avec moi à préciser l'intention dans le chapeau du 13 notamment, ça a été très riche pour ma formation.

**Stéphane**, merci d'être passionné et de rendre le travail avec toi passionnant. Cela a été un plaisir de pouvoir travailler ponctuellement avec toi, c'était toujours très riche. Merci de ta rigueur inspirante, de ton éloge de la lenteur qui assurément guide ma manière de travailler pas à pas, de m'avoir répété "qu'elle est la question" pour que je puisse intégrer cette interrogation et m'en servir comme une boussole pour guider le travail de précision de ma question de recherche. Je suis très reconnaissante des conseils que tu m'as donnés, notamment quand je commençais à rédiger ma thèse. Merci de m'avoir expliqué l'importance de préciser une

problématique. C'est devenu un des axes de mon travail et ça a été une période éclairante et passionnante pour moi. En discutant avec toi, je me suis rendue compte qu'en quatrième année de doctorat, je ne savais pas ce que signifiait d'écrire une thèse et que je confondais la thèse avec le manuscrit. Je suis alors allée me documenter et ça a complètement impacté ma manière d'écrire, comme tu me l'avais expliqué. Toutes les lignes servent à préciser ou répondre à la problématique (i.e. la question ou la série de questions précises à laquelle on cherche à répondre). La thèse est devenue l'idée que je défends (la réponse que j'apporte à la problématique), elle n'est pas une synthèse de mes travaux, la synthèse doit être au service de l'argumentation. Le manuscrit n'est que le support écrit de la thèse, qui se défendait d'ailleurs uniquement à l'oral pendant le Moyen-Age.

**Jean-Louis**, nous nous sommes rencontrés lors de mon premier Roffiac en 2021, sous l'impulsion de Stéphane. Merci d'avoir toujours su écouter mes propositions et de m'avoir accompagnée dans le travail. Tu m'as encouragée, sans jamais m'empêcher de quoi que ce soit, en me mettant simplement en garde face au temps qui passe quand il le fallait et en me faisant confiance. Merci de m'avoir guidée avec autant de bienveillance et d'empathie.

Je me souviens du jour où je suis venue te voir avec des feuilles pleines de calculs qui dataient de quelques mois en arrière. Stéphane venait tout juste de nous faire un cours sur le découpage-raccordement (celui qu'on a décortiqué ensuite ensemble dans tous les sens), je sentais un lien avec cet ancien travail, sans vraiment savoir le préciser. Tu as pris le temps de me relire et tu m'as dit de ne pas lâcher cette idée, même si pour toi non plus le lien n'était pas clair pour le moment. La période de juillet/août 2022 qui a suivie a été pleine d'émulation et de partage entre nous. J'en garde un très bon souvenir, merci beaucoup. Je me souviens avoir passé des heures au tableau à faire des découpages dans tous les sens (ça en devient addictif...) jusqu'à enfin réussir à comprendre comment extraire les sensibilités que je voulais. Nous partagions nos avancements plusieurs fois par semaine, tes retours sur mon travail m'ont énormément aidée et tes vérifications numériques avec Miniker ont fini par nous soulager, nous avons pu partir tranquillement en vacances.

**Jean-Yves**, merci de nous avoir rejoints à partir de Roffiac 2022. Tes questions et tes retours très précis nous ont toujours permis de travailler à préciser la pensée (je pense par exemple à la différence entre sensibilité isolée vs découplée). Merci de m'avoir signalé quand j'étais trop dans la couche mathématique et quand il était important que je revienne dans le modèle, maintenant j'en comprends beaucoup mieux le sens.

Nos séances de travail en visio du mercredi matin avec toi et Jean-Louis m'ont énormément permis de travailler la matière de l'article. Je vous suis très reconnaissante de ces moments de partage et de travail collectif pendant lesquels vous avez toujours accepté que je vous présente l'état de mon travail (avec des photos de tableaux, de feuilles et quelques sorties de codes). Je mesure la chance que j'ai eu à être formée par tous les deux au découpage-raccordement et à l'emboîtement. Vous m'avez également enseigné l'importance du choix des notations. Je crois que c'est devenu un réflexe... en tout cas j'y veille vous pouvez me croire !

**Shailendra**, merci de tous les retours positifs que tu m'as fait et qui m'ont permis de prendre davantage confiance en mon travail. Merci de ta disponibilité à chaque fois que j'ai eu besoin de me mettre en situation de transmission pour préciser ma pensée (tu l'auras

entendu le mot "découplé" !). Quand je n'arrivais pas à t'expliquer quelque chose, je savais où travailler. Il y en a eu des séances de travail à ton bureau les vendredis soirs jusque tard... Ces moments ont été l'occasion pour nous d'apprendre à travailler ensemble depuis nos formations différentes. Enfin, je te suis reconnaissante d'avoir pris le temps de répondre à mes questions de neuroimagerie, jusqu'à minuit par téléphone à quelques jours du rendu ! Merci d'avoir fait partie de la team des relecteurs de la thèse qui se reconnaîtront... cette période de relecture à la chaîne aura été intense, pleine de partage et de bonne humeur. Je ne l'aurais pas aussi bien vécue sans vous, quelle chance j'ai eu d'avoir été si bien entourée jusqu'à la fin !

Je tiens évidemment à remercier le **collectif de Roffiac** pour tout l'accueil, la bienveillance, le partage, la formation et le sentiment d'appartenance au groupe. J'ai eu la chance de participer à trois sessions du séminaire (en septembre 2021, 2022 et 2024), j'ai pu y présenter à chaque fois ma thèse et y proposer des ateliers de travail. Merci beaucoup à toutes les personnes qui ont pris le temps de m'écouter et de travailler avec moi, vos questions et vos divers retours m'ont permis de préciser ma pensée. Merci également pour les retours positifs, qui encouragent, motivent et permettent d'apprendre à prendre confiance en son travail. Ma première "conférence" grand public du jeudi soir reste un moment unique de partage et de joyeuse effervescence : la séance de travail avec les habitants du village pour construire avec eux un modèle de la dynamique de population sur le plateau, avec en parallèle les ateliers de mise en équations et de codage. C'était drôle de voir le travail se faire dans ces trois couches en aller-retour (merci les messagers) ! Je suis également reconnaissante des jolis moments suspendus du jeudi soir, au rythme des guitares, des chants et de la danse, ainsi que de la grande diversité de sujets de conversation, à n'importe quelle heure du jour ou de la nuit. Merci enfin de votre écoute, soutien et de vos conseils, en particulier aux deux **Léa, Guillaume, Mégane, Frédéric, Nada, Thomas, Richard, Jacques, Stéphane, Jean-Louis, Lionel, Vincent, Mathieu et Najda**.

**Coline**, merci d'avoir accepté de travailler avec nous pour ton stage de master 2 qui a marqué le début officiel d'une collaboration entre nos deux laboratoires en février 2022. Je tiens à te remercier de tes mots réconfortants, de tes encouragements et de tous les retours positifs que tu m'as fait, y compris à ce dernier Roffiac. C'est gratifiant de savoir que tu t'es lancée dans une thèse à l'interface entre la physique et les neurosciences après avoir travaillé avec nous. Merci aussi de ton humour, j'adore l'idée des phrases "entendues à Roffiac". Je vais en écrire une version de ma thèse bientôt !

Je remercie les **personnes du programme 13 novembre** pour les échanges auxquels j'ai pu assister à l'occasion de quelques séminaires et qui ont été formateurs pour moi. Merci également de m'avoir écoutée et fait des retours à l'occasion de deux présentations. Pouvoir s'adresser à un auditorat aussi varié (historiens, sociologues, politologues, neuropsychologues, neuroscientifiques...) a constitué un travail de pédagogie très enrichissant, plaisant à mener et utile pour me permettre d'apprendre à situer ma question de recherche dans les travaux de la communauté neuroscientifiques. Merci particulièrement à **Carine Klein Peschanski** pour ta bienveillance, ton énergie et tes retours à l'école d'automne sur ma présentation. Même si nous ne nous sommes vues que quelques fois en cinq ans, cela a toujours été un plaisir d'échanger avec toi.

Les rencontres avec des étudiant.e.s stagiaires de licence et master, principalement de biologie ou psychologie, ont été autant d'opportunités d'expliquer ma thèse et d'affiner mon discours. **Laura Masson** et **Paola**, je vous remercie de votre enthousiasme, et de votre curiosité. Ca a été un plaisir de vous partager mon travail. **Margot** et **Caroline**, merci également pour les quelques ateliers d'algorithmique. Je me suis bien amusée à préparer les séances et à travailler avec vous.

Pendant mon doctorat, j'ai eu l'opportunité d'enseigner pendant 130h auprès d'étudiants en L1 de physique-chimie en parcours adapté, c'est-à-dire venant de formations très différentes (baccalauréats S, L mais aussi professionnels spécialisés dans la vente, le bâtiment ou l'esthétique par exemple). Je tiens à remercier **Michaël Fromager** de m'avoir fait confiance en me confiant les enseignements d'optique géométrique et de remise à niveau en dynamique newtonienne durant deux années consécutives. Merci de m'avoir laissée la liberté de préparer mes cours, tout en me confiant tes supports pour m'y aider, et de m'avoir même laissée préparer certains sujets d'examens. L'autonomie que tu m'as accordée a été très motivante pour moi et cela a été un plaisir de travailler avec toi, en se partageant les exercices et en se tenant au courant de nos avancées en classe.

Cette première expérience d'enseignement m'a beaucoup apporté en ce qu'elle m'a poussée à devoir travailler diverses dimensions de la pédagogie (la didactique, mais aussi les aspects communicationnels et organisationnels), peut être encore plus que si j'avais commencé avec un public apprenant aux niveaux et pré-requis plus homogènes. Les formations à la pédagogie mises en place par l'école doctorale m'ont été de ce point de vue très utiles. Je souhaite remercier tout particulièrement **Thierry Piot** pour le partage de sa passion pour la transmission et pour tous les repères et outils d'analyses dont j'ai pu bénéficier pour ne pas me sentir démunie devant les étudiant.e.s et être en mesure d'améliorer mes enseignements.

Evidemment, je ne peux pas m'empêcher de penser à toi **Cassandra**. Ca a été un plaisir de commencer à enseigner en même temps que toi et de pouvoir partager la préparation de nos cours (c'était chouette de se montrer nos diapos pendant la période covid des cours en visio), de pouvoir débriefer de nos séances et de se donner des conseils. Plus largement, merci de ton authenticité, de ta sensibilité et de ta générosité. On s'est soutenue, on a pleuré, on a ri et on s'en est sortie ! Cette thèse sans toi à proximité aurait été assurément moins drôle, je ne suis pas prête d'oublier le fameux drive de courgettes, carottes et poivrons ou encore ton inscription à la "danse en ligne" :p

**Najlaa**, quelle joie d'avoir appris récemment ta merveilleuse surprise ! Petite cachotière, je n'ai rien vu sur aucune photo de profil pour annoncer son arrivée ! Ca a été un plaisir d'apprendre à te connaître, je suis contente qu'on se soit soutenue ensemble et qu'on ait ri ensemble. Je me souviens de nos pauses marrantes au début de ma thèse, se terminant inévitablement par la question existentielle de Shail (Scratty) "qu'est-ce que je vais manger ce soir ?" et sa réponse "du poulet". Merci d'avoir été la super camerawoman qui a immortalisé mon cadeau d'anniversaire pour le soin de ses cheveux. Par contre, je ne sais pas si je dois te remercier pour avoir encouragé le surnom de Josie... mais c'est trop tard parce qu'après toutes ces années il est plus qu'adopté. A toi de voir au bout de combien de temps il y a

prescription pour les défis de goûteurs de fromage avec soda.

**David Corredor**, tu as été mon binôme de thèse. J'ai pris beaucoup de plaisir à prendre le temps de travailler devant un tableau avec toi. Merci de ce partage, de nos aller-retours constants entre les différents plans du discours pour savoir nous dire quand ça allait trop vite, quand on percevait chez l'autre un implicite et quand on avait besoin que l'autre précise sa pensée. Je te remercie aussi de toutes les nouvelles notions en neurosciences ou en épistémologie auxquelles tu m'as introduite. J'ai trouvé nos discussions sur la phénoménologie passionnantes ! Je garde l'image inspirante de ton bureau avec tous tes livres annotés dans tous les sens. Je te suis évidemment reconnaissante de ton soutien et de tes conseils. Merci de tes encouragements, de l'intérêt que tu as eu pour mon travail, de tes questions et de tes retours positifs.

Merci à **Mikaël Naveau** pour ton soutien, tes conseils, et pour avoir veillé à distance en demandant de mes nouvelles à Shail et David. Les séances de travail avec toi, David, Gina, Shail et Nicolas Delcroix ont été des moments de partage motivants et formatteurs. Merci à ce groupe de m'avoir permis de présenter mon travail. Vos retours ont notamment alimenté le chapitre 14 sur les limites du modèle hémodynamique.

**Anne-Lise** et **Géraldine** merci de votre soutien, d'avoir pris des nouvelles même quand j'étais au chômage et que je ne travaillais plus au labo. Anne-Lise, je tiens aussi à te remercier de tes conseils et paroles rassurantes pendant une grosse période de doutes.

Merci **Thibaut** de ta disponibilité et réactivité à chaque fois qu'il a été question de percevoir un salaire, c'est important. Ça a été un plaisir **Thomas Vallée** d'échanger avec toi et de t'aider à préparer, avec Margot et Caroline, quelques tâches de rappel d'images pour REMEMBER, c'était long mais marrant. **Antoine** et **Geoffrey** merci d'être toujours très rapidement disponibles quand on a besoin de vous pour l'informatique. Vous m'avez bien aidée plus d'une fois.

Je tiens à remercier les personnes de la cantine, et en particulier **Marina**, **Nina** et **Martial**, pour votre accueil, votre bonne humeur, vos encouragements et d'avoir pris de mes nouvelles à travers Shail pendant la dernière ligne droite de rédaction. Merci **Arnaud** de ton soutien et de ta bonne compagnie du midi (à 13h30).

Mes remerciements vont aux doctorants, post-doctorants et neuropsychologues de Cyron et du PFRS pour les bons moments, et en particulier à **Pierre**, **Thomas**, **Anaïs**, **Léa**, **Edelweiss**, **Francesca**, **Asrar**, **Laura Charretier**, **Kenza**, les deux **Lucie**, **Mathilde**, **Gwendolyn**, **Pauline**, **Elodie**, **Giovanni**, **Paolo**, **Joy**, **Gina** et **Marion** pour votre écoute, vos conseils et vos encouragements à différents moments. Merci Elodie de ton empathie et ton soutien, Anaïs de ta sincérité et de tes retours qui font chaud au coeur. Pierre, je ne peux pas ne pas mentionner notre première année de west coast swing en tant que leaders. C'était très amusant aussi de danser ensemble pendant les pratiques !

Merci aux **copains et copines de la bachata, la salsa et du west coast swing**, pour le partage, la bonne humeur et la cohésion. Les cours, stages et soirées ont été mes

bulles de déconnexion.

**Nicolas Cabé**, j'ai été très contente de me dire que ta thèse aussi aurait pu être présentée en plusieurs volumes et qu'elle contenait des remerciements à rallonge. Merci de toutes nos discussions scientifiques, dont certaines ont directement alimenté les chapitres 2 et 14 de ma thèse, de ta pédagogie, et d'avoir intégré très rapidement qu'on parlait deux langages différents. J'ai pris beaucoup de plaisir à échanger sur mon travail avec toi. Merci de ta générosité, de ton soutien et des nombreuses soirées passées chez toi, avec Shail et tes deux adorables minous. Tes idées loufoques d'emballage cadeau de la porte d'Anne-Lise pour Noël, de décorations égyptiennes pour son bureau, avec une session découpage chez toi jusqu'à une heure du matin m'auront bien fait rire. Merci d'ailleurs à la troupe Alcool de m'avoir intégrée à ces événements festifs, ça a été un plaisir de les partager avec vous.

**Alice**, tes cours de neuropsychologie et nos discussions sur le contrôle inhibiteur m'ont bien rassurée quand j'avais l'impression d'être perdue au milieu de tous les modèles de fonctions exécutives. Merci de ton aide pour cette partie introductive de ma thèse, de ton soutien et des discours positifs que tu as su m'adresser dans des moments difficiles, ils m'ont soulagée et encouragée.

Les Loustiques, ces années de thèse ont été rythmées par les nombreux repas et sorties toujours très chouettes en votre compagnie. Je suis reconnaissante de vous avoir eu si présents à mes côtés et de ses partages des moments de nos vies, les moins bons comme les meilleurs. Merci de m'avoir soutenue, écoutée et conseillée.

**Shail** (ou Diabloustik) si tu croyais que je n'allais pas te mentionner, et bien c'est fait, deux fois. J'ai passé tellement de temps avec toi pendant mon doctorat que je n'arrive même pas à imaginer ce qu'il aurait pu être sans toi. Merci de m'avoir autant fait rire (même quand c'est malgré toi avec des erreurs de traductions qui ne veulent plus dire pareil du tout...), taquinée et fait de blagues. Merci aussi de m'avoir rappelée que j'avais le droit de pleurer quand ça n'allait pas. Je te suis très reconnaissante de tout le soutien que tu m'as apporté, depuis le début. Il m'a été d'une aide précieuse au quotidien pour que je sois en capacité de continuer. Ta capacité d'analyse et ton empathie te conduisent à savoir parfois trouver les mots très justes qui résonnent chez moi. Je te remercie d'avoir veillé sur moi, y compris sur mes repas pendant les deux/trois dernières semaines de rédaction quand j'étais en ermite dans mon appartement. Shail et Nicolas, j'ai adoré vous entendre parler de vos expériences de rédaction de manuscrit. Elles m'ont permis je pense de prendre le temps de profiter encore plus de la balade pendant ma propre rédaction. Shail, merci pour tout le partage professionnel, amical et en tant que partenaire de danse ! C'est un plaisir de continuer à aller au west coast swing avec toi chaque semaine.

**Charlotte** (ou Charloustik), merci de la finesse de tes analyses, de ta générosité et de ton empathie. Comment oublier de te remercier pour ton texto d'il y a cinq ans me prévenant que ma thèse avait démarré et que je pouvais venir ! Je suis contente de pouvoir compter sur l'une de mes plus fidèles spectatrices de danse ! J'ai hâte de te voir au flamenco. Promis, on va avoir plus de temps pour aller danser maintenant et pour aller à Disney.

**Matthieu** (ou Bigloustik), nous sommes arrivés au labo en même temps. J'ai appris à te connaître, à savoir qu'il ne faut pas t'emmener au cinéma après 20h et que si je veux des



recommandations de tiramisu c'est à toi que je dois demander. Merci de ton humour et de ta sensibilité. C'est un plaisir de passer du temps avec toi, **Rosa** et les enfants. Merci aussi à tous les quatre pour mes nouveaux cactus de rendu de manuscrit (ils sont encore vivants!).

**Stéphane Réhel** (ou Fafanou le Magnifique), je suis bien contente que tu sois revenu au labo parce qu'on peut passer plus de temps avec toi. Tu me fais bien rire et j'aime beaucoup discuter avec toi. Merci de partager avec nous les aventures de tes deux petits bouts de chou et de ta famille.

Je tiens évidemment à prendre le temps de remercier mes ami.e.s de longue date : **Laura** et **Allan** (depuis la licence à Montpellier), **Francine** et **David** (depuis nos années de master à Toulouse). Merci à tous les quatre de faire encore aujourd'hui partie de ma vie, malgré le peu de temps que j'ai eu à vous consacrer ces dernières années. Promis, on devrait réussir à se voir plus souvent maintenant ! Merci de votre patience, de votre compréhension, de m'avoir soutenue et encouragée. Laura, ça a été un réel plaisir d'être en colloc' avec toi lors de mes passages toulousains pour le travail, merci beaucoup de m'avoir accueillie les bras ouverts (avec l'accord de tes deux chats).

Mes remerciements vont à mes parents, **Valérie** et **Frédéric**, ainsi qu'à ma soeur **Cindy**. Merci de tout votre amour, vos encouragements et de l'intérêt que vous avez porté à mon travail au quotidien. Maman et papa, merci de m'avoir enseigné la valeur du travail et le goût de l'effort. Papa, merci de nos discussions scientifiques et d'être passionné. Maman, merci de ton empathie et tes conseils avisés à chaque fois que j'ai besoin. Merci Cindy des questions que tu m'as posées et qui m'ont entraînée à dire ma problématique de thèse en quelques minutes, c'était un super exercice. J'ai adoré partir en Espagne avec toi juste après mon rendu de manuscrit ! J'espère qu'on va se voir plus régulièrement maintenant.

Je tiens également à remercier mes deux mamies, **Carmen** et **Annie**, pour votre amour et votre soutien.

Merci tatie **Jeanine** pour ton accueil à chaque fois que je redescends à Montpellier.

Tatie **Stépanie**, merci d'avoir pris régulièrement de mes nouvelles et de tes encouragements, particulièrement pendant les dernières périodes compliquées et pour la fin de la rédaction. Merci à mes cousines **Anaïs** et **Emilie** de leur soutien. A **Gregory** de me faire autant rire quand on se voit. A **Emy** et **Raphaël** de votre curiosité et des questions que vous m'avez posées sur ma recherche, à **Alicia** et **Noé** d'avoir été si enthousiastes en voyant en photo mon "livre", c'était adorable. Merci aussi à **Célia**, **Noa**, **Bastien** et au reste de ma famille.

Je remercie enfin mon comité de supporters à l'affût de l'envoi de mon manuscrit aux alentours du 15 août (les émetteurs de textos : "c'est bon c'est envoyé ?" se reconnaîtront). Un grand merci de m'avoir permis de vivre l'expérience des derniers jours de rédaction avec la sensation d'un tel soutien de votre part. J'ai hâte de partager avec vous ma soutenance.

Il est évident que j'oublie des personnes... merci à toutes celles et ceux qui d'une manière ou d'une autre ont partagé un bout de ces cinq années et ont participé à en faire une expérience si riche.

## Résumé étendu

En psychologie, le contrôle inhibiteur (ou inhibition) est un mécanisme cognitif qui permet de stopper une réponse motrice, émotionnelle ou cognitive non adaptée pour la réalisation d'un but désiré. Au niveau cérébral, le contrôle inhibiteur est associé au fonctionnement en réseau de régions cérébrales, qui peut être mesuré avec l'Imagerie par Résonance Magnétique fonctionnelle (IRMf). Des régions de contrôle préfrontales abaissent l'activité BOLD de régions cibles, qui diffèrent en fonction de la réponse à inhiber (motrice, cognitive ou émotionnelle, [Apšvalka et al., 2022]).

Il existe une variabilité interindividuelle des capacités de contrôle inhibiteur aux niveaux comportemental et cérébral. Les régions cérébrales sont constituées de neurones excitateurs et inhibiteurs, qui ont pour effets respectifs d'augmenter ou d'abaisser l'activité d'autres neurones. La métaphore de la balance locale Excitation/Inhibition (E/I) permet de se représenter les effets compensatoires qu'induisent la variation du poids excitateur ou du poids inhibiteur au sein d'une région cérébrale. La variabilité interindividuelle de l'état des balances E/I dans les régions cibles et/ou dans les régions de contrôle pourrait expliquer la variabilité interindividuelle des capacités de contrôle inhibiteur [Schmitz et al., 2017].

L'IRMf permet de mesurer de manière indirecte l'activité des neurones. La caractérisation des mécanismes cérébraux qui assurent la communication entre régions reste très limitée en raison de l'absence de modèle neural précis en IRMf, qui explicite les contributions des neurones excitateurs et inhibiteurs. Comment peut-on alors inférer à partir de données d'IRMf des propriétés excitatrices et inhibitrices (E/I) neurales au sein de régions cérébrales impliquées dans une tâche de contrôle inhibiteur ?

Nous partons d'un modèle biophysique non linéaire, hiérarchique qui décrit les évolutions temporelles des activités neurales excitatrice et inhibitrice par région ([Naskar et al., 2021]). Ces variations d'activité produisent des changements hémodynamiques puis BOLD dans chaque région cérébrale. Les outils d'analyses des systèmes non linéaires nous permettent de : 1) identifier des paramètres neuraux de la balance E/I; 2) montrer que, dans le modèle de [Naskar et al., 2021], l'augmentation d'activité BOLD d'une région de contrôle ne permet pas d'abaisser l'activité BOLD d'une région cible, quelque soit le jeu de paramètres neuraux qui caractérisent les balances E/I, parce que les régions sont connectées par leurs neurones excitateurs uniquement ; 3) proposer une nouvelle architecture de connectivité pour le permettre ; 4) étudier comment la baisse d'activité de la région cible dépend de la balance E/I dans la cible.

Nous estimons ensuite les paramètres neuraux de chaque région à partir des signaux BOLD recueillis pendant une tâche de contrôle inhibiteur. Notre procédure d'inversion, éclairée par l'analyse préalable du modèle direct, comprend deux étapes : une inversion bayésienne variationnelle ([Daunizeau et al., 2014]) suivie d'une inversion algébrique. Nous en vérifions la fiabilité avec des simulations, avant de présenter une preuve de concept sur les données d'un sujet pendant une tâche de Think/NoThink, un paradigme d'étude du contrôle inhibiteur des intrusions mnésiques [Mary et al., 2020].



# Table des matières

|           |                                                                                                                                                                              |           |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b>  | <b>Présentation de la thèse</b>                                                                                                                                              | <b>31</b> |
| 1.1       | Contexte . . . . .                                                                                                                                                           | 32        |
| 1.2       | Choix de modélisation . . . . .                                                                                                                                              | 34        |
| 1.3       | Organisation du travail de thèse . . . . .                                                                                                                                   | 36        |
| 1.3.1     | Première partie : Analyse du modèle direct (chapitre 9) . . . . .                                                                                                            | 36        |
| 1.3.2     | Deuxième partie : Estimation de paramètres (chapitres 10 et 11) . . . . .                                                                                                    | 36        |
| 1.4       | Perspectives d'application de la thèse : étude du contrôle inhibiteur dans le trouble de stress post-traumatique . . . . .                                                   | 37        |
| <b>I</b>  | <b>Introduction</b>                                                                                                                                                          | <b>39</b> |
| <b>2</b>  | <b>Le contrôle inhibiteur, un processus psychologique adaptatif associé au fonctionnement de réseaux cérébraux qui impliquent le cortex préfrontal et des régions cibles</b> | <b>41</b> |
| 2.1       | Différents niveaux d'analyse . . . . .                                                                                                                                       | 42        |
| 2.2       | Contrôle inhibiteur en psychologie . . . . .                                                                                                                                 | 42        |
| 2.2.1     | Le contrôle inhibiteur : une des principales fonctions exécutives . . . . .                                                                                                  | 43        |
| 2.2.1.1   | Définition des fonctions exécutives . . . . .                                                                                                                                | 43        |
| 2.2.1.2   | Principales fonctions exécutives . . . . .                                                                                                                                   | 43        |
| 2.2.1.3   | Contrôle inhibiteur : un des principaux processus exécutifs . . . . .                                                                                                        | 43        |
| 2.2.2     | Différentes formes d'inhibition et tâches associées . . . . .                                                                                                                | 44        |
| 2.2.2.1   | Inhibition d'une réponse motrice . . . . .                                                                                                                                   | 45        |
| 2.2.2.2   | Inhibition émotionnelle . . . . .                                                                                                                                            | 45        |
| 2.2.2.3   | Inhibition cognitive . . . . .                                                                                                                                               | 46        |
| 2.2.2.3.1 | Contrôle inhibiteur de l'attention . . . . .                                                                                                                                 | 46        |
| 2.2.2.3.2 | Contrôle inhibiteur de pensées . . . . .                                                                                                                                     | 47        |
| 2.2.3     | Contrôle inhibiteur des intrusions mnésiques . . . . .                                                                                                                       | 47        |
| 2.2.3.1   | Oubli adaptatif et suppression mnésique . . . . .                                                                                                                            | 47        |
| 2.2.3.2   | Paradigme Think/No Think . . . . .                                                                                                                                           | 48        |
| 2.2.3.3   | Résultats comportementaux . . . . .                                                                                                                                          | 49        |
| 2.2.4     | Contrôle inhibiteur proactif et réactif . . . . .                                                                                                                            | 50        |
| 2.3       | Contrôle inhibiteur à l'échelle des régions cérébrales . . . . .                                                                                                             | 51        |
| 2.3.1     | Différentes méthodes pour étudier les mécanismes cérébraux du contrôle inhibiteur . . . . .                                                                                  | 51        |

|           |                                                                                                                                  |           |
|-----------|----------------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.3.2     | Etude en Imagerie par Résonance Magnétique fonctionnelle . . . . .                                                               | 51        |
| 2.3.2.1   | Régions de contrôle . . . . .                                                                                                    | 52        |
| 2.3.2.2   | Régions cibles . . . . .                                                                                                         | 52        |
| 2.3.2.3   | Du signal BOLD à l'interprétation neurale ? . . . . .                                                                            | 53        |
| <b>3</b>  | <b>Neurophysiologie de l'inhibition</b>                                                                                          | <b>55</b> |
| 3.1       | Fonctionnement biologique d'un neurone . . . . .                                                                                 | 55        |
| 3.1.1     | Quelques éléments sur la physiologie des neurones et de leurs membranes                                                          | 55        |
| 3.1.2     | Transmission chimique de l'influx nerveux : synapses excitatrice et inhibitrice . . . . .                                        | 56        |
| 3.1.2.1   | Neurone excitateur : diminue la ddp <sub>m</sub> du neurone post-synaptique . . . . .                                            | 58        |
| 3.1.2.1.1 | Récepteur AMPA . . . . .                                                                                                         | 58        |
| 3.1.2.1.2 | Récepteur NMDA . . . . .                                                                                                         | 58        |
| 3.1.2.2   | Neurone inhibiteur : augmente la ddp <sub>m</sub> du neurone post-synaptique . . . . .                                           | 58        |
| 3.1.2.3   | Intégration des courants par le neurone post-synaptique . . . . .                                                                | 58        |
| 3.1.3     | Transmission électrique de l'influx nerveux : les potentiels d'actions . . . . .                                                 | 59        |
| 3.1.3.1   | Naissance d'un potentiel d'action . . . . .                                                                                      | 59        |
| 3.1.3.2   | Propagation du potentiel d'action par dépolarisation vers l'avant des régions voisines . . . . .                                 | 60        |
| 3.2       | Deux types d'inhibition GABAergique : phasique et tonique . . . . .                                                              | 61        |
| 3.2.1     | Récepteurs GABA-A synaptiques et extra-synaptiques . . . . .                                                                     | 61        |
| 3.2.2     | Inhibition phasique synaptique . . . . .                                                                                         | 62        |
| 3.2.3     | Inhibition tonique extrasynaptique . . . . .                                                                                     | 62        |
| 3.3       | Inhibition GABAergique tonique dans l'hippocampe . . . . .                                                                       | 63        |
| 3.3.1     | Effets d'une levée de l'inhibition tonique sur les apprentissages dépendant de l'hippocampe (modèles animaux) . . . . .          | 63        |
| 3.3.2     | Métaphore de la balance Excitation/Inhibition . . . . .                                                                          | 64        |
| 3.3.3     | La balance Excitation/Inhibition pour penser les apprentissages dépendant de l'hippocampe . . . . .                              | 65        |
| 3.4       | Contrôle inhibiteur des intrusions mnésiques . . . . .                                                                           | 67        |
| 3.4.1     | Mesures de concentrations GABA par spectroscopie . . . . .                                                                       | 67        |
| 3.4.2     | Hypothèse de la variabilité interindividuelle de l'état des balances E/I . . . . .                                               | 68        |
| <b>4</b>  | <b>Signaux BOLD obtenus avec l'IRMf : des mesures très indirectes de l'activité des neurones sous-jacents</b>                    | <b>69</b> |
| 4.1       | Principe physique de l'IRM : obtenir une mesure de la relaxation de l'aimantation macroscopique locale dans les tissus . . . . . | 70        |
| 4.1.1     | Définition du système, de l'observable macroscopique et des directions . . . . .                                                 | 70        |
| 4.1.2     | Condition initiale : application d'un champ magnétique constant $\vec{B}_0$ et description de l'état d'équilibre . . . . .       | 71        |
| 4.1.3     | Perturbation du système par des impulsions radio fréquences et résonance magnétique . . . . .                                    | 72        |
| 4.1.4     | Arrêt des impulsions radio-fréquence : phase de relaxation transverse . . . . .                                                  | 73        |

|           |                                                                                                                                                                                                                               |           |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 4.1.4.1   | Relaxation $T_2$ . . . . .                                                                                                                                                                                                    | 74        |
| 4.1.4.2   | Relaxation $T_2^*$ . . . . .                                                                                                                                                                                                  | 75        |
| 4.2       | Variation de la concentration locale en déoxyhémoglobine due au couplage neurovasculaire . . . . .                                                                                                                            | 76        |
| 4.3       | Principe de l'IRM fonctionnelle : $T_2^*$ dépend des variations locales de la concentration en déoxyhémoglobine . . . . .                                                                                                     | 77        |
| 4.3.1     | Oxyhémoglobine et déoxyhémoglobine : des propriétés magnétiques différentes . . . . .                                                                                                                                         | 77        |
| 4.3.1.1   | Composition . . . . .                                                                                                                                                                                                         | 77        |
| 4.3.1.2   | Propriété magnétique . . . . .                                                                                                                                                                                                | 78        |
| 4.3.2     | Principe de la mesure en IRMf : un signal dépend du niveau d'oxygénation du sang (BOLD) . . . . .                                                                                                                             | 79        |
| 4.3.2.1   | Présence de déoxyhémoglobine et variations de $T_2^*$ . . . . .                                                                                                                                                               | 79        |
| 4.3.2.2   | Mesures du signal BOLD dans un voxel . . . . .                                                                                                                                                                                | 79        |
| 4.3.2.2.1 | Signal BOLD pondéré en $T_2^*$ . . . . .                                                                                                                                                                                      | 80        |
| 4.3.2.2.2 | Série temporelle de mesures . . . . .                                                                                                                                                                                         | 80        |
| 4.3.2.3   | Importance des contrastes . . . . .                                                                                                                                                                                           | 81        |
| 4.3.2.4   | Signal BOLD et activité neurale sous-jacente . . . . .                                                                                                                                                                        | 81        |
| 4.3.3     | Nature des données recueillies en IRMf . . . . .                                                                                                                                                                              | 82        |
| 4.3.4     | Description d'une réponse . . . . .                                                                                                                                                                                           | 82        |
| 4.3.5     | Limites . . . . .                                                                                                                                                                                                             | 82        |
| 4.3.5.1   | Mesures indirectes . . . . .                                                                                                                                                                                                  | 82        |
| 4.3.5.2   | Données très bruitées . . . . .                                                                                                                                                                                               | 83        |
| 4.3.5.3   | Résolution temporelle . . . . .                                                                                                                                                                                               | 84        |
| 5         | <b>Etudier la communication entre régions cérébrales à partir de signaux BOLD : nécessité de considérer un modèle dynamique du signal BOLD, des variations de l'activité neurale et des couplages inter-régions</b> . . . . . | <b>85</b> |
| 5.1       | Ségrégation et intégration fonctionnelle . . . . .                                                                                                                                                                            | 86        |
| 5.1.1     | Etude de la ségrégation fonctionnelle . . . . .                                                                                                                                                                               | 86        |
| 5.1.1.1   | Principe . . . . .                                                                                                                                                                                                            | 86        |
| 5.1.1.2   | Exemple du contrôle inhibiteur des intrusions mnésiques . . . . .                                                                                                                                                             | 87        |
| 5.1.2     | Etude de l'intégration fonctionnelle . . . . .                                                                                                                                                                                | 88        |
| 5.1.2.1   | Connectivité fonctionnelle . . . . .                                                                                                                                                                                          | 88        |
| 5.1.2.2   | Connectivité effective . . . . .                                                                                                                                                                                              | 89        |
| 5.2       | Analyses de connectivité fonctionnelle non basée sur un modèle dynamique du signal . . . . .                                                                                                                                  | 89        |
| 5.2.1     | Connectivité fonctionnelle entre deux régions . . . . .                                                                                                                                                                       | 89        |
| 5.2.2     | Carte de connectivité fonctionnelle . . . . .                                                                                                                                                                                 | 89        |
| 5.2.3     | Connectome et réseau fonctionnel . . . . .                                                                                                                                                                                    | 90        |
| 5.2.4     | Limites . . . . .                                                                                                                                                                                                             | 90        |
| 5.3       | Les modèles dynamiques des variations du signal BOLD . . . . .                                                                                                                                                                | 91        |
| 5.3.1     | De la stimulation expérimentale aux variations de l'activité neurale . . . . .                                                                                                                                                | 91        |
| 5.3.1.1   | Un modèle des variations de l'activité neurale pendant une tâche : la Stickfunction . . . . .                                                                                                                                 | 91        |

|             |                                                                                                                              |     |
|-------------|------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.3.1.2     | Modèle phénoménologique linéaire de l'activité neurale (DCM)                                                                 | 92  |
| 5.3.1.2.1   | Modèle sans stimulation                                                                                                      | 92  |
| 5.3.1.2.2   | Modèle de l'effet produit par une condition expérimentale                                                                    | 92  |
| 5.3.1.2.3   | Modèle des effets produits par plusieurs conditions expérimentales                                                           | 93  |
| 5.3.2       | De l'activité neurale au signal BOLD                                                                                         | 94  |
| 5.3.2.1     | Sans modèle explicite du couplage neurovasculaire                                                                            | 94  |
| 5.3.2.1.1   | Fonction de réponse hémodynamique ou HRF                                                                                     | 94  |
| 5.3.2.1.2   | Du neural au BOLD par convolution                                                                                            | 95  |
| 5.3.2.2     | Avec un modèle explicite du couplage neurovasculaire                                                                         | 97  |
| 5.3.2.2.1   | Modèle hémodynamique                                                                                                         | 97  |
| 5.3.2.2.1.1 | Motivations et objectifs                                                                                                     | 97  |
| 5.3.2.2.1.2 | Présentation du modèle et des variables d'état                                                                               | 97  |
| 5.3.2.2.1.3 | Structure du modèle                                                                                                          | 98  |
| 5.3.2.2.1.4 | Propriétés                                                                                                                   | 99  |
| 5.3.2.2.1.5 | 1er bloc : Comment les variations de l'activité neurale provoquent une variation de flux en entrée du compartiment veineux ? | 99  |
| 5.3.2.2.1.6 | Balloon-Windkessel model                                                                                                     | 100 |
| 5.3.2.2.2   | Modèle BOLD non linéaire                                                                                                     | 103 |
| 5.3.2.2.2.1 | En fonction des signaux intra et extra-vasculaires                                                                           | 103 |
| 5.3.2.2.2.2 | En fonction du volume et de la quantité de déoxyhémoglobine                                                                  | 105 |
| 5.3.2.2.3   | Modèle complet et simulations numériques                                                                                     | 106 |
| 5.4         | Analyse de connectivité fonctionnelle basée sur un modèle dynamique : PPI (Psychophysiological Interactions)                 | 111 |
| 5.4.1       | Objectif de l'analyse                                                                                                        | 111 |
| 5.4.2       | Les régresseurs construits à partir du modèle de convolution linéaire                                                        | 112 |
| 5.4.2.1     | Régresseur PPI au niveau BOLD                                                                                                | 112 |
| 5.4.2.2     | Réponse BOLD théorique de la cible due uniquement à sa stimulation pendant la tâche                                          | 113 |
| 5.4.3       | Modèle statistique linéaire général (ou GLM)                                                                                 | 113 |
| 5.4.4       | Exemple du contrôle inhibiteur des intrusions mnésiques                                                                      | 114 |
| 5.4.4.1     | Résultats d'imagerie                                                                                                         | 114 |
| 5.4.4.2     | Liens avec la concentration GABA                                                                                             | 114 |
| 5.4.4.2.1   | Dans les régions de contrôle                                                                                                 | 114 |
| 5.4.4.2.2   | Dans les régions cibles                                                                                                      | 115 |
| 5.5         | Analyse de connectivité effective basée sur un modèle dynamique : DCM                                                        | 115 |
| 5.5.1       | Présentation de la méthode d'analyse                                                                                         | 115 |
| 5.5.1.1     | Modèle dynamique                                                                                                             | 116 |
| 5.5.1.2     | Estimation des paramètres de ce modèle à partir de données                                                                   | 116 |
| 5.5.2       | Exemple du contrôle inhibiteur des intrusions mnésiques                                                                      | 117 |
| 5.5.2.1     | Quel modèle direct ?                                                                                                         | 117 |
| 5.5.2.2     | Sélection de modèles à l'échelle d'un groupe de participants                                                                 | 119 |

|           |                                                                                                       |     |
|-----------|-------------------------------------------------------------------------------------------------------|-----|
| 5.5.2.3   | Modulation de couplage du DLPFC vers l'hippocampe en lien avec les données comportementales . . . . . | 119 |
| 5.5.2.3.1 | Performances de rappel post-TNT . . . . .                                                             | 119 |
| 5.5.2.3.2 | Reports d'intrusions pendant le TNT . . . . .                                                         | 120 |
| 5.5.2.4   | Un mécanisme adaptatif . . . . .                                                                      | 120 |
| 5.5.2.5   | Forçage dans le DLPFC et liens avec les concentration de GABA . . . . .                               | 121 |
| 5.6       | Le positionnement de la thèse . . . . .                                                               | 121 |
| 5.6.1     | Limites des études actuelles sur le contrôle inhibiteur des intrusions mnésiques . . . . .            | 121 |
| 5.6.2     | La question que se pose la communauté . . . . .                                                       | 122 |

## II Choix de modélisation 123

|          |                                                                                                                                  |            |
|----------|----------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>6</b> | <b>Modèle du signal BOLD construit à partir des variations de l'activité neurale excitatrice et inhibitrice</b>                  | <b>125</b> |
| 6.1      | Modélisation de la dynamique d'association-dissociation entre récepteurs et neurotransmetteurs au niveau d'une synapse . . . . . | 125        |
| 6.1.1    | A l'échelle d'un récepteur . . . . .                                                                                             | 126        |
| 6.1.1.1  | Modélisation de l'état d'un récepteur par une chaîne de Markov entre $t$ et $t + dt$ . . . . .                                   | 126        |
| 6.1.1.2  | Evolution temporelle de la probabilité de l'état lié du récepteur                                                                | 126        |
| 6.1.2    | A l'échelle de la synapse . . . . .                                                                                              | 128        |
| 6.1.2.1  | Evolution temporelle de la fraction post-synaptique de récepteurs liés . . . . .                                                 | 128        |
| 6.1.2.2  | Evolution temporelle de la fraction post-synaptique de canaux ouverts . . . . .                                                  | 128        |
| 6.1.2.3  | Décharge présynaptique et libération de neurotransmetteurs                                                                       | 129        |
| 6.1.2.4  | Canaux excitateurs et canaux inhibiteurs . . . . .                                                                               | 129        |
| 6.2      | Modélisation de la dynamique neurale à l'échelle des régions : approche en champ moyen . . . . .                                 | 131        |
| 6.2.1    | Activités synaptiques moyennes . . . . .                                                                                         | 132        |
| 6.2.1.1  | Equations d'évolution temporelle des fractions moyennes de canaux ouverts . . . . .                                              | 132        |
| 6.2.1.2  | Bruit . . . . .                                                                                                                  | 133        |
| 6.2.1.3  | Valeurs des paramètres . . . . .                                                                                                 | 133        |
| 6.2.2    | Fréquences moyennes de potentiels d'action . . . . .                                                                             | 134        |
| 6.2.2.1  | Fonctions des courants ioniques . . . . .                                                                                        | 134        |
| 6.2.2.2  | Valeurs des paramètres . . . . .                                                                                                 | 135        |
| 6.2.3    | Courants ioniques moyens . . . . .                                                                                               | 135        |
| 6.2.3.1  | Courant ionique moyen en entrée d'un pool excitateur . . .                                                                       | 135        |
| 6.2.3.2  | Courant ionique moyen en entrée d'un pool inhibiteur . . .                                                                       | 136        |
| 6.2.3.3  | Valeurs des paramètres . . . . .                                                                                                 | 137        |
| 6.2.4    | Modèle de stimulation pendant une tâche de contrôle inhibiteur . . .                                                             | 137        |



|           |                                                                                                                                         |            |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------|------------|
| 6.2.5     | Cas de deux régions : une région de contrôle et une région cible . . . .                                                                | 138        |
| 6.3       | Modèle de la mesure : modèles hémodynamique et BOLD . . . . .                                                                           | 139        |
| 6.3.1     | Modèle hémodynamique : extended Balloon-Windkessel Model . . . .                                                                        | 139        |
| 6.3.2     | Modèle BOLD . . . . .                                                                                                                   | 140        |
| 6.4       | Traduction de la question que pose la communauté étant donné ces choix de modélisation . . . . .                                        | 140        |
| <b>7</b>  | <b>Problématique de thèse et organisation du travail</b>                                                                                | <b>143</b> |
| 7.1       | Etude du passage entre les niveaux de description neurale et BOLD . . . . .                                                             | 144        |
| 7.1.1     | Comparaison entre les temps d'installation du modèle neural et du modèle hémodynamique . . . . .                                        | 144        |
| 7.1.1.1   | Temps caractéristiques du modèle hémodynamique . . . . .                                                                                | 144        |
| 7.1.1.2   | Temps caractéristiques du modèle neural pour un système à deux régions cérébrales . . . . .                                             | 145        |
| 7.1.1.2.1 | Etude dynamique par simulations . . . . .                                                                                               | 146        |
| 7.1.1.2.2 | Analyse linéaire au voisinage de l'état stationnaire . . . . .                                                                          | 147        |
| 7.1.1.3   | Que devient le modèle ? . . . . .                                                                                                       | 149        |
| 7.1.2     | Relation entre le signal BOLD $y_i^*$ émis par une région cérébrale et son activité synaptique à l'état stationnaire $sn_i^*$ . . . . . | 150        |
| 7.1.3     | Conclusion pour la précision de la problématique . . . . .                                                                              | 152        |
| 7.1.3.1   | Pour le modèle direct . . . . .                                                                                                         | 152        |
| 7.1.3.2   | Pour l'inversion . . . . .                                                                                                              | 152        |
| 7.2       | Etude d'une région isolée : stimulation d'un des deux pools . . . . .                                                                   | 153        |
| 7.2.1     | Formulation mathématique du modèle au régime stationnaire . . . . .                                                                     | 153        |
| 7.2.2     | Etude de sensibilité par rapport au forçage sur le pool exciteur ou inhibiteur . . . . .                                                | 154        |
| 7.2.3     | Conclusion pour la précision de la problématique . . . . .                                                                              | 156        |
| 7.3       | Identification de paramètres permettant de modifier l'état de la balance E/I locale . . . . .                                           | 156        |
| 7.3.1     | Notion de balance locale, paramètres locaux E et I . . . . .                                                                            | 156        |
| 7.3.2     | Approche analytique . . . . .                                                                                                           | 157        |
| 7.3.2.1   | Etude de sensibilité par rapport au forçage exciteur sur le pool exciteur et par rapport à $J_{gaba}$ . . . . .                         | 157        |
| 7.3.3     | Choix des paramètres pour le système à deux régions . . . . .                                                                           | 162        |
| 7.3.3.1   | Guidé par la question de la variabilité interindividuelle . . . .                                                                       | 162        |
| 7.3.3.2   | Hypothèse sur l'effet d'une augmentation de $I_{stim}$ dans la RC . . . .                                                               | 162        |
| 7.3.3.3   | Hypothèse sur l'effet de $J_{gaba_c}$ . . . . .                                                                                         | 162        |
| 7.3.3.4   | Précision des hypothèses en lien avec la question de la variabilité interindividuelle . . . . .                                         | 162        |
| 7.4       | Précision des questions de recherche : explicitation des objectifs précis de la thèse . . . . .                                         | 163        |
| 7.5       | Présentation des étapes du travail . . . . .                                                                                            | 163        |

|              |                                                                                                              |            |
|--------------|--------------------------------------------------------------------------------------------------------------|------------|
| <b>III</b>   | <b>Analyse du modèle direct</b>                                                                              | <b>165</b> |
| <b>8</b>     | <b>Préambule de l'article</b>                                                                                | <b>167</b> |
| <b>9</b>     | <b>Analyse du comportement du modèle pour 2 régions</b>                                                      | <b>169</b> |
| 9.1          | Article . . . . .                                                                                            | 169        |
| 9.2          | Annexes de l'article . . . . .                                                                               | 209        |
| <b>IV</b>    | <b>Inversion</b>                                                                                             | <b>229</b> |
| <b>10</b>    | <b>Construction de la procédure d'inversion</b>                                                              | <b>231</b> |
| 10.1         | Le modèle direct . . . . .                                                                                   | 232        |
| 10.1.1       | Deux paramètres d'intérêt par région : $J_{gaba_i}$ et $I_{stim_i}$ . . . . .                                | 232        |
| 10.1.2       | Formulation mathématique du modèle complet . . . . .                                                         | 232        |
| 10.2         | Présentation du découplage en 2 étapes . . . . .                                                             | 234        |
| 10.2.1       | Première étape . . . . .                                                                                     | 235        |
| 10.2.2       | Deuxième étape . . . . .                                                                                     | 235        |
| 10.3         | Première étape de l'inversion . . . . .                                                                      | 235        |
| 10.3.1       | Principe de l'approche bayésienne . . . . .                                                                  | 235        |
| 10.3.1.1     | Exemple de l'estimation de paramètres . . . . .                                                              | 235        |
| 10.3.1.2     | Exemple de l'estimation de paramètres et d'états cachés . . . . .                                            | 237        |
| 10.3.2       | Réécriture des modèles hémodynamique et BOLD préalable à l'utilisation de la toolbox VBA . . . . .           | 238        |
| 10.3.2.1     | Modèle hémodynamique et BOLD non transformés . . . . .                                                       | 238        |
| 10.3.2.2     | Changement d'écriture des paramètres à inverser . . . . .                                                    | 239        |
| 10.3.2.3     | Changement d'écriture des variables d'état . . . . .                                                         | 240        |
| 10.3.2.3.1   | Définition des changement de variable . . . . .                                                              | 240        |
| 10.3.2.3.2   | Modèle hémodynamique transformé . . . . .                                                                    | 241        |
| 10.3.2.3.3   | Modèle BOLD transformé . . . . .                                                                             | 243        |
| 10.3.3       | Définition des fonctions d'évolution $\vec{f}$ et d'observation $\vec{g}$ . . . . .                          | 244        |
| 10.3.3.1     | Fonction d'évolution $\vec{f}$ . . . . .                                                                     | 244        |
| 10.3.3.2     | Fonction d'observation $\vec{g}$ . . . . .                                                                   | 245        |
| 10.3.4       | Définition du modèle statistique direct génératif de la mesure BOLD . . . . .                                | 248        |
| 10.3.4.1     | Paramètres hémodynamiques : $\tilde{s}n_{B_i}$ et $\tilde{s}n_{stim_i}$ . . . . .                            | 248        |
| 10.3.4.1.1   | Définition des variables aléatoires . . . . .                                                                | 248        |
| 10.3.4.1.2   | Distributions de probabilités marginales ou a priori . . . . .                                               | 248        |
| 10.3.4.2     | Processus stochastiques d'états cachés et d'états observables . . . . .                                      | 248        |
| 10.3.4.2.1   | Définition des variables aléatoires . . . . .                                                                | 248        |
| 10.3.4.2.2   | Etats cachés : Définition des probabilités conditionnelles et des probabilités marginales/a priori . . . . . | 249        |
| 10.3.4.2.2.1 | Probabilités de transition entre états cachés . . . . .                                                      | 249        |
| 10.3.4.2.2.2 | Loi de probabilité de $\vec{X}_0$ . . . . .                                                                  | 250        |
| 10.3.4.2.2.3 | Représentation de la chaîne de Markov associée aux états cachés . . . . .                                    | 250        |

|              |                                                                                                              |     |
|--------------|--------------------------------------------------------------------------------------------------------------|-----|
| 10.3.4.2.2.4 | Densité de probabilité d'émission d'une séquence d'états cachés . . . . .                                    | 251 |
| 10.3.4.2.3   | Etats observables . . . . .                                                                                  | 252 |
| 10.3.4.2.3.1 | Probabilités d'observation . . . . .                                                                         | 252 |
| 10.3.4.2.3.2 | Représentation du Modèle de Markov Caché . . . . .                                                           | 252 |
| 10.3.4.2.3.3 | Densité de probabilité d'émission d'une séquence de mesures . . . . .                                        | 253 |
| 10.3.4.3     | Hyperparamètres de bruit sur l'observable . . . . .                                                          | 254 |
| 10.3.4.3.1   | Variables aléatoires et distributions de probabilité . . . . .                                               | 254 |
| 10.3.4.3.2   | Que devient la probabilité d'observation ? . . . . .                                                         | 255 |
| 10.3.4.3.3   | Loi marginale a priori sur l'inverse de la variance du bruit . . . . .                                       | 255 |
| 10.3.4.3.4   | Représentation du Modèle de Markov Caché avec bruit sur l'observable . . . . .                               | 256 |
| 10.3.5       | Méthode d'inversion . . . . .                                                                                | 257 |
| 10.3.5.1     | Méthode bayésienne variationnelle : un problème d'optimisation fonctionnelle . . . . .                       | 257 |
| 10.3.5.2     | Hypothèses de travail pour chercher une loi approchante optimale $q^{opt}$ . . . . .                         | 259 |
| 10.3.5.2.1   | Approximation en champ moyen . . . . .                                                                       | 259 |
| 10.3.5.2.2   | Approximation de Laplace . . . . .                                                                           | 261 |
| 10.3.5.3     | Principe de l'algorithme d'estimation des paramètres . . . . .                                               | 261 |
| 10.3.5.4     | Entrées et sorties de VBA . . . . .                                                                          | 262 |
| 10.3.6       | Extension à d'autres possibilités d'inversion . . . . .                                                      | 263 |
| 10.3.6.1     | Trois paramètres par région . . . . .                                                                        | 264 |
| 10.3.6.2     | Quatre paramètres par région . . . . .                                                                       | 264 |
| 10.3.7       | Récapitulatif . . . . .                                                                                      | 266 |
| 10.4         | Test de fiabilité de la première étape : inversion de simulations avec différents niveaux de bruit . . . . . | 267 |
| 10.4.1       | Simulations . . . . .                                                                                        | 267 |
| 10.4.1.1     | Une simulation . . . . .                                                                                     | 267 |
| 10.4.1.2     | Tester différents niveaux de bruit . . . . .                                                                 | 267 |
| 10.4.2       | Inversion des simulations . . . . .                                                                          | 268 |
| 10.4.2.1     | Densités de probabilité a priori . . . . .                                                                   | 268 |
| 10.4.2.1.1   | Paramètres hémodynamiques . . . . .                                                                          | 268 |
| 10.4.2.1.2   | Paramètres de bruit . . . . .                                                                                | 269 |
| 10.4.2.1.3   | Condition initiale . . . . .                                                                                 | 269 |
| 10.4.2.2     | Fixer et inverser la condition initiale ? . . . . .                                                          | 270 |
| 10.4.2.3     | Sorties d'inversion pour une condition initiale contenue dans l'a priori . . . . .                           | 270 |
| 10.4.2.3.1   | 2 paramètres par région . . . . .                                                                            | 272 |
| 10.4.2.3.2   | 3 paramètres par région . . . . .                                                                            | 273 |
| 10.4.2.3.3   | 4 paramètres par région . . . . .                                                                            | 274 |
| 10.4.2.4     | Sorties d'inversion pour une condition initiale non contenue dans l'a priori . . . . .                       | 275 |

|            |                                                                                     |            |
|------------|-------------------------------------------------------------------------------------|------------|
| 10.4.2.4.1 | 2 paramètres par région . . . . .                                                   | 276        |
| 10.4.2.4.2 | 3 paramètres par région . . . . .                                                   | 277        |
| 10.4.2.4.3 | 4 paramètres par région . . . . .                                                   | 278        |
| 10.5       | Lancer la première étape sur toutes les régions en même temps ? . . . . .           | 279        |
| 10.6       | Deuxième étape de l'inversion . . . . .                                             | 281        |
| 10.6.1     | Introduction . . . . .                                                              | 281        |
| 10.6.2     | Nouvelle formulation du modèle . . . . .                                            | 281        |
| 10.6.3     | Principe de la méthode d'inversion algébrique des paramètres neuraux . . . . .      | 282        |
| 10.6.4     | Inversion . . . . .                                                                 | 283        |
| 10.6.4.1   | Inversion des $J_{gaba}$ hors stimulation . . . . .                                 | 283        |
| 10.6.4.1.1 | Modèle hors stimulation . . . . .                                                   | 283        |
| 10.6.4.1.2 | 1ère étape : travail sur le pool inhibiteur pour trouver les $sg_{B_i}^*$ . . . . . | 283        |
| 10.6.4.1.3 | 2ème étape : travail sur le pool exciteur pour inverser $J_{gaba_i}$ . . . . .      | 284        |
| 10.6.4.2   | Inversion des $I_{stim}$ en période de stimulation . . . . .                        | 285        |
| 10.6.4.2.1 | Modèle en période de stimulation . . . . .                                          | 285        |
| 10.6.4.2.2 | 1ère étape : travail sur le pool inhibiteur pour trouver $sg_{stim_i}^*$ . . . . .  | 286        |
| 10.6.4.2.3 | 2ème étape : travail sur le pool exciteur pour inverser $I_{stim_i}$ . . . . .      | 286        |
| 10.6.5     | Schéma récapitulatif de la deuxième étape . . . . .                                 | 288        |
| <b>11</b>  | <b>Preuve de concept : Inversion des données d'un participant</b>                   | <b>289</b> |
| 11.1       | Première étape de l'inversion : du BOLD à $sn^*$ . . . . .                          | 290        |
| 11.2       | Deuxième étape . . . . .                                                            | 293        |
| 11.2.1     | $J_{gaba}$ par région . . . . .                                                     | 293        |
| 11.2.2     | $I_{stim}$ par région . . . . .                                                     | 294        |
| <b>V</b>   | <b>Discussion</b>                                                                   | <b>297</b> |
| <b>12</b>  | <b>Synthèse</b>                                                                     | <b>299</b> |
| 12.1       | Rappel des objectifs de la thèse . . . . .                                          | 299        |
| 12.2       | Principaux résultats . . . . .                                                      | 300        |
| 12.2.1     | Un nouveau modèle de connectivité . . . . .                                         | 300        |
| 12.2.2     | Structure de connectivité pertinente pour du contrôle inhibiteur . . . . .          | 300        |
| 12.2.3     | Effets de $I_{stim}$ sur la baisse d'activité $sn_c^*$ . . . . .                    | 300        |
| 12.2.4     | Effets de $J_{gaba_c}$ dans la cible sur la baisse d'activité $sn_c^*$ . . . . .    | 301        |
| 12.2.5     | Procédure d'inversion des paramètres neuraux à partir de mesures BOLD . . . . .     | 301        |
| 12.3       | Méthode de travail . . . . .                                                        | 301        |
| 12.3.1     | Séparation de la modélisation et de la technique/méthode d'inversion . . . . .      | 301        |
| 12.3.2     | Méthode d'analyse du modèle direct . . . . .                                        | 302        |

|                                                                                                                                                          |            |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>13 Préciser le langage pour partager le questionnement</b>                                                                                            | <b>305</b> |
| 13.1 Balance E/I locale                                                                                                                                  | 306        |
| 13.1.1 Comment définit-on l'effet excitateur ou inhibiteur : Etude de l'effet d'un paramètre                                                             | 307        |
| 13.1.1.1 Etude en perturbation infinitésimale : effet excitateur ou inhibiteur infinitésimal d'un paramètre                                              | 307        |
| 13.1.1.2 Cas de la balance E/I locale : caractérisation de l'effet excitateur ou inhibiteur infinitésimal d'un paramètre                                 | 308        |
| 13.1.1.3 Etude en variation finie : effet excitateur ou inhibiteur fini d'un paramètre                                                                   | 310        |
| 13.1.1.4 Augmentation de la variation du paramètre : prédiction de l'effet ?                                                                             | 311        |
| 13.1.1.5 Cas de la balance E/I : caractérisation de l'effet excitateur ou inhibiteur fini d'un paramètre                                                 | 312        |
| 13.1.2 Etude de l'effet de plusieurs paramètres                                                                                                          | 318        |
| 13.1.2.1 Perturbation simultanée de plusieurs paramètres                                                                                                 | 318        |
| 13.1.2.2 Variation simultanée de plusieurs paramètres                                                                                                    | 318        |
| 13.1.3 Résumé de la métaphore de la balance E/I locale pour penser le comportement du modèle d'une région isolée                                         | 321        |
| 13.1.3.1 Au régime stationnaire                                                                                                                          | 321        |
| 13.1.3.2 Changement du poids sur le plateau excitateur                                                                                                   | 321        |
| 13.1.3.3 Changement du poids sur le plateau inhibiteur                                                                                                   | 322        |
| 13.1.3.4 Variation de plusieurs poids                                                                                                                    | 322        |
| 13.1.4 Limites de la métaphore de la balance                                                                                                             | 323        |
| 13.1.4.1 Effet de la variation d'un paramètre sur $sn^*$                                                                                                 | 323        |
| 13.1.4.1.1 Propriété de linéarité ?                                                                                                                      | 323        |
| 13.1.4.1.2 Déplacements identiques du plateau E et du plateau I ?                                                                                        | 325        |
| 13.1.4.2 Effet de la variation de plusieurs paramètres : penser la compensation                                                                          | 325        |
| 13.1.4.2.1 En perturbation                                                                                                                               | 326        |
| 13.1.4.2.2 En variation                                                                                                                                  | 326        |
| 13.1.5 Poids synaptiques GABAergique et glutamatergique : quels effets sur la balance E/I ?                                                              | 328        |
| 13.1.5.1 Poids GABAergiques                                                                                                                              | 329        |
| 13.1.5.2 Poids glutamatergiques                                                                                                                          | 329        |
| 13.1.6 Bilan                                                                                                                                             | 330        |
| 13.2 Comment définir le contrôle inhibiteur d'une région cible ? Que signifie augmenter l'inhibition d'une région cible ?                                | 331        |
| 13.2.1 Contrôle inhibiteur d'une région cible : une pensée en variations qui implique des régions de contrôle et des régions cibles couplées entre elles | 332        |
| 13.2.1.1 Implication de régions différentes des cibles : les régions de contrôle                                                                         | 332        |
| 13.2.1.2 Une pensée causale découplée ?                                                                                                                  | 332        |

|              |                                                                                                                                                 |     |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 13.2.1.2.1   | Régulation top-down : absence de rétroaction . . .                                                                                              | 332 |
| 13.2.1.2.2   | Mesures de connectivité effective : rétroactions négligées . . . . .                                                                            | 333 |
| 13.2.1.2.3   | Un signal "inhibiteur" ? . . . . .                                                                                                              | 335 |
| 13.2.2       | Modélisation du contrôle inhibiteur dans la thèse . . . . .                                                                                     | 336 |
| 13.2.2.1     | Etude de l'effet sur la cible d'une augmentation du forçage sur le pool exciteur de la région de contrôle . . . . .                             | 336 |
| 13.2.2.1.1   | En perturbation . . . . .                                                                                                                       | 336 |
| 13.2.2.1.2   | En variation . . . . .                                                                                                                          | 338 |
| 13.2.2.2     | Etude de l'effet sur la région de contrôle d'une augmentation du forçage sur le pool exciteur de la région de contrôle . .                      | 339 |
| 13.2.2.2.1   | En perturbation . . . . .                                                                                                                       | 339 |
| 13.2.2.2.2   | En variation . . . . .                                                                                                                          | 339 |
| 13.2.2.3     | Conclusion : importance de préciser la variable d'observation pour caractériser l'effet exciteur ou inhibiteur d'un paramètre . . . . .         | 340 |
| 13.2.2.4     | Notions de <i>contrôle</i> et contrôle <i>inhibiteur</i> . . . . .                                                                              | 340 |
| 13.2.3       | Comment augmenter l'inhibition de l'activité d'une région cible ? . .                                                                           | 341 |
| 13.2.3.1     | Prediction sur l'effet d'une augmentation de $I_{stim}$ dans la région de contrôle . . . . .                                                    | 342 |
| 13.2.3.1.1   | Confirmation donnée par l'étude de sensibilité . . .                                                                                            | 342 |
| 13.2.3.1.2   | Hypothèse de linéarité . . . . .                                                                                                                | 344 |
| 13.2.3.2     | Prédiction sur l'effet d'une augmentation de $J_{gaba_c}$ dans la cible . . . . .                                                               | 345 |
| 13.2.3.2.1   | Sans rétroaction de la cible vers la région de contrôle                                                                                         | 345 |
| 13.2.3.2.1.1 | Raisonnement causal et usage de la métaphore de la balance E/I . . . . .                                                                        | 345 |
| 13.2.3.2.1.2 | Variations des points fixes : définitions formelles de l' <i>inhibition</i> et de la <i>sur-inhibition</i> .                                    | 348 |
| 13.2.3.2.2   | Avec rétroaction : cas de deux régions couplées en EI                                                                                           | 351 |
| 13.2.3.2.2.1 | Peut-on prévoir le comportement du modèle couplé avec la métaphore de la balance E/I ?                                                          | 351 |
| 13.2.3.2.2.2 | Effet d'une augmentation de $J_{gaba_c}$ dans la cible sur l'activité $sn_c^*$ de la cible stimulée .                                           | 356 |
| 13.2.3.2.2.3 | Variations des points fixes : définitions formelles de l' <i>inhibition</i> et de la <i>sur-inhibition</i> .                                    | 357 |
| 13.2.3.2.2.4 | Note sur la variation totale d'activité $\Delta sn_c^*$ : deux chemins possibles pour la construire . .                                         | 359 |
| 13.2.3.2.3   | Prendre en compte les mesures entre participants .                                                                                              | 360 |
| 13.2.3.2.3.1 | Contrôle inhibiteur : $J_{gaba_c}$ fixe pour un participant . . . . .                                                                           | 360 |
| 13.2.3.2.3.2 | Exemple avec deux participants . . . . .                                                                                                        | 360 |
| 13.2.3.2.4   | Effet de $J_{gaba_c}$ sur la variation d'activité de la cible $\Delta sn_{c/B_{Er_c}}$ due à une stimulation de la région de contrôle . . . . . | 362 |

|              |                                                                                                                                      |            |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------|------------|
| 13.2.3.2.4.1 | Définir un effet sur-inhibiteur ou sous-inhibiteur<br>infinitésimal de $J_{gaba_c}$ . . . . .                                        | 362        |
| 13.2.3.2.4.2 | Un résultat de la thèse : la courbe en "U" . . . . .                                                                                 | 363        |
| 13.2.3.2.4.3 | Traduction mathématique et explorations de<br>l'origine du profil en "U" . . . . .                                                   | 364        |
| 13.2.3.2.5   | Synthèse : comment traduire formellement l'augmen-<br>tation de l'inhibition de la cible en fonction de la<br>question ? . . . . .   | 366        |
| 13.2.4       | Bilan . . . . .                                                                                                                      | 369        |
| <b>14</b>    | <b>Perspectives</b> . . . . .                                                                                                        | <b>371</b> |
| 14.1         | Modèle direct . . . . .                                                                                                              | 371        |
| 14.1.1       | Extension à plus de régions . . . . .                                                                                                | 371        |
| 14.1.2       | Explorer d'autres architectures de connectivité . . . . .                                                                            | 371        |
| 14.1.3       | Modéliser d'autres conditions expérimentales . . . . .                                                                               | 372        |
| 14.1.3.1     | Think . . . . .                                                                                                                      | 372        |
| 14.1.3.2     | No Think intrusive . . . . .                                                                                                         | 373        |
| 14.1.3.3     | Les pistes avec DCM . . . . .                                                                                                        | 374        |
| 14.1.4       | Un modèle neural dynamique ? . . . . .                                                                                               | 374        |
| 14.1.5       | Quel forçage en entrée du modèle hémodynamique ? . . . . .                                                                           | 374        |
| 14.1.6       | Prendre en compte des régions d'une autre nature que les régions de<br>contrôle et les régions cibles : les régions relais . . . . . | 375        |
| 14.1.7       | Augmenter l'inhibition dans un contexte de contrôle excitateur . . . . .                                                             | 376        |
| 14.1.7.1     | Contrôle . . . . .                                                                                                                   | 376        |
| 14.1.7.2     | Contrôle Excitateur . . . . .                                                                                                        | 377        |
| 14.1.7.3     | Définition et traduction du concept de moindre excitation<br>dans le modèle . . . . .                                                | 377        |
| 14.1.7.3.1   | Cas de la variation d'un paramètre $\pi_E$ . . . . .                                                                                 | 377        |
| 14.1.7.3.2   | Cas de la variation de deux paramètres $\pi_E$ et $\pi_I$ . . . . .                                                                  | 378        |
| 14.1.7.3.2.1 | Activité de la cible $sn_c^*$ . . . . .                                                                                              | 378        |
| 14.1.7.3.2.2 | Variation d'activité de la cible due à la varia-<br>tion de $\pi_E : \Delta sn_c / \pi_E$ . . . . .                                  | 380        |
| 14.1.7.4     | Que signifie augmenter l'inhibition dans le contexte d'un contrôle<br>excitateur ? . . . . .                                         | 381        |
| 14.1.7.5     | Comment le travail dans la thèse permet déjà d'explorer cette<br>question ? . . . . .                                                | 381        |
| 14.1.7.5.1   | Un modèle du contrôle excitateur . . . . .                                                                                           | 381        |
| 14.1.7.5.2   | Effet d'une augmentation de $J_{gaba_C}$ sur $\Delta sn_c / B_{Erc}$<br>dans le cas d'un contrôle excitateur . . . . .               | 381        |
| 14.2         | Inversion . . . . .                                                                                                                  | 382        |
| 14.2.1       | Perspective d'application de la méthode en l'état actuel . . . . .                                                                   | 382        |
| 14.2.1.1     | Etude de pathologies psychiatriques comme le Trouble de<br>Stress Post-Traumatique . . . . .                                         | 382        |
| 14.2.1.1.1   | Enjeux cliniques de l'étude du contrôle inhibiteur . . . . .                                                                         | 382        |

|            |                                                                                                  |            |
|------------|--------------------------------------------------------------------------------------------------|------------|
| 14.2.1.1.2 | Symptômes intrusifs dans le Trouble de Stress Post-Traumatique . . . . .                         | 382        |
| 14.2.1.1.3 | Liens avec le travail présenté dans la thèse . . . . .                                           | 383        |
| 14.2.1.2   | D'autres types de données . . . . .                                                              | 383        |
| 14.2.2     | Première étape de la procédure d'inversion : variabilité de la réponse hémodynamique ? . . . . . | 384        |
| 14.2.3     | Deuxième étape de la procédure d'inversion . . . . .                                             | 385        |
| 14.2.3.1   | Interprétation des paramètres ? . . . . .                                                        | 385        |
| 14.2.3.2   | Propagation d'incertitudes . . . . .                                                             | 385        |
| 14.2.3.3   | Sources d'incertitudes ? . . . . .                                                               | 386        |
| <b>VI</b>  | <b>Bibliographie</b>                                                                             | <b>387</b> |
| <b>VII</b> | <b>Annexes</b>                                                                                   | <b>399</b> |
| <b>15</b>  | <b>Annexes de la partie I : Introduction</b>                                                     | <b>401</b> |
| 15.1       | Etude du contrôle inhibiteur des intrusions mnésiques par contrastes . . . . .                   | 401        |
| 15.1.1     | Dans les régions de contrôle . . . . .                                                           | 401        |
| 15.1.1.1   | Entre Think et No Think . . . . .                                                                | 401        |
| 15.1.1.2   | Entre No Think non intrusif et No Think intrusif . . . . .                                       | 402        |
| 15.1.2     | Dans les régions cibles . . . . .                                                                | 403        |
| 15.1.2.1   | Entre Think et No Think . . . . .                                                                | 403        |
| 15.1.2.2   | Think par rapport à baseline . . . . .                                                           | 403        |
| 15.1.2.3   | No Think par rapport à baseline . . . . .                                                        | 403        |
| 15.1.2.4   | Entre No Think non intrusif, No Think intrusif et baseline . . . . .                             | 404        |
| 15.1.3     | Corrélations d'activité entre régions de contrôle et régions cibles . . . . .                    | 405        |
| 15.1.4     | Limites de l'utilisation des contrastes pour l'étude du contrôle inhibiteur . . . . .            | 405        |
| <b>16</b>  | <b>Annexes de la partie II : Choix de modélisation</b>                                           | <b>407</b> |
| 16.1       | Le modèle neural à une région isolée . . . . .                                                   | 407        |
| 16.1.1     | Evolution des activités synaptiques . . . . .                                                    | 407        |
| 16.1.2     | Analyse linéaire au voisinage du Point Fixe. . . . .                                             | 408        |
| 16.1.3     | Détermination numérique des temps caractéristiques . . . . .                                     | 409        |
| 16.1.4     | Estimation des temps caractéristiques par voie analytique . . . . .                              | 409        |
| 16.1.5     | Paramètres du modèle influant sur ces temps . . . . .                                            | 410        |
| <b>17</b>  | <b>Annexes de la partie IV : Inversion</b>                                                       | <b>413</b> |
| 17.1       | Première étape . . . . .                                                                         | 413        |
| 17.1.1     | Loi Gamma de paramètres $\alpha$ et $\beta$ . . . . .                                            | 413        |
| 17.1.2     | Densité de probabilité d'un paramètre hémodynamique non transformé $sn$ . . . . .                | 413        |
| 17.1.3     | Densité de probabilité d'une variable hémodynamique non transformée . . . . .                    | 416        |



|            |                                                                                                   |     |
|------------|---------------------------------------------------------------------------------------------------|-----|
| 17.1.4     | Calcul des jacobienues du modèle différentiel hémodynamique transformé . . . . .                  | 418 |
| 17.1.4.1   | Modèle hémodynamique transformé . . . . .                                                         | 418 |
| 17.1.4.2   | Calculs des dérivées partielles par rapport aux variables d'état                                  | 419 |
| 17.1.4.2.1 | Dérivées partielles par rapport à $z_i$ . . . . .                                                 | 419 |
| 17.1.4.2.2 | Dérivées partielles par rapport à $\tilde{f}_i$ . . . . .                                         | 420 |
| 17.1.4.2.3 | Dérivées partielles par rapport à $\tilde{v}_i$ . . . . .                                         | 421 |
| 17.1.4.2.4 | Dérivées partielles par rapport à $\tilde{q}_i$ . . . . .                                         | 421 |
| 17.1.4.3   | Calculs des dérivées partielles par rapport aux paramètres à inverser . . . . .                   | 421 |
| 17.1.4.3.1 | Dérivées partielles par rapport à $\tilde{s}n_{B_i}$ . . . . .                                    | 421 |
| 17.1.4.3.2 | Dérivées partielles par rapport à $\tilde{s}n_{stim_i}$ . . . . .                                 | 422 |
| 17.1.4.4   | Jacobienues à coder dans la toolbox VBA . . . . .                                                 | 422 |
| 17.1.5     | Calcul des jacobienues du modèle algébrique BOLD transformé . . . . .                             | 423 |
| 17.1.5.1   | Modèle BOLD transformé . . . . .                                                                  | 423 |
| 17.1.5.2   | Calculs des dérivées partielles par rapport aux variables d'état                                  | 423 |
| 17.1.5.3   | Jacobienne à coder dans la toolbox VBA . . . . .                                                  | 424 |
| 17.1.6     | Sorties d'inversion des $sn$ sur simulation . . . . .                                             | 424 |
| 17.1.6.1   | Cas où la condition initiale est contenue dans l'a-priori . . . . .                               | 424 |
| 17.1.6.1.1 | 2 paramètres transformés . . . . .                                                                | 425 |
| 17.1.6.1.2 | 3 paramètres transformés . . . . .                                                                | 426 |
| 17.1.6.1.3 | 4 paramètres transformés . . . . .                                                                | 427 |
| 17.1.6.2   | Cas où la condition initiale est éloignée de l'a-priori . . . . .                                 | 428 |
| 17.1.6.2.1 | 2 paramètres transformés . . . . .                                                                | 428 |
| 17.1.6.2.2 | 3 paramètres transformés . . . . .                                                                | 429 |
| 17.1.6.2.3 | 4 paramètres transformés . . . . .                                                                | 430 |
| 17.2       | Deuxième étape de la procédure d'inversion . . . . .                                              | 431 |
| 17.2.1     | Exploration des points fixes pour un pool inhibiteur . . . . .                                    | 431 |
| 17.2.1.1   | Recherche de point(s) fixe(s) . . . . .                                                           | 431 |
| 17.2.1.2   | Existence d'un unique point fixe . . . . .                                                        | 431 |
| 17.2.1.3   | Etude de la stabilité . . . . .                                                                   | 432 |
| 17.2.2     | Propriétés des fonctions $hn$ et $hg$ qui donnent les fréquences de potentiels d'action . . . . . | 433 |
| 17.2.2.1   | Etude du signe . . . . .                                                                          | 433 |
| 17.2.2.2   | Etude des variations . . . . .                                                                    | 433 |
| 17.2.2.3   | Fréquences de potentiels d'action pour les pools exciteur et inhibiteur . . . . .                 | 435 |

## 18 Annexes de la partie V : Discussion 437

|          |                                                                        |     |
|----------|------------------------------------------------------------------------|-----|
| 18.1     | Balance E/I : une région isolée . . . . .                              | 437 |
| 18.1.1   | Analyse de sensibilité : perturbation des paramètres dans les courants | 437 |
| 18.1.1.1 | Formulation mathématique du modèle pour une région isolée              | 437 |
| 18.1.1.2 | Analyse de sensibilité . . . . .                                       | 438 |
| 18.1.2   | Variation simultanée de deux paramètres . . . . .                      | 441 |
| 18.1.2.1 | Construction analytique de la variation d'activité totale . . . . .    | 441 |

|            |                                                                                                                          |     |
|------------|--------------------------------------------------------------------------------------------------------------------------|-----|
| 18.1.2.1.1 | Première procédure . . . . .                                                                                             | 441 |
| 18.1.2.1.2 | Deuxième procédure : . . . . .                                                                                           | 442 |
| 18.1.2.1.3 | Résumé . . . . .                                                                                                         | 442 |
| 18.1.2.2   | Limites de la balance E/I pour penser l'effet de la variation<br>de plusieurs paramètres . . . . .                       | 443 |
| 18.2       | Deux régions : Effet d'une augmentation du forçage dans la région de contrôle                                            | 444 |
| 18.2.1     | Expressions analytiques des sensibilités par rapport à une perturbation<br>du forçage de la région de contrôle . . . . . | 444 |
| 18.2.1.1   | Avec rétroaction . . . . .                                                                                               | 444 |
| 18.2.1.2   | Sans la rétroaction de la cible vers la région de contrôle . . .                                                         | 445 |
| 18.2.2     | Variation du forçage sur le pool exciteur de la région de contrôle de<br>$\Delta B_{Erc} = I_{stim}$ . . . . .           | 446 |
| 18.2.2.1   | Effet sur la région cible . . . . .                                                                                      | 446 |
| 18.2.2.1.1 | Cas de la connectivité EI . . . . .                                                                                      | 446 |
| 18.2.2.1.2 | Sans la rétroaction de la cible vers la région de contrôle<br>. . . . .                                                  | 447 |
| 18.2.2.2   | Effet sur la région de contrôle . . . . .                                                                                | 448 |
| 18.2.2.2.1 | Cas de la connectivité EI . . . . .                                                                                      | 449 |
| 18.2.2.2.2 | Sans la rétroaction de la cible vers la région de contrôle<br>. . . . .                                                  | 450 |
| 18.2.2.3   | Hypothèse de linéarité de la variation d'activité de la cible .                                                          | 451 |
| 18.2.2.3.1 | Conditions de vérification de la linéarité . . . . .                                                                     | 451 |
| 18.2.2.3.2 | Condition non validée . . . . .                                                                                          | 451 |
| 18.3       | Deux régions couplées : perturbation des $J_{gaba}$ . . . . .                                                            | 452 |
| 18.3.1     | Formulation en $x_{inter}$ et $x_{intra}$ . . . . .                                                                      | 452 |
| 18.3.2     | Elimination des variables intermédiaires . . . . .                                                                       | 453 |
| 18.3.3     | Expression des sensibilités par rapport aux couplages $J_{gaba}$ . . . . .                                               | 454 |
| 18.3.3.1   | Modèle perturbé et expression canonique . . . . .                                                                        | 454 |
| 18.3.3.2   | Sensibilités découplées par rapport à $J_{gaba_1}$ et $J_{gaba_2}$ . . . . .                                             | 455 |
| 18.3.3.3   | En utilisant les sensibilités de régions découplées par rapport<br>aux forçages sur les pools excitateurs . . . . .      | 456 |
| 18.3.3.4   | Sensibilités dans le système couplé en fonction des sensibilités<br>découplées . . . . .                                 | 458 |
| 18.3.4     | Sensibilités dans un système contrôle-cible . . . . .                                                                    | 460 |
| 18.3.4.1   | Extraction des cas d'intérêt . . . . .                                                                                   | 460 |
| 18.3.4.2   | Ecriture générique de la connectivité . . . . .                                                                          | 461 |
| 18.3.4.3   | Perturbation de $J_{gaba_c}$ uniquement . . . . .                                                                        | 461 |
| 18.3.4.3.1 | Dans le système couplé . . . . .                                                                                         | 461 |
| 18.3.4.3.2 | Cas de la connectivité EI . . . . .                                                                                      | 463 |
| 18.3.4.3.3 | Sans la rétroaction de la cible vers la région de contrôle                                                               | 464 |
| 18.3.4.4   | Perturbation de $J_{gaba_{rc}}$ uniquement . . . . .                                                                     | 465 |
| 18.3.4.4.1 | Dans le système couplé . . . . .                                                                                         | 465 |
| 18.3.4.4.2 | Cas de la connectivité EI . . . . .                                                                                      | 466 |
| 18.3.4.4.3 | Sans la rétroaction de la cible vers la région de contrôle                                                               | 466 |

|                                                                                                                          |     |
|--------------------------------------------------------------------------------------------------------------------------|-----|
| 18.4 Deux régions : Effet d'une augmentation de $I_{stim}$ et $J_{gaba_c}$ sur l'activité $sn_c^*$ de la cible . . . . . | 467 |
| 18.4.1 En perturbation . . . . .                                                                                         | 467 |
| 18.4.2 En variation . . . . .                                                                                            | 467 |
| 18.4.2.1 Première procédure : . . . . .                                                                                  | 467 |
| 18.4.2.2 Deuxième procédure : . . . . .                                                                                  | 468 |
| 18.4.2.3 Conclusion . . . . .                                                                                            | 469 |
| 18.4.2.4 Une autre méthode pour construire $\Delta sn_{c/B_{Er_c}}$ en fonction de $J_{gaba_c}$ . . . . .                | 470 |

## Table des abréviations

- IRMf : Imagerie par Résonance Magnétique fonctionnelle
- MRS : Magnetic Resonance Spectroscopy
- DTI : Diffusion Tensor Imaging
- TNT : paradigme Think/No Think
- T : condition expérimentale Think
- NT : condition expérimentale No Think
- DLPFC : dorso-lateral prefrontal cortex
- VLPFC : ventro-lateral prefrontal cortex
- SFG : superior frontal gyrus
- MFG : middle frontal gyrus
- IFG : inferior frontal gyrus
- CG : cingulate gyrus
- Hipp : hippocampus
- PhG : parahippocampal gyrus
- PCun : precuneus
- AMPA :  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid
- NMDA : N-Methyl-D-aspartic acid
- GABA :  $\gamma$ -amino butyric acid
- balance E/I : balance Excitation/Inhibition
- PA : potentiel d'action
- PLT : Potentialisation à long terme
- BOLD : Blood Oxygen Level Dependent
- HRF : Hemodynamic Response Function
- PPI : Psychophysiological Interactions
- DCM : Dynamic Causal Modeling
- pool E/I : pool Excitateur/Inhibiteur
- v.a. : variable aléatoire
- REMEMBER : Protocole de recherche « REsilience and Modification of brain control network following novEMBER 13 »
- DSM : Diagnostic and Statistical Manual of mental disorders
- TSPT : Trouble de Stress Post-Traumatique
- i.e. : du latin id est, qui signifie "c'est-à-dire"
- e.g. : du latin exempli gratia, qui signifie "par exemple"
- resp. : respectivement



# Chapitre 1

## Présentation de la thèse

*Je suis physicienne de formation et j'ai mené ma thèse dans un laboratoire de neuropsychologie et de neurosciences. Ces derniers deux ans et demi ont été rythmés par une dynamique de travail collectif qui s'est peu à peu construite entre mes directeurs de thèse Francis Eustache, Pierre Gagnepain (neuroscientifiques au NIMH<sup>1</sup>, spécialisés en neuropsychologie et neuroimagerie), Shailendra Segobin (ingénieur de recherche en neuroimagerie au NIMH), Jacques Gautrais (chargé de recherche au CRCA<sup>2</sup> spécialisé en neurosciences computationnelles), Jean-Louis Joly, Stéphane Blanco (physiciens du non linéaire au La-Place<sup>3</sup>) et Jean-Yves Grandpeix (physicien du non linéaire au LMD<sup>4</sup>). Ce manuscrit reflète cette dynamique et je présente une thèse interdisciplinaire, à l'interface de plusieurs communautés : neurosciences expérimentales (neuropsychologie, neurobiologie, neuroimagerie), neurosciences computationnelles et physique du non linéaire.*

*Je m'adresse à ces trois types de lectorat, et je porte une attention particulière à leur adresser un exposé compréhensible. En effet, pour favoriser le dialogue entre communautés, il est important que chaque communauté puisse saisir ce qui se raconte dans l'autre. Pour cela, la thèse est organisée en quatre parties, chacune contenant plusieurs chapitres. Le niveau de détails pourra parfois paraître superflu et le propos évident pour certains lecteurs et certaines lectrices qui pourront saisir le message principal en lisant rapidement. Ces chapitres s'adressent alors davantage au lectorat appartenant à une autre des trois communautés, qui n'est pas familier avec les notions présentées.*

*Par exemple, les chapitres de la partie d'introduction s'adressent notamment à des physiciens qui ne travaillent pas en neurosciences expérimentales, pour leur permettre de comprendre la signification et la justification de mon travail dans ce domaine. Les détails du chapitre 6 pourront intéresser davantage des personnes ne travaillant pas dans la modélisation physique. Ils présentent une démarche qui n'est pas encore beaucoup répandue en neurosciences expérimentales et qui consiste à faire le lien entre le signal BOLD obtenu en Imagerie par Résonance Magnétique fonctionnelle et le signal neural sous-jacent. Au début de chaque chapitre, j'essaierai de guider le lecteur à travers les quelques lignes d'un chapeau (écrit en italique comme le présent premier chapeau).*

- 
1. Neuropsychologie et Imagerie de la Mémoire Humaine (UMR-S 1077), Université Caen Normandie
  2. Centre de Recherches sur la Cognition Animale (UMR 5169), Université Toulouse III - Paul Sabatier
  3. Laboratoire Plasma et Conversion d'Énergie (UMR 5213), Université Toulouse III - Paul Sabatier
  4. Laboratoire de Météorologie Dynamique (UMR 8539), Université Paris Sciences et Lettres

*Par le présent document, j'expose le contenu qui s'est construit à l'interface de plusieurs domaines et je l'écris dans l'idée d'alimenter la dynamique de travail collectif entre ces communautés.*

## 1.1 Contexte

### Vers l'étude des mécanismes neuraux du contrôle inhibiteur (chapitre 2)

En psychologie, le *contrôle inhibiteur* (ou *inhibition*) est un mécanisme cognitif qui permet de stopper une réponse automatique, prédominante ou naturelle mais non adaptée pour la réalisation d'un but désiré. Cette réponse peut être de nature motrice, émotionnelle ou cognitive (attention portée aux entrées sensorielles, pensée, souvenir ou intrusions mnésiques). Elle constitue la cible du contrôle inhibiteur.

Au niveau cérébral, le contrôle inhibiteur est associé au fonctionnement en réseau de différentes régions cérébrales, dont l'activité est notamment mesurée avec l'Imagerie par Résonance Magnétique fonctionnelle. Des régions du cortex préfrontal abaissent l'activité BOLD de régions cérébrales particulières, appelées régions cibles ([Apšvalka et al., 2022]). Celles-ci diffèrent en fonction du type de réponse que cible le contrôle inhibiteur (motrice, cognitive ou émotionnelle). Par exemple, pendant l'inhibition d'une réponse motrice (tâche de Go/No Go), les régions motrices subissent une baisse d'activité BOLD sous l'effet de l'activation de régions de contrôle préfrontales. Pendant le contrôle inhibiteur d'une intrusion mnésique (qui consiste à stopper un événement mental associé à un stimulus, e.g. visualisation mentale d'une image associée à un mot dans le cadre d'une tâche de Think/No Think), l'hippocampe droit subit une baisse d'activité BOLD sous l'effet de l'activation de régions de contrôle préfrontales.

Une variabilité interindividuelle des capacités de contrôle inhibiteur, notamment de contrôle inhibiteur des intrusions mnésiques, est mesurée aux niveaux comportemental et cérébral. Au niveau comportemental, certains sujets ont ainsi plus de difficultés à bloquer l'intrusion d'une image associée à un mot dans le cadre du paradigme Think/No Think, et ceci est associé à un moindre oubli de cette image lors de performances de rappel réalisées après la tâche. Au niveau cérébral, ceci est en lien avec une moindre baisse de l'activité BOLD de l'hippocampe par les régions de contrôle préfrontales.

Bien que mesurés de manière indirecte avec l'IRMf, les échanges d'activité entre régions sont d'abord des échanges d'activité entre neurones. Pour améliorer notre compréhension des mécanismes cérébraux qui sous-tendent la variabilité interindividuelle du contrôle inhibiteur, il semble important de préciser au niveau neural comment un groupe de régions de contrôle préfrontales peut inhiber l'activité de régions cibles.

### Source de variabilité interindividuelle du contrôle inhibiteur à l'échelle neurale (chapitre 3)

Les biologistes ont étudié le fonctionnement de neurones de deux types, excitateur ou inhibiteur, qui se distinguent par l'effet produit sur un neurone post-synaptique (respectivement une augmentation ou une baisse d'activité). La communication entre régions cérébrales passerait par des neurones excitateurs uniquement. Au contraire, des neurones excitateurs et inhibiteurs seraient présents localement dans les régions cérébrales.

[Schmitz et al., 2017] montrent que des sujets avec une moindre concentration de neurotransmetteurs inhibiteurs, réalisant une tâche de Think/No Think (TNT), présentent une moindre baisse de l'activité BOLD de l'hippocampe par les régions de contrôle préfrontales. Cela souligne l'importance du poids relatif des neurones excitateurs et inhibiteurs localement présents dans l'hippocampe pour l'abaissement efficace de son activité par les régions de contrôle. Nous utilisons la métaphore de la balance Excitation/Inhibition (E/I) locale pour nous représenter ce poids relatif et les effets compensatoires locaux entre variations du poids excitateur ou du poids inhibiteur. L'état de la balance E/I de l'hippocampe semble ainsi présenter un effet modulateur sur la baisse d'activité de celui-ci en réponse à l'augmentation d'activité des régions de contrôle.

Plus généralement, la variabilité interindividuelle de l'état des balances E/I locales dans les régions cibles et/ou dans les régions de contrôle pourrait expliquer la variabilité interindividuelle des capacités de contrôle inhibiteur. Il apparaît alors nécessaire de pouvoir caractériser les propriétés neurales (excitatrice et inhibitrice) locales au sein de régions cérébrales impliquées dans une tâche de contrôle inhibiteur. L'utilisation de l'IRMf permet de mesurer l'activité cérébrale de manière non invasive pendant une tâche de contrôle inhibiteur, mais présente des limites en tant que mesure indirecte de l'activité neurale.

### Nature des données obtenues en IRM fonctionnelle et méthodes d'analyses standard de la communication entre régions cérébrales (chapitres 4 et 5)

L'activité neurale ne peut en effet pas être mesurée de manière directe avec l'une des techniques d'imagerie couramment utilisée chez l'homme. Par exemple, l'Imagerie par Résonance Magnétique fonctionnelle (ou IRMf) est une méthode non invasive qui ne fournit qu'une mesure des variations du niveau d'oxygénation des tissus (signal BOLD : Blood Oxygen Level Dependent), en lien avec des variations hémodynamiques provoquées par l'évolution des besoins métaboliques des neurones sous-jacents.

Les méthodes d'analyse classiques de signaux BOLD se basent sur des modèles statistiques de la mesure qui prennent en compte ou non un modèle dynamique de l'activité cérébrale. Parmi les modèles dynamiques standards, certains modélisent l'activité neurale et les échanges entre régions, puis le couplage neurovasculaire et enfin la mesure BOLD (comme pour DCM<sup>5</sup>, [Friston et al., 2003]). L'activité neurale est pensée comme le résultat global de processus excitateurs et inhibiteurs non explicites et cela ne permet pas d'étudier le rôle des neurones inhibiteurs localement présents dans les régions. La caractérisation des mécanismes neuraux reste donc très limitée en raison de l'absence de modèle neural précis en IRMf.

---

5. Dynamic Causal Modeling



## La question que se pose la communauté/le positionnement de la thèse :

Comment inférer à partir de données d'IRMf des propriétés excitatrices et inhibitrices (E/I) neurales locales au sein de régions cérébrales impliquées dans une tâche de contrôle inhibiteur ?

## 1.2 Choix de modélisation

Pour étudier la balance neurale E/I à partir d'un signal BOLD, il est nécessaire de travailler à partir d'un modèle dynamique qui fournit de manière explicite les évolutions temporelles des activités neurales excitatrice et inhibitrice.

### Présentation du modèle direct (chapitre 6)

Nous partons d'un modèle biophysique de l'activité neurale proposé par [Naskar et al., 2021]. Il se base sur des hypothèses biologiques du fonctionnement des neurones (intégration de courants ioniques au niveau des dendrites, décharge de potentiels d'action, libération de neurotransmetteurs dans la fente post-synaptique, dynamiques de liaison/dissociation entre neurotransmetteurs et récepteurs). Des approximations en champ moyen permettent de modéliser l'activité synaptique moyenne de neurones excitateurs et inhibiteurs en interaction au sein de chaque région cérébrale, ainsi que les échanges d'activité excitatrice entre régions. Les variations de l'activité synaptique neurale produisent alors des changements hémodynamiques puis BOLD dans chaque région cérébrale. Ce modèle hiérarchique se traduit mathématiquement dans un système non linéaire d'équations différentielles couplées. La seule observable est la mesure BOLD de chaque région. Etant donné ces choix de modélisation, nous traduisons la question que pose la communauté en une première liste de questions.

### Questions de recherche préliminaires

1. Peut-on obtenir une baisse du signal BOLD dans la région cible quand l'activité BOLD de la région de contrôle augmente ?
2. Peut-on identifier des paramètres qui permettent de modifier l'état de la balance neurale locale E/I ?
3. Comment les propriétés E/I neurales locales<sup>6</sup> influencent-elles les variations de signaux BOLD des régions impliquées dans une tâche de contrôle inhibiteur, en particulier la baisse d'activité des régions cibles due à l'augmentation d'activité des régions de contrôle ?
4. Comment inférer les paramètres neuraux qui représentent l'état des balances E/I locales de chaque région à partir de signaux BOLD recueillis pendant une tâche de contrôle inhibiteur ?

---

6. ou balances E/I locales

5. Nous prenons comme cas d'étude le contrôle inhibiteur des intrusions mnésiques (testé avec le paradigme TNT). Est-il possible de présenter une preuve de concept du point 4, avec la mise en oeuvre numérique de la procédure d'inversion sur les données de participants ?

Pour préciser ces questions et pouvoir les traduire dans le modèle (i.e. en terme de variables et paramètres d'intérêt), nous en étudions les propriétés dynamiques.

## **Etude du passage entre les niveaux de description neurale et BOLD (chapitre 7)**

L'étude de certaines propriétés du modèle hiérarchique permet de répondre aux questions 1, 2 et 3 en se focalisant uniquement sur le modèle neural au régime stationnaire. Nous testerons l'effet de deux catégories de paramètres sur la baisse d'activité de régions cibles. Nous prendrons le temps d'exposer complètement, dans l'infinie précision de chaque détail, comment nous identifions, pour le modèle que nous nous donnerons à l'échelle neurale, les paramètres qui traduiront notre question.

Nous montrons également que nous pouvons décomposer la méthode d'inversion en 2 étapes pour traiter de manière séparée l'inversion du modèle BOLD puis l'inversion du modèle neural.

### **Précision des questions de recherche : objectifs de la thèse**

Les objectifs de la thèse portent alors sur :

- A) Comment connecter les régions au niveau neural pour obtenir une baisse de l'activité neurale excitatrice au régime stationnaire dans la région cible, quand l'activité neurale excitatrice au régime stationnaire dans la région de contrôle augmente ?
- B) Pour une architecture de connectivité fixée, comment la variation du forçage dans les régions de contrôle et l'auto-inhibition dans les régions cibles influencent-elles les variations des activités neurales excitatrices au régime stationnaire des régions impliquées dans une tâche de contrôle inhibiteur ; en particulier la baisse d'activité des régions cibles due à l'augmentation d'activité des régions de contrôle ?
- C) Comment l'analyse du modèle direct éclaire-t-elle le problème de l'inversion et nous permet de construire une procédure d'inversion (en 2 étapes) pour inférer les paramètres neuraux de chaque région à partir de signaux BOLD recueillis pendant une tâche de contrôle inhibiteur (quel choix de modèle direct pertinent, sur quelle observable porte l'inversion, comment construire une procédure d'inversion etc. ) ?
- D) La mise en oeuvre numérique du point C) nous permet d'estimer les paramètres neuraux à partir de mesures BOLD pendant une tâche de contrôle inhibiteur. Nous prenons comme cas d'étude le contrôle inhibiteur des intrusions mnésiques (testé avec le paradigme TNT). Est-il possible de présenter une preuve de concept de l'inversion sur les données d'un sujet (un participant non exposé de l'étude REMEMBER) ?

## 1.3 Organisation du travail de thèse

La thèse s’organise en deux parties principales :

### 1.3.1 Première partie : Analyse du modèle direct (chapitre 9)

— Pour la question A) :

Nous proposons d’abord un (nouveau) modèle neural du contrôle inhibiteur. Cette proposition se base sur une extension de l’architecture de connectivité inter-régions proposée par la communauté ([Naskar et al., 2021]). L’analyse du comportement de ce modèle avec la mise en œuvre des outils d’analyses des systèmes non linéaires (régime stationnaire, temps caractéristiques, étude des sensibilités, gains de rétroaction, analyse par découpage-raccordement, emboîtement) amène à proposer une architecture de connectivité différente de celle du modèle de départ pour pouvoir rendre compte des mécanismes d’inhibition (i.e. une baisse de l’activité neurale et BOLD de régions cibles suite à une augmentation de l’activité de régions de contrôle). Dans le modèle original, l’activité des régions cibles ne peut qu’augmenter.

— Pour la question B) :

Nous étudions l’influence des propriétés E/I locales sur l’effet inhibiteur de régions de contrôle sur des régions cibles pendant une tâche de contrôle inhibiteur (i.e. l’amplitude de la baisse d’activité des régions cibles). Dans un système composé de deux régions, nous mettons en évidence une relation non linéaire entre la baisse d’activité dans la région cible et le paramètre inhibiteur de celle-ci. Il existe une valeur de ce paramètre inhibiteur au delà de laquelle plus on augmente l’inhibition dans la cible et moins on abaisse l’activité de la cible. Ce résultat contre-intuitif est important à prendre en considération dans la perspective de donner du sens au rôle des paramètres du modèle.

L’ensemble du travail à deux régions est contenu dans l’article présenté à la partie III. Grâce à ce travail analytique, nous pouvons fixer une architecture de connectivité qui permet l’abaissement de l’activité (neurale et BOLD) des régions cibles quand les régions de contrôle sont stimulées. Nous nous assurons également de la possibilité de pouvoir inverser sur des décours temporels du signal BOLD.

### 1.3.2 Deuxième partie : Estimation de paramètres (chapitres 10 et 11)

Nous cherchons à estimer les paramètres neuraux de chaque région à partir des décours temporels des signaux BOLD recueillis pendant une tâche de contrôle inhibiteur.

— Pour la question C) :

Nous présentons la procédure d’inversion en deux étapes (formalisation de l’inversion et présentation des outils numériques pour sa mise en œuvre) :

1. Inversion des signaux BOLD vers les points fixes neuraux correspondant aux états non stimulés/ stimulés. La méthode d’inversion est une méthode bayésienne variationnelle. L’implémentation numérique de cette étape se fait en adaptant les codes de la toolbox VBA (matlab).

2. Inversion des paramètres neuraux à partir de ces jeux de points fixes par résolution algébrique et numérique (python)

La fiabilité de la procédure est d'abord vérifiée à partir de simulations.

- Pour la question [D](#)) : Preuve de concept

Nous prenons comme cas d'étude le contrôle inhibiteur des intrusions mnésiques. Nous utilisons les données d'un participant non exposé de l'étude REMEMBER, obtenues en IRMf pendant une tâche de Think/NoThink. Nous menons la procédure d'inversion précédemment décrite en sélectionnant 15 régions cérébrales impliquées dans la tâche (11 régions de contrôle et 4 cibles). Nous obtenons une estimation des paramètres neuraux dans chaque région.

## 1.4 Perspectives d'application de la thèse : étude du contrôle inhibiteur dans le trouble de stress post-traumatique

Les propriétés E/I locales influencent l'effet inhibiteur de régions de contrôle sur des régions cibles pendant une tâche de contrôle inhibiteur (i.e. l'amplitude de la baisse d'activité des régions cibles). La perspective d'application de la thèse pourra être de chercher à savoir si des modifications de ces propriétés E/I peuvent être en lien avec un dysfonctionnement du contrôle inhibiteur de la mémoire. Pour cela, il sera nécessaire de poursuivre le travail d'inversion sur les données de tous les participants de l'étude REMEMBER, puis de conduire des analyses statistiques pour comparer les valeurs des paramètres estimés entre les groupes. Afin de pouvoir proposer une interprétation clinique des résultats en sortie d'inversion sur tous les participants, la relation non linéaire entre la baisse d'activité des cibles et les paramètres dans les cibles est une question importante à explorer. Nous l'avons fait dans un système composé de 2 régions uniquement, avec 2 paramètres libres. Il serait nécessaire de le poursuivre dans un système composé de plus de régions (ici 15 régions) et donc avec plus de paramètres libres.



# Première partie

## Introduction

Dans le chapitre 2, je définis le concept de contrôle inhibiteur à l'échelle des processus cognitifs et à l'échelle de l'activité de régions cérébrales, principalement mesurée avec l'Imagerie par Résonance Magnétique fonctionnelle (IRMf). Bien que mesurés de manière indirecte avec l'IRMf, les échanges d'activité entre régions sont d'abord des échanges d'activité entre neurones. Le chapitre 3 présente le fonctionnement d'un neurone, son excitation et son inhibition dans le corpus de la neurophysiologie, ainsi que le concept de balance Excitation/Inhibition et son importance pour l'étude des mécanismes neuraux du contrôle inhibiteur. Le chapitre 4 permet de mieux comprendre la nature du signal BOLD mesuré en IRMf, qui reflète de manière très indirecte l'activité des neurones contenus dans une région cérébrale. Au chapitre 5, j'exposerai un état de l'art des principaux modèles et méthodes d'analyse de signaux BOLD pour l'étude des mécanismes cérébraux du contrôle inhibiteur. Nous verrons que l'absence de modèle neural précis en IRMf rend la caractérisation des mécanismes neuraux du contrôle inhibiteur très limitée, et notamment l'étude des balances Excitation/Inhibition au sein des régions cérébrales impliquées.

Tout au long de cette première partie, je prendrai comme cas d'exposition le contrôle inhibiteur des intrusions mnésiques, particulièrement étudié dans le laboratoire de Neuropsychologie et Imagerie de la Mémoire Humaine où j'ai travaillé.

## Chapitre 2

# Le contrôle inhibiteur, un processus psychologique adaptatif associé au fonctionnement de réseaux cérébraux qui impliquent le cortex préfrontal et des régions cibles

*Par souci de clarté pédagogique d'abord, je fais le choix de séparer dans ce chapitre l'étude du contrôle inhibiteur en psychologie et en neuroimagerie.*

*En tant que physicienne se formant à de nouveaux corpus de connaissances, j'ai eu besoin de prendre le temps de séparer les résultats obtenus à partir de données comportementales<sup>1</sup>, de ceux obtenus à partir de données d'Imagerie par Résonance Magnétique fonctionnelle. Je présenterai également au fur et à mesure les liens mesurés entre les données comportementales issues du premier corpus et les données d'imagerie issues du second .*

*Dans un cadre épistémologique réductionniste, la fonction cognitive émerge de l'activation de différentes régions cérébrales et de leurs interactions dans le temps. Ce discours donne une origine physiologique des comportements et des cognitions, et a émergé pour l'étude de l'inhibition à partir des années 1870 avec les travaux de Wundt en psychologie physiologique et de Ferrier en neuropsychologie ([Bari and Robbins, 2013]). Un lien causal est alors formulé entre les données d'imagerie et les données comportementales<sup>2</sup> .*

*Cependant, [Bari and Robbins, 2013] rapportent qu'il n'est pas si facile de relier l'interaction inhibitrice entre différentes régions cérébrales au concept d'inhibition au niveau comportemental<sup>3</sup>. Il conseille de distinguer les différents niveaux d'analyse.*

*En tant que physicienne de formation, je n'ai pas la prétention de pouvoir défendre un lien*

---

1. performances de rappel, temps de réaction, reports phénoménologique d'intrusions etc.

2. par exemple dans [Levy and Anderson, 2012] : "The nature of phenomenal awareness itself remains unsolved (Chalmers, 1995), but the current linkage of first-person experience to the regulation of neural activity suggests that the physical basis by which people voluntarily stop this experience may ultimately be witnessed and understood"

3. [Bari and Robbins, 2013] : "it is not easy to relate synaptic inhibition to inhibitory interactions between different brain areas or to the concept of inhibition at the behavioral level (Ursin, 1976)"



*causal ou non entre ces corpus. Pour écrire de manière plus sereine, je fais donc le choix d'une présentation basée sur la séparation des niveaux d'analyse et je rapporte les liens mis en lumière entre les différents types de mesures.*

## 2.1 Différents niveaux d'analyse

Le concept d'*inhibition* apparaît dans les travaux scientifiques à partir du début du 19e siècle ([Bari and Robbins, 2013]). Il a été employé par différentes communautés (psychologie expérimentale, psychiatrie, biologie, psychanalyse, neurologie, neurosciences etc. ), qui s'intéressent à des niveaux de descriptions et d'analyses différents. En fonction du contexte dans lequel il est employé, il a ainsi des significations très différentes.

[Pilkington and McKELLAR, 1960] recommandent de l'accompagner d'un adjectif pour préciser le niveau d'analyse auquel on souhaite se rapporter. Nous distinguerons pour la suite trois échelles d'analyses pour parler d'inhibition :

1. l'échelle des processus cognitifs (contrôle inhibiteur d'une action, d'une pensée ou d'une émotion par exemple), échelle que je désignerai par le qualificatif "neuropsychologique",
2. l'échelle des régions cérébrales (baisse de l'activité de régions cérébrales due aux variations d'activité d'autres régions cérébrales, toutes connectées dans un réseau), échelle que je dénoterai "cérébrale",
3. l'échelle des neurones (transmission synaptique inhibitrice entre deux neurones), que je qualifierai de "neurophysiologique".

Nous préciserons les notions d'inhibition à l'échelle "neuropsychologique" et à l'échelle "cérébrale" dans ce chapitre. L'inhibition à l'échelle "neurophysiologique" sera traitée dans le chapitre 3.

Dans chaque cas, nous décrirons plus en détails le contrôle inhibiteur des intrusions mnésiques.

## 2.2 Contrôle inhibiteur en psychologie

En neuropsychologie, l'*inhibition* (ou *contrôle inhibiteur*) est un mécanisme cognitif qui permet de stopper une réponse automatique, prédominante ou naturelle. Il nous permet de ne pas agir uniquement selon nos habitudes, impulsions ou fausses croyances ([Diamond, 2013]). Le contrôle inhibiteur fait partie des principales fonctions exécutives, qui assurent un avantage adaptatif à l'individu en lui permettant d'atteindre un but désiré dans des situations complexes et non routinières. La réponse à inhiber peut être de nature différente (motrice, émotionnelle ou cognitive).

## 2.2.1 Le contrôle inhibiteur : une des principales fonctions exécutives

### 2.2.1.1 Définition des fonctions exécutives

L'étude des fonctions exécutives a commencé dès la fin du 19e siècle avec les travaux des neuropsychologues sur les fonctions du lobe frontal ([Seron, 2009]). Ces travaux regroupent principalement des observations anatomocliniques de patients cérébrolésés après un accident (ex : le patient Phinéas Gage de Harlow en 1868, des blessés de la 1ère guerre mondiale par Luria) ou une chirurgie (ex : résections de tumeurs pour le patient A de Brickner et KM de Hebb ; lobotomies frontales comme traitement de troubles psychiatriques à partir de 1936) et des expériences de lésions frontales sur animaux (chiens, chats, singes...).

Les fonctions exécutives regroupent un ensemble de processus cognitifs qui, selon Lezak ([Lezak, 1982], [Lezak, 1994]), permettent de réguler les conduites cognitives (ex : verbales, mnésiques, perceptives...), comportementales (ex : régulation des mouvements) et sociales. Elles permettent de s'adapter de manière intentionnelle aux situations non routinières (inhabituelles, conflictuelles ou complexes) pour lesquelles les schémas d'actions et de pensées automatiques ne sont pas adaptés, afin d'atteindre un but et de résoudre des problèmes.

Les *fonctions exécutives* peuvent aussi être désignées par *contrôle exécutif* ou *contrôle cognitif* ([Diamond, 2013]).

### 2.2.1.2 Principales fonctions exécutives

Selon le modèle de Luria (1966-1973), repris par Lézac (1982,1994), toute activité complexe se décompose en 4 étapes principales : 1) analyse de la situation et formulation d'un but 2) élaboration d'un programme ordonnant les différentes étapes 3) exécution séquentielle et organisée de ce plan 4) comparaison continue entre le résultat de l'action réalisée et les attentes du plan pour d'éventuelles corrections.

Les opérations cognitives en jeu pendant ces étapes peuvent regrouper de manière non exhaustive : la sélection de stratégies, la génération d'hypothèses, la planification, l'initiation de l'action, le maintien de l'attention, l'inhibition des réponses dominantes, la flexibilité, la prise de décision, le rétrocontrôle des effets de l'action, la déduction de règles etc. [Godefroy, 2009] y inclut également les stratégies d'encodage et de récupération en mémoire épisodique, la sélection de souvenirs pertinents, les processus de changement de perspectives (ego vs exocentrée) dans la théorie de l'esprit et les aptitudes métacognitives.

Il n'existe actuellement pas de définition consensuelle ni de liste exhaustive des fonctions exécutives ([Broto, 2020]).

### 2.2.1.3 Contrôle inhibiteur : un des principaux processus exécutifs

[Miyake et al., 2000] proposent un découpage des fonctions exécutives par processus cognitifs. Il distingue trois fonctions exécutives principales non indépendantes : la mémoire de travail, la flexibilité mentale et le contrôle inhibiteur. Elles permettraient l'émergence de fonctions exécutives plus complexes comme la planification, la résolution de problème, le raisonnement etc. ([Diamond, 2013]).

La *mémoire de travail* permet de garder en tête des informations pertinentes et d'effectuer

sur elles une manipulation mentale (autrement dit, travailler à partir d’une information qui n’est plus accessible à nos sens). Cette deuxième caractéristique est importante et c’est ce qui la distingue de la mémoire à court-terme. La mémoire de travail existe sous deux types en fonction de la nature des informations manipulées : la mémoire de travail verbale et la mémoire de travail non verbale ou visuo-spatiale ([Diamond, 2013]).

La *flexibilité mentale* permet le passage volontaire d’une opération cognitive à une autre ou d’une tâche à une autre, lors de l’exécution d’un plan ou d’un changement de stratégie mentale par exemple.

Le *contrôle inhibiteur* (ou *inhibition*) permet de bloquer/stopper une réponse automatique, prédominante/prépondérante ou naturelle (comme une impulsion) mais non adaptée pour la réalisation du but désiré. Cette réponse peut être de nature motrice (ie un déplacement physique du corps), cognitive (attention portée aux entrées sensorielles, pensée, souvenir ou intrusion mnésique) ou émotionnelle. Elle constitue la cible du contrôle inhibiteur. Les capacités de contrôle inhibiteur se développent pendant l’enfance et l’adolescence, pour décliner durant le vieillissement normal ([Diamond, 2013]).

## 2.2.2 Différentes formes d’inhibition et tâches associées

Le contrôle inhibiteur peut être impliqué dans une variété de situations et il n’existe pas de liste exhaustive ni de classification consensuelle des différentes formes d’inhibition (e.g. [Bari and Robbins, 2013], [Diamond, 2013]).

Nous faisons le choix d’une présentation basée sur la nature de la réponse cible à inhiber. Nous distinguerons ainsi trois dimensions :

- la dimension de l’agir : *inhibition motrice* (actions)
- la dimension du ressentir : *inhibition émotionnelle*
- la dimension du penser : *inhibition cognitive* (attention, pensées, souvenirs, intrusions)

Ces différentes formes d’inhibition peuvent fonctionner ensemble pour conduire à l’inhibition d’un comportement non adapté et au contraire à l’observation d’un comportement orienté vers un but. Par comportement, nous entendons ici tout mouvement du corps qui est observable. Nous n’utiliserons pas le terme d’inhibition comportementale (comme [Bari and Robbins, 2013], qui l’oppose à inhibition cognitive) pour éviter toute confusion entre ce qui relèverait directement de l’inhibition comportementale et ce qui la permettrait plus indirectement.

Nous décrirons pour chaque forme d’inhibition quelques tâches utilisées en laboratoire pour les tester (on parle de paradigmes expérimentaux), bien que certaines tâches puissent être discutées quant au manque de spécificité de la forme d’inhibition testée ou à l’implication d’autres processus exécutifs. Cette dimension sort du cadre de cette thèse et nous nous contenterons alors d’une présentation de tâches standards dans une catégorie où la forme d’inhibition testée peut au moins y être impliquée. Nous pourrions aussi donner quelques exemples en vie quotidienne.

### 2.2.2.1 Inhibition d’une réponse motrice

La suppression d’une réponse prépondérante se manifeste souvent sur le plan moteur. Nous pouvons inhiber une action automatique (e.g. une habitude ou un réflexe) ou planifiée. Elle peut être non commencée ou en cours d’exécution.

Dans la vie quotidienne, un piéton s’apprêtant à traverser une route doit par exemple interrompre sa marche quand le feu du passage piéton passe au rouge. [Anderson et al., 2016] donnent également l’exemple d’un pot de plantes tombant du rebord d’une fenêtre, que la personne doit éviter de rattraper lorsqu’elle s’aperçoit qu’il s’agit d’un cactus.

Le mouvement ou son absence est un comportement facilement observable et objectivement mesurable en laboratoire. Les expériences d’inhibition de la réponse motrice se basent généralement sur les mouvements des membres et sur des pressions de boutons, même si n’importe quelle action peut être étudiée. La performance et les temps de réaction sont facilement mesurables ([Bari and Robbins, 2013]). Différents types de tâches sont utilisées pour inhiber une action et ne rien faire à la place. Nous pouvons citer les tâches de :

- Go/No Go :

un stimulus "go" indique de faire une action (ex : 2 points sur un écran indiquent de taper avec deux doigts dessus), un stimulus "no go" indique de retenir/ne pas faire l’action (ex : 1 point sur l’écran indique de ne pas taper dessus). Ce test mesure l’inhibition d’une action planifiée.

- Stop-Signal Task (SST) :

Des stimuli visuels ou sonores sont utilisés comme signaux "stop" après le début de la réponse motrice pour indiquer au participant d’arrêter son action déjà commencée.

Le concept d’inhibition de la réponse motrice a mis moins de temps à s’imposer que celui d’inhibition émotionnelle ou cognitive ([Bari and Robbins, 2013]), dû au fait que son observation et sa mesure en laboratoire sont plus directes.

### 2.2.2.2 Inhibition émotionnelle

Nous pouvons d’abord inhiber une émotion pour éviter sur l’instant l’inconfort de son ressenti (cas d’une émotion négative comme la tristesse, la colère, la peur, la honte, etc.). Cela peut aussi permettre de contrôler une action, lorsqu’il s’agit de résister aux tentations de récompenses immédiates ou de ne pas agir de manière impulsive et prématurée (cf définition du self-control de [Diamond, 2013]).

Inhiber le désir de récompenses immédiates peut avoir lieu dans différents contextes. Quand au moins deux désirs entrent en compétition, il est possible d’accéder à une plus grande récompense après un délai ou un effort (e.g. aller jusqu’au bout d’une tâche longue et complexe malgré l’envie d’abandonner pour une activité plus simple ou plus intéressante mais qui, à plus long terme, nous sera bénéfique, comme courir un marathon, rédiger une dissertation, etc.). Une expérience assez connue a consisté à présenter à de jeunes enfants une friandise dans une salle de test en laboratoire. L’expérimentateur s’absentait quelques

minutes. Pendant ce temps, l'enfant pouvait manger la friandise, ou non et alors en recevoir davantage au retour de l'expérimentateur ([Eigsti et al., 2006]). Le contrôle inhibiteur se développant tout au long de l'enfance et de l'adolescence, les réactions (filmées) de certains enfants montrent à quel point il leur était compliqué de contrôler leur gourmandise.

Une récompense immédiate peut aussi conduire à des conséquences négatives plus tard (e.g. consommation de substances psychoactives comme drogues ou alcool). Inhiber le désir de la récompense immédiate a alors pour objectif de les éviter dans le futur. Dans le cas où la personne ne parvient pas à inhiber sa tentation, [Bechara et al., 1994] utilisent l'expression de "myopia for the future".

Enfin, nous pouvons inhiber une émotion afin de ne pas agir de manière impulsive et prématurée. Cela rend possible la capacité d'attendre, en prenant le temps de la collecte d'informations pertinentes et de l'analyse pour répondre et ne pas réagir. Lors d'un conflit par exemple, le contrôle de la colère aide à ne pas réagir de manière physiquement violente. Pour éviter un raisonnement biaisé, il est parfois nécessaire de contrôler l'envie de répondre rapidement en donnant la première réponse à laquelle on pense.

### 2.2.2.3 Inhibition cognitive

#### 2.2.2.3.1 Contrôle inhibiteur de l'attention

Un stimulus saillant (e.g. bruit fort, objet en mouvement) attire notre attention de manière automatique, que nous le voulions ou non. La plupart du temps, ce processus permet de s'adapter efficacement à l'environnement.

Néanmoins, nos ressources attentionnelles étant limitées, lorsque le stimulus avantagé n'est pas pertinent pour la réalisation d'un but que nous nous sommes fixé, nous pouvons choisir d'inhiber volontairement notre attention de celui-ci (i.e. d'ignorer le stimulus) pour la focaliser sur d'autres stimuli pertinents ([Broto, 2020]). En psychologie cognitive, le stimulus qui a activé automatiquement une représentation non pertinente pour la tâche à résoudre est une interférence. Le contrôle inhibiteur de l'attention permet la suppression de l'interférence dans un contexte de compétition pour l'utilisation des ressources attentionnelles et facilite ainsi la réorientation du focus attentionnel ([Bari and Robbins, 2013]). On parle aussi d'*inhibition attentionnelle*, de *contrôle attentionnel*, d'*attention exécutive*, *sélective* ou *focalisée* ([Diamond, 2013]).

Par exemple, pendant une fête, le contrôle inhibiteur de l'attention permet de bloquer en dehors du champ attentionnel le bruit environnant pour se concentrer sur la voix de la personne avec laquelle nous discutons (effet "cocktail party").

En laboratoire, le test du Stoop permet de tester l'inhibition attentionnelle. Le participant doit donner le plus rapidement possible la couleur de l'encre (vert, bleu, rouge, jaune, noir) de plusieurs mots successifs. Ces mots sont des adjectifs qui désignent ces mêmes couleurs. Dans le cas congruent, le mot désigne la même couleur que celle de l'encre qui a servi à l'écrire. Dans le cas incongruent, le mot désigne une autre couleur que celle de l'encre. Si la lecture s'est automatisée, la signification du mot lu constitue alors une interférence qui entre en compétition avec la couleur de l'encre. Le participant doit inhiber cette information non

pertinente pour énoncer uniquement la couleur de l'encre et les temps de réponse mesurés sont plus longs par rapport au cas congruent([Broto, 2020], [Diamond, 2013]).

Les stimuli sur lesquels portent l'attention peuvent être externes ou internes (pensées, souvenirs, intrusions, émotions).

### 2.2.2.3.2 Contrôle inhibiteur de pensées

L'inhibition de pensées permet en premier lieu de stopper un schéma mental dominant ([Diamond, 2013]) et d'éviter dans ce contexte les raisonnements biaisés par des habitudes de penser ou des fausses croyances.

Nous présentons deux tâches utilisées de manière courante en laboratoire pour évaluer ce processus cognitif :

- Version cognitive du go/no-go ([Bari and Robbins, 2013]) :  
Les participants doivent compter silencieusement pendant un signal "go" et arrêter pendant un signal "no go". Dans une autre version, ils imaginent une réponse motrice pendant un signal "go" et son arrêt pendant le "no-go".
- Paradigmes d'apprentissage inversé (Reversal learning paradigms) comme le Test de classement de cartes du Wisconsin (WCST pour Wisconsin Card Sorting Test) :  
Le participant doit trouver le bon critère d'appariement de cartes parmi plusieurs possibilités (couleur, forme, nombre). Il procède par essais et l'expérimentateur indique à chaque tentative si l'appariement est correct ou non. Lorsque plusieurs essais consécutifs sont corrects, le participant a trouvé et a mémorisé la règle d'appariement. L'expérimentateur en change alors, sans avertir le participant. Celui-ci doit inhiber la règle précédemment apprise pour réussir à trouver et mémoriser la nouvelle règle qui sert à classer les cartes. Cette expérience se répète plusieurs fois de suite.

Le contrôle inhibiteur de pensées permet également de résister à des événements mentaux non désirés, comme par exemple des souvenirs, inquiétudes, ruminations, hallucinations, intrusions ([Schmitz et al., 2017]). Il a été montré que bloquer une intrusion participerait à son oubli, de manière active et volontaire ([Diamond, 2013], [Anderson and Hanslmayr, 2014]). On parle aussi de "intentional forgetting" en anglais.

## 2.2.3 Contrôle inhibiteur des intrusions mnésiques

### 2.2.3.1 Oubli adaptatif et suppression mnésique

L'oubli a longtemps été considéré comme une limite de la mémoire. De manière involontaire, les souvenirs de nos expériences passées s'estompent et disparaissent avec le temps qui passe ou l'interférence provoquée par l'encodage de nouvelles expériences similaires. Cependant, ces dernières années, de nouvelles recherches sur l'oubli montrent qu'il peut aussi résulter d'un mécanisme de contrôle inhibiteur de la mémoire, au stade de l'encodage ou de la récupération ([Anderson and Hanslmayr, 2014]). Celui-ci peut s'exercer de manière active

et adaptative ([Fawcett and Hulbert, 2020]) : on parle d'oubli motivé ou "motivated forgetting".

Le contrôle inhibiteur de la récupération (aussi appelé "retrieval suppression") permet de "repousser" et "bloquer" en dehors de la conscience les souvenirs ou intrusions mnésiques liées à des événements de vie négatifs. Il diffère d'une simple substitution de pensées ([Benoit and Anderson, 2012]). Son effet sur les contenus mnésiques est étudié scientifiquement avec le paradigme Think/No Think (TNT), proposé par [Anderson and Green, 2001], qui constitue un modèle expérimental du contrôle des intrusions. Les performances de rappel post TNT suggèrent une altération de la trace mnésique de l'intrusion lorsque celle-ci a été inhibée pendant la tâche (cf 2.2.3.3). Ainsi, le contrôle inhibiteur d'une intrusion est également désigné par l'expression "suppression mnésique".

### 2.2.3.2 Paradigme Think/No Think

Le paradigme Think/No Think (TNT) permet de mimer en laboratoire une intrusion mnésique, c'est-à-dire une représentation mentale automatique associée à un stimulus. Il se déroule en plusieurs étapes.

Le participant apprend d'abord par cœur des associations de mot/image<sup>4</sup>. Après cette phase d'apprentissage, la présentation du mot seul déclenche automatiquement la visualisation mentale de l'image associée.

Ensuite, pendant la deuxième phase dite de "Think/No Think", les mots sont présentés de manière successive soit en vert (condition Think) soit en rouge (condition No Think). En condition *Think*, le participant doit visualiser l'image associée au mot avec le plus de détails possibles. En condition *No Think*, il doit maintenir son attention sur le mot et dans le cas où l'image fait intrusion dans son esprit, tenter de chasser la représentation mentale en "vidant" son esprit, pour se reconcentrer sur le mot. Ce processus de blocage de l'intrusion diffère de la simple substitution de pensées ([Benoit and Anderson, 2012]). Après chaque mot rouge, le participant peut rapporter s'il a eu ou non une intrusion. Pour cela, il appuie dans l'IRM sur les boutons d'un boîtier qui correspondent à la fréquence des intrusions pendant la présentation du mot qui a précédé : "jamais", "brièvement", "souvent" ([Levy and Anderson, 2012]). Ceci permet de différencier deux autres conditions expérimentales : *No Think Intrusive* ou *No Think Non Intrusive*.

La troisième phase dite de Test consiste en des tâches de rappel pour mesurer la mémoire des images associées aux différents mots. Le participant doit tenter de se souvenir de l'image à partir de la présentation du mot (Same Probe), ou à partir d'un indice de catégorie et de la première lettre désignant le contenu de l'image (Independent Probe)<sup>5</sup>. Certaines études utilisent aussi une tâche de priming : le participant doit reconnaître le plus rapidement possible l'image bruitée, qui devient de moins en moins bruitée au cours du temps ([Anderson and Hanslmayr, 2014]).

---

4. mot/image : mot/objet, mot/scène, mot/visage. Il peut aussi s'agir d'association mot/mot ou image/image (visage-scène). cf [Anderson and Hanslmayr, 2014], [Anderson et al., 2016], [Schmitz et al., 2017]

5. pour une description plus détaillée, on pourra se référer à [Anderson and Green, 2001], "trained cue" et "independent cue" sont les équivalents de "same probe" et "independent probe" employés dans l'étude de [Anderson and Levy, 2009] pour laquelle nous rapportons les résultats à la figure 2.1



Parmi les mots dont les participants ont appris l'association avec une image, certains ne sont pas présentés pendant la phase de "Think/No Think". Ils permettent de tester la récupération de la trace mnésique en l'absence de toute manipulation expérimentale, ce qui constitue la condition Baseline.

### 2.2.3.3 Résultats comportementaux

Les résultats chez des sujets sains montrent que :

- les items No Think sont moins bien récupérés que les items Think (cf figure 2.1). Le contrôle inhibiteur empêche alors l'effet de consolidation mnésique liée au rappel.
- les items No Think sont moins bien rappelés que les items Baseline, dans la tâche de rappel comme illustré sur la figure 2.1, ainsi que de priming (e.g. [Gagnepain et al., 2014]).

Ces effets sont cumulatifs et augmentent avec le nombre de suppressions réalisées ([Anderson and Hanslmayr, 2014]).

De plus, la fréquence des intrusions diminue au cours de la tâche (cf figure 2.2). Plus la fréquence des intrusions a diminué rapidement au cours de la tâche et plus les participants ont oublié les images associés aux mots rouges (No Think), en comparaison des items de Baseline ([Levy and Anderson, 2012]).

[Anderson and Hanslmayr, 2014] définissent deux composantes importantes permettant la récupération mnésique : la possibilité d'accéder au contenu stocké (accessibilité), et la présence de ce contenu dans le système de mémoire (disponibilité). Ne pas réussir à se souvenir peut être dû à un déficit d'accès à la trace mnésique ou au fait que celle-ci ne soit plus disponible.

Les études sur le contrôle inhibiteur des intrusions mnésiques montrent que celui-ci permettrait l'oubli ("suppression induced forgetting") en agissant sur ces deux composantes. Le contrôle inhibiteur rendrait la trace des images sur lesquelles il s'exerce moins accessible (hypothèse testée avec la tâche de rappel Independent Probe, cf [Anderson and Green, 2001]) et moins disponible en la dégradant (hypothèse testée avec la tâche de priming)<sup>6</sup>. Cet affaiblissement de la trace mnésique permettrait une diminution de sa réactivation ultérieure (i.e. une diminution des intrusions).

Le paradigme Think/No-Think constituerait ainsi une méthode permettant de supprimer à long-terme les souvenirs intrusifs, incluant jusqu'à leurs caractéristiques visuelles les plus inconscientes ([Gagnepain et al., 2014]). Les capacités de contrôle de nos souvenirs et pensées joueraient un rôle adaptatif et prépondérant dans la préservation de la santé mentale des individus ([Cole et al., 2014]).

---

6. [Anderson and Green, 2001] explicitent d'autres hypothèses infirmées pour expliquer cette diminution du rappel des items No Think par rapport au Baseline. Elle ne résulte pas d'associations alternatives avec le mot et qui interfèrent ("associative interference"), ou d'un affaiblissement de l'association entre le mot et l'image ("unlearning"). La trace serait moins disponible et moins accessible



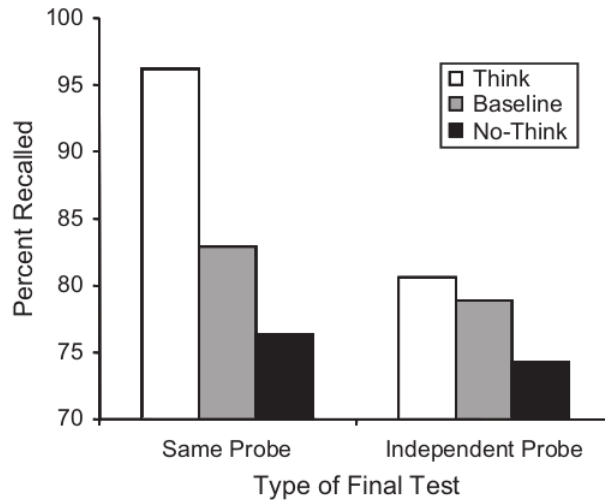


**Figure 2.1** – Performances de rappel dans le paradigme TNT (issu de [Anderson and Levy, 2009])

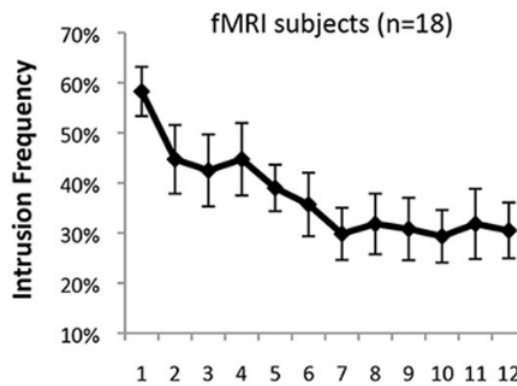


**Figure 2.2** – Diminution de la fréquence des intrusions au cours de la tâche TNT ([Levy and Anderson, 2012])

## 2.2.4 Contrôle inhibiteur proactif et réactif

Le contrôle inhibiteur pourrait être utilisé de deux manières différentes et complémentaires : proactif (ou prédictif) et réactif.

Le *contrôle proactif* agit avant que la réponse (motrice, cognitive ou émotionnelle) n'ait eu lieu pour pouvoir la prévenir. Le niveau de contrôle nécessaire à mettre en place dépendrait d'une estimation personnelle de la probabilité de réalisation de la réponse (ex : [Leone et al., 2022]). S'il est suffisant, la réponse ne se produit pas. Ce mécanisme nécessiterait une représentation soutenue des objectifs de la tâche en cours ([Bari and Robbins, 2013]). Si le contrôle proactif ne suffit pas, le *contrôle réactif* agit après le début de la réponse pour la stopper. Il s'agit donc d'un mécanisme de correction et il nécessiterait la réactivation transitoire des objectifs de la tâche en cours ([Bari and Robbins, 2013]).

Alors que le contrôle proactif pourrait être conscient ou non, le contrôle réactif s'exercerait de manière consciente. Un contrôle actif, conscient ou volontaire, demanderait un ef-

fort, contrairement à un processus passif, inconscient ou automatique ([[Diamond, 2013](#)], [[Bari and Robbins, 2013](#)]).

## 2.3 Contrôle inhibiteur à l'échelle des régions cérébrales

A partir de la seconde moitié du 19e siècle, les neurosciences posent la question du fonctionnement cérébral associé au fait de stopper une action ou un processus mental. Diverses approches et méthodes d'imagerie ont été utilisées pour mieux identifier les mécanismes cérébraux du contrôle inhibiteur.

### 2.3.1 Différentes méthodes pour étudier les mécanismes cérébraux du contrôle inhibiteur

Chez l'homme, les études ont porté sur des populations saines et des populations pathologiques (lésions cérébrales ou pathologies psychiatriques sans atteinte cérébrale) qui présentent des déficits cognitifs, notamment de contrôle inhibiteur. Aux premières méthodes d'observations anatomocliniques de patients cérébrolésés dont le cas princeps est Phineas Gage en 1868 (cf paragraphe 2.2.1.1), se sont rajoutées plus récemment différentes techniques d'imagerie des structures cérébrales et de leurs activations pendant une tâche de contrôle inhibiteur. Nous pouvons citer l'Imagerie par Résonance Magnétique (IRM), l'Imagerie du Tenseur de Diffusion (DTI), la Tomographie à Emission de Positons (TEP), la magnétoencéphalographie, l'électroencéphalographie, l'Imagerie par Résonance Magnétique fonctionnelle (IRMf). En plus des techniques précédentes, l'expérimentation animale utilise des méthodes plus invasives : lésions focales, stimulations électriques et enregistrement de l'activité neuronale locale à l'aide d'électrodes implantées par exemple ([[Bari and Robbins, 2013](#)]). Dans la suite de la thèse, nous nous intéresserons plus particulièrement à l'utilisation de l'IRMf chez l'homme pour étudier les mécanismes de contrôle inhibiteur. Le signal d'une région cérébrale mesuré en IRMf est appelé BOLD pour *Blood Oxygen Level Dependent*, il dépend du niveau d'oxygénation des tissus qui environnent les neurones de cette région cérébrale (chapitre 4). Il reflète de manière indirecte l'activité moyenne du grand nombre de neurones présents dans cette région, ou activité neurale.

### 2.3.2 Etude en Imagerie par Résonance Magnétique fonctionnelle

Avec l'approche localisationniste, on a cherché d'abord une région cérébrale précise dont la stimulation provoquerait systématiquement une inhibition. Cette hypothèse sera abandonnée au profit d'un fonctionnement en interaction de différentes régions cérébrales : des régions cérébrales sont connectées entre elles et fonctionnent ensemble à travers un réseau.

Les mesures effectuées en IRMf pendant une tâche de contrôle inhibiteur montrent que l'augmentation d'activité BOLD de régions du cortex préfrontal (dites *régions de contrôle*) conduit à un abaissement de l'activité de régions cérébrales particulières (dites *régions cibles*),

en passant par d'autres *régions relais*. Ces régions relais comprennent des ganglions de la base (noyau caudé droit et putamen), le thalamus et l'hypothalamus. Les mécanismes de contrôle inhibiteur fonctionnent de plus en interaction avec d'autres processus cognitifs ancillaires pendant la réalisation de la tâche (comme l'attention soutenue, la détection de conflit, la mémoire de travail<sup>7</sup> etc.). En fonction de la tâche, d'autres régions corticales et sous-corticales (e.g. cortex cingulaire antérieur ou des régions pariétales droites) associées à l'implication de ces processus ancillaires peuvent moduler l'effet des régions de contrôle sur les régions cibles ([Bari and Robbins, 2013], [Anderson et al., 2016]).

Le contrôle inhibiteur est un mécanisme dit "top-down" où des régions de contrôle de haut niveau dans le traitement de l'information agissent sur les régions cibles de plus bas niveau ([Anderson et al., 2016],[Schmitz et al., 2017]). Dans la suite, nous nous focaliserons sur les régions de contrôle et les régions cibles.

### 2.3.2.1 Régions de contrôle

Différentes formes d'inhibition partageraient des bases neurales similaires au niveau des régions de contrôle, dans le cortex préfrontal ([Munakata et al., 2011], [Depue et al., 2016], [Hunt et al., 2023]).

Le cortex préfrontal est divisé en quatre parties : les cortex préfrontal dorso-latéral, ventro-latéral, orbito-frontal et fronto-médian. Le cortex préfrontal dorso-latéral (DLPFC) et ventro-latéral (VLPFC) droit s'activent particulièrement pendant des tâches de contrôle inhibiteur moteur et mnésique ([Bari and Robbins, 2013], [Anderson et al., 2016], [Schmitz et al., 2017], [Apšvalka et al., 2022]). Le cortex orbito-frontal et fronto-médian s'activent particulièrement dans des tâches où il s'agit d'inhiber une composante émotionnelle ([Bonelli and Cummings, 2007]).

### 2.3.2.2 Régions cibles

L'augmentation d'activité (ou activation) de régions de contrôle préfrontales conduit à un abaissement de l'activité de régions cérébrales particulières, dites régions cibles. Celles-ci diffèrent en fonction du type de réponse que cible le contrôle inhibiteur (motrice, cognitive ou émotionnelle présentées à la section 2.2.2).

Par exemple, pendant l'inhibition d'une réponse motrice (tâche de go/no-go), les régions motrices M1 subissent une baisse d'activité suite à l'activation de régions de contrôle préfrontales ([Schmitz et al., 2017],[Apšvalka et al., 2022]). Pendant l'inhibition d'une intrusion mnésique (tâche de Think/No Think), l'hippocampe droit subit une baisse d'activité suite à l'activation de régions de contrôle préfrontales ([Levy and Anderson, 2012], [Benoit and Anderson, 2012]). Si les intrusions à inhiber présentent une valence émotionnelle négative, les régions de contrôle inhibent également l'amygdale ([Depue et al., 2007], [Gagnepain et al., 2017]).

Au contraire, l'activation des régions cibles est associée à la réalisation de la réponse en l'absence de contrôle inhibiteur. Ainsi, les régions motrices s'activent lors d'un mouvement, l'hippocampe est impliqué dans des processus de récupération en mémoire épisodique

---

7. online representation of task rules

([Squire, 1992],[Eldridge et al., 2000],[Eichenbaum et al., 2007]), et l'amygdale dans les processus émotionnels.

Au chapitre 5, nous présenterons plus en détails les résultats obtenus en IRMf pendant une tâche de Think/No Think chez des participants qui présentent une variabilité interindividuelle de leur capacité de contrôle inhibiteur des intrusions mnésiques. Nous verrons que cette variabilité interindividuelle comportementale est associée à une baisse plus ou moins importante de l'activité BOLD des régions cibles hippocampiques due à l'augmentation d'activité BOLD des régions de contrôle préfrontales.

### 2.3.2.3 Du signal BOLD à l'interprétation neurale ?

Améliorer notre compréhension des mécanismes neuraux qui sous-tendent le contrôle inhibiteur, notamment celui des intrusions mnésiques, d'abord dans le continuum sain puis dans des cas pathologiques (e.g. Trouble de Stress Post-Traumatique<sup>8</sup>), pourrait permettre à plus long terme d'améliorer la prévention de difficultés, le diagnostic de cas pathologiques et les traitements ([Bari and Robbins, 2013]).

Bien que le signal BOLD constitue une mesure indirecte de l'activité neurale sous-jacente (comme nous le verrons au chapitre 4), les échanges d'activité entre régions sont d'abord des échanges d'activité entre neurones, présents en très grands nombre dans les régions cérébrales<sup>9</sup>. Pour améliorer notre compréhension des mécanismes cérébraux du contrôle inhibiteur, il apparaît alors important de préciser au niveau neural 1) comment une région de contrôle peut abaisser l'activité neurale d'une région cible. 2) comment cet effet peut être modulé.

Dans le chapitre 3, nous descendrons à une échelle de description plus petite que la région cérébrale, pour présenter le fonctionnement d'un neurone, son inhibition et son excitation dans le corpus de la neurophysiologie. Le fonctionnement à l'échelle neurale sera par la suite décrit à partir d'éléments issus du fonctionnement d'un neurone à l'échelle individuelle.

Dans la thèse, le qualificatif *neural* fait référence à un niveau de description moyenné sur un grand nombre de neurones, contenus dans une région cérébrale. Il est différent du qualificatif *neuronal* qui désigne une description du fonctionnement individuel non moyenné de neurones (libération de neurotransmetteurs par un neurone pré-synaptique dans une fente synaptique, intégration des courants ioniques et décharge de potentiels d'action par le neurone post-synaptique...).

---

8. cf perspective d'application de la thèse

9. Des ordres de grandeurs sont donnés à la section 6.2



# Chapitre 3

## Neurophysiologie de l'inhibition

*Pour mieux comprendre les bases neurales des mécanismes cérébraux du contrôle inhibiteur, il est nécessaire de s'intéresser au fonctionnement des neurones et aux mécanismes biologiques qui permettent d'abaisser leur activité. Je présente donc quelques éléments de la neurophysiologie de l'inhibition, à l'échelle neuronale. Je décris brièvement à la section 3.1 quelques bases biologiques des transmissions synaptiques excitatrice et inhibitrice, qui permettent de définir le concept de balance Excitation/Inhibition (E/I) d'abord à l'échelle neuronale.*

*Je prendrai comme cas d'exposition le contrôle inhibiteur des intrusions mnésiques et à cet effet, je détaillerai à la section 3.2 deux types d'inhibition possibles : phasique et tonique. Dans la partie 3.3, nous verrons l'importance de l'inhibition tonique dans l'hippocampe pour des tâches qui dépendent de l'hippocampe, et présenterons la métaphore de la balance E/I à l'échelle neurale pour aider à penser les effets compensatoires locaux des propriétés excitatrice et inhibitrice dans une région cérébrale. Nous questionnons alors au 3.4 la variabilité interindividuelle du contrôle inhibiteur à partir de la variabilité interindividuelle de l'état des balances E/I en émission et en réception (i.e. dans les régions de contrôle et dans les régions cibles).*

### 3.1 Fonctionnement biologique d'un neurone

Le neurone constitue l'unité fonctionnelle de base dans le cerveau. C'est seulement au début du 20<sup>e</sup> siècle que le concept d'inhibition est largement accepté en neurophysiologie : en plus de neurones excitateurs, il existe des neurones inhibiteurs qui ont pour effet d'abaisser l'activité des neurones post-synaptiques ([Bari and Robbins, 2013]).

#### 3.1.1 Quelques éléments sur la physiologie des neurones et de leurs membranes

Un neurone est une cellule constituée d'un corps (ou soma, qui contient le noyau) et de prolongements cytoplasmiques plus ou moins longs, comme schématisé sur la figure 3.1. Ce type cellulaire est spécialisé dans le traitement et la transmission de fluctuations électriques. Chaque neurone reçoit par ses dendrites des fluctuations émises par d'autres neurones et,

après leur intégration au niveau du soma, émet en réponse, par son axone, des trains de fluctuations brèves et stéréotypées (les "potentiels d'action") vers d'autres neurones, ou d'autres types cellulaires (comme les muscles).

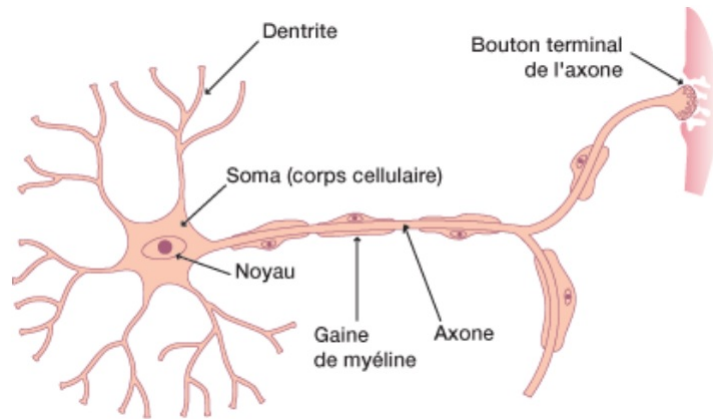


**Figure 3.1** – Schéma d'un neurone

#### Le "potentiel de membrane" au repos

Comme toutes les cellules de l'organisme, les neurones sont caractérisés par une différence de potentiel membranaire, du fait de la différence de composition chimique des milieux intra et extra cellulaires (cf figure 3.2). A l'extérieur de la cellule se trouvent principalement des ions sodium  $\text{Na}^+$  (un peu de  $\text{K}^+$  aussi) et chlorure  $\text{Cl}^-$ , avec un surplus de cations. L'intérieur de la cellule est surtout constitué d'ions potassium  $\text{K}^+$  (peu de  $\text{Na}^+$ ) et phosphate ( $\text{PO}_3^-$ ), avec un surplus d'anions. La différence de charges portée par les deux milieux crée une différence de potentiel au repos qui vaut environ  $-70 \text{ mV}$ . Par facilité de langage, cette différence de potentiel 'membranaire' entre les milieux intra- et extra-cellulaire est souvent dénommée simplement 'potentiel de membrane'. Nous garderons l'abréviation "ddpm" dans la suite, pour éviter toute confusion.

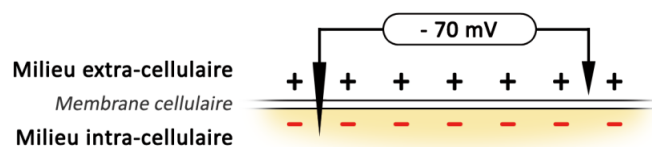


**Figure 3.2** – Schéma d'une membrane nerveuse au repos. Issu de [Corson, 2010]

### 3.1.2 Transmission chimique de l'influx nerveux : synapses excitatrice et inhibitrice

La fente synaptique désigne l'espace de jonction entre deux neurones consécutifs. Le neurone présynaptique se situe en amont de la fente et transmet, par son axone, le message

nerveux à un dendrite du neurone post-synaptique.

La transmission du message nerveux se fait grâce à la libération de neurotransmetteurs par le neurone présynaptique. Ceux-ci diffusent dans la fente synaptique et peuvent se fixer aux récepteurs membranaires du neurone post-synaptique (liaisons non covalentes, de type Van der Waals, liaisons hydrogènes ou électrostatiques). La fixation d'un neurotransmetteur sur un récepteur va permettre de modifier localement la perméabilité ionique de la membrane : un récepteur est couplé à un canal ionique<sup>1</sup> qui va laisser passer des ions du milieu extérieur vers le milieu intérieur et donc modifier la ddpm.

Il existe deux types de neurones présynaptiques, excitateur ou inhibiteur, nommés d'après les effets qu'ils produisent sur la ddpm du neurone post-synaptique. Ils diffèrent par le type de neurotransmetteurs qu'ils libèrent dans la fente synaptique (cf figure 3.3).

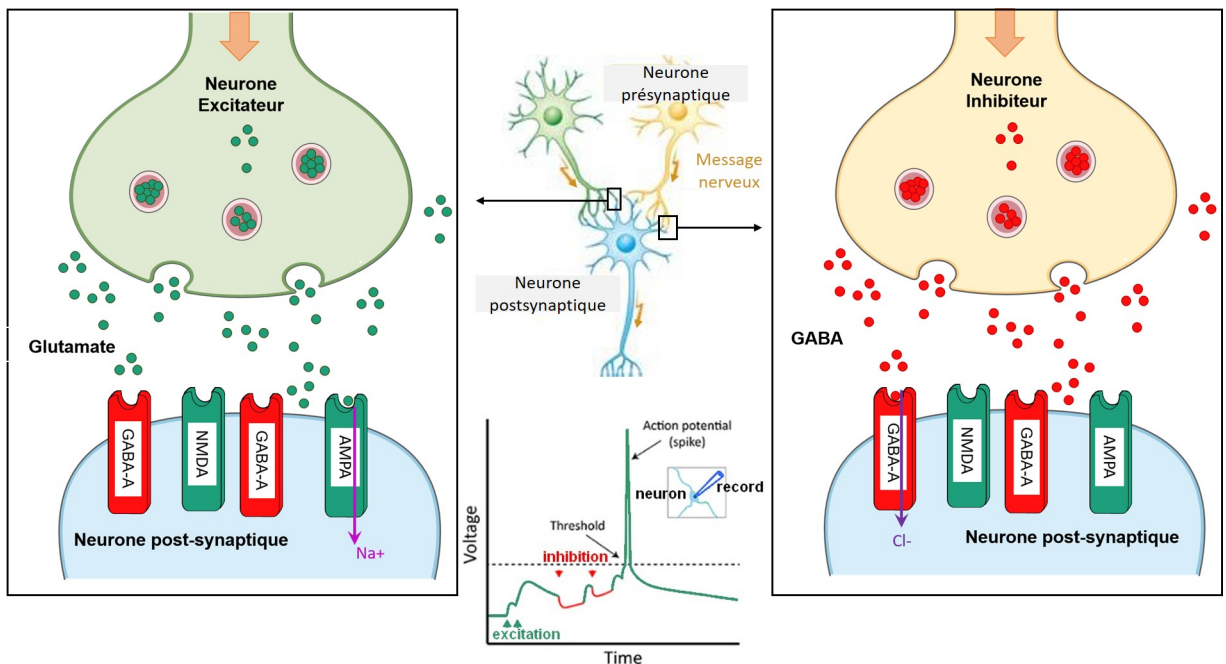


**Figure 3.3** – Principe de la transmission synaptique. Le neurone post-synaptique possède plusieurs types de récepteurs : NMDA, AMPA, GABA. Selon que le neurone présynaptique est excitateur ou inhibiteur, il va libérer dans la fente synaptique des neuro-transmetteurs différents, et activer seulement certains types de récepteurs, ce qui aura pour effet d'augmenter ou diminuer la différence de potentiel au niveau de la membrane du soma au cours du temps (au milieu en bas). Si celle-ci dépasse un seuil (Threshold), un potentiel d'action est généré, qui sera propagé par l'axone, et transmis aux cellules en aval (neurones ou autres).

1. On parle de "récepteur ionotrope".



### 3.1.2.1 Neurone excitateur : diminue la ddp<sub>m</sub> du neurone post-synaptique

Le glutamate est le neurotransmetteur libéré par un neurone excitateur. Il peut se fixer sur deux types de récepteurs : les récepteurs AMPA ( $\alpha$ -amino-3-hydroxy-5-méthyl-4-isoxazole propionique acid) et les récepteurs NMDA (N-Méthyl-D-aspartic acid).

#### 3.1.2.1.1 Récepteur AMPA

Ce récepteur est couplé à un canal ionique sodique, une protéine membranaire qui laisse spécifiquement passer les ions  $\text{Na}^+$ . L'entrée d'ions  $\text{Na}^+$  de l'extérieur vers l'intérieur de la cellule entraîne une diminution de la ddp<sub>m</sub> : on parle de dépolarisation membranaire.

#### 3.1.2.1.2 Récepteur NMDA

Ce récepteur est couplé à un canal calcique, bloqué par un ion  $\text{Mg}^{2+}$  tant que la différence de potentiel membranaire est inférieure à -50 mV environ. Si les récepteurs AMPA ont permis de dépolariser suffisamment la membrane, des ions  $\text{Ca}^{2+}$  entrent dans le neurone post-synaptique. Ceci a pour effet d'accélérer la dépolarisation membranaire due aux AMPA, et déclenche un ensemble de réactions cellulaires qui ont pour effet de rendre la synapse plus sensible au glutamate. L'augmentation d'amplitude de la réponse post-synaptique pour une même stimulation est appelée potentialisation à long terme (ou PLT). Elle s'opère à la suite d'un entraînement avec une intense activation présynaptique, généralement courte et de fréquence élevée. Les récepteurs NMDA, en rendant les synapses plus efficaces, assurent les phénomènes d'apprentissage.

### 3.1.2.2 Neurone inhibiteur : augmente la ddp<sub>m</sub> du neurone post-synaptique

Le neurotransmetteur libéré par un neurone inhibiteur est le GABA ( $\gamma$ -amino butyric acid). Il se fixe principalement sur les récepteurs GABA-A (que nous décrirons plus en détails à la section 3.2). Ce récepteur est couplé à un canal chlorure qui laisse passer des ions  $\text{Cl}^-$  dans le neurone post-synaptique. Ceci a pour effet d'augmenter la ddp<sub>m</sub> : on parle d'hyperpolarisation de la membrane.

### 3.1.2.3 Intégration des courants par le neurone post-synaptique

Un neurone reçoit continuellement des dépolarisations/hyperpolarisations de plusieurs synapses au cours du temps (cf figure 3.3, le nombre de synapses par neurone est estimé de l'ordre de  $10^4$  en moyenne dans le cerveau humain), et des mécanismes actifs (la "pompe  $\text{Na}/\text{K}$ ") tendent, en permanence, à faire tendre la ddp<sub>m</sub> vers sa valeur de repos. L'intégration temporelle de ces différents processus peut amener la ddp<sub>m</sub> du soma à dépasser une valeur seuil (de l'ordre de -50 mV) et conduire au déclenchement d'une réponse brève et stéréotypée (le potentiel d'action représenté à la figure 3.4), suivie d'un retour rapide au potentiel de repos.

Les synapses inhibitrices rendent donc plus difficile le déclenchement d'un potentiel d'action : davantage d'ions  $\text{Na}^+$  devront entrer dans le neurone pour atteindre le seuil de -50

mV, i.e. il faudra une excitation plus grande. Il existe donc au niveau synaptique des phénomènes de compensation entre courants excitateurs et inhibiteurs qui conduisent à l'activation des neurones. Nous pouvons parler d'une balance entre Excitation et Inhibition à l'échelle neuronale.

### 3.1.3 Transmission électrique de l'influx nerveux : les potentiels d'actions

#### 3.1.3.1 Naissance d'un potentiel d'action

Si la stimulation est assez forte et que la ddpm atteint le seuil, tous les canaux sodiques s'ouvrent. Avec l'entrée massive de  $\text{Na}^+$  permise alors par cette diffusion restreinte à cet ion, la charge cytoplasmique totale subit un changement brusque jusqu'à une ddpm qui devient positive (jusque 30-40 mV). L'intérieur de la cellule a désormais une charge supérieure à celle du milieu extérieur.

A ce transfert d'ions  $\text{Na}^+$  suit le passage différé d'ions  $\text{K}^+$  de l'intérieur vers l'extérieur de la cellule, par l'intermédiaire des canaux potassiques. La membrane tend alors à retrouver sa différence de potentiel au repos : on parle de repolarisation.

Cependant comme les canaux  $\text{K}^+$  réagissent lentement, ils restent ouverts en plus grand nombre qu'à l'état de repos, laissant sortir  $\text{K}^+$ , aussi la ddpm va passer sous le potentiel de repos (ie  $< -70$  mV). La membrane s'hyperpolarise donc au-delà du potentiel de repos le temps que tous les canaux se referment. La pompe Na/K rétablit alors les concentrations en ions  $\text{Na}^+$  à l'extérieur de la cellule et les ions  $\text{K}^+$  à l'intérieur et on retrouve la ddpm de repos. Un dérouls temporels typique d'un potentiel d'action est représenté à la figure 3.4.

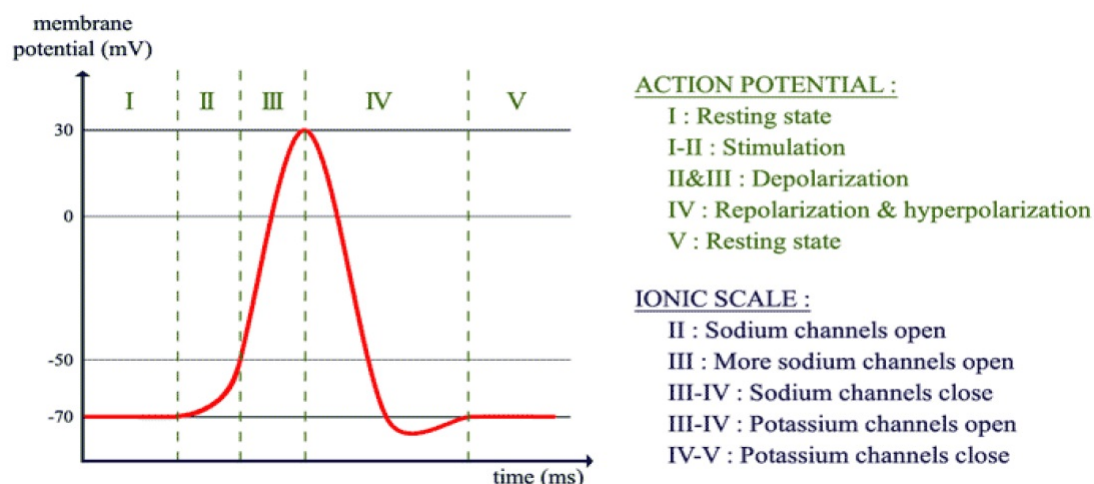


**Figure 3.4** – Dérouls temporels typique de la ddpm du soma au cours d'un potentiel d'action. L'échelle temporelle est de l'ordre de 2 ms. ([Corson, 2010])

### 3.1.3.2 Propagation du potentiel d'action par dépolarisation vers l'avant des régions voisines

Des mécanismes ioniques basés sur des différences de temporalités d'ouvertures de canaux ioniques assurent que le potentiel d'action se propage vers l'avant le long de l'axone et de ses ramifications. Sans entrer dans les détails, l'axone du neurone est constitué de zones de dépolarisation et de repolarisation successives, à l'origine de la propagation du signal nerveux (cf figure 3.5).

Si la ddp<sub>m</sub> atteint sa valeur seuil au niveau du soma, chacune de ces zones va assurer que le PA se propagera sur toute la longueur de l'axone, et de ses ramifications, sans perdre d'amplitude et ainsi propager la réponse du neurone considéré.

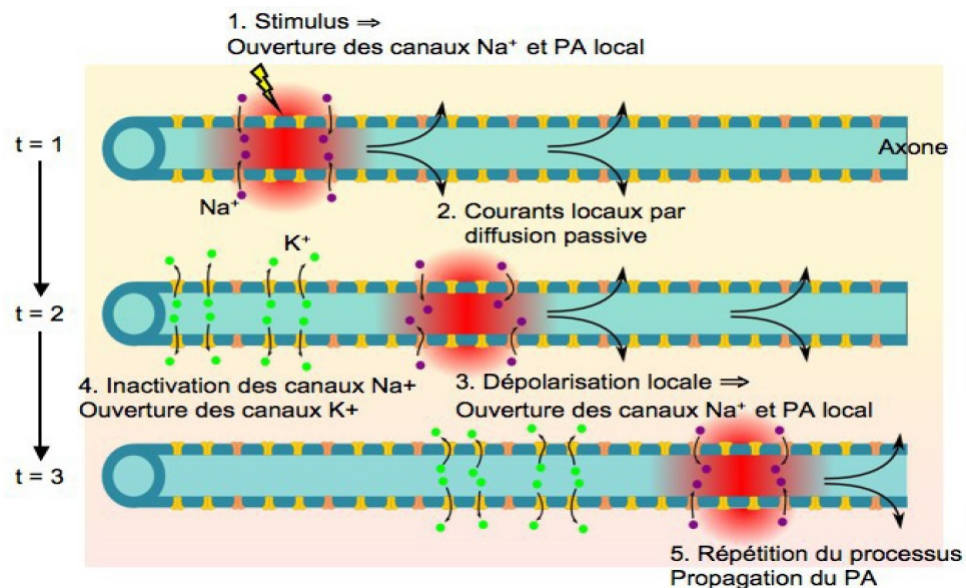


**Figure 3.5** – Schéma fonctionnel de propagation du potentiel d'action le long d'un axone (<https://sites.google.com/site/aphysionado/home>, bases neuronales)

## 3.2 Deux types d'inhibition GABAergique : phasique et tonique

### 3.2.1 Récepteurs GABA-A synaptiques et extra-synaptiques

Le GABA (de l'anglais **gamma-aminobutyric acid**) est le principal neurotransmetteur inhibiteur dans le système nerveux central des mammifères. Il se fixe sur deux types de récepteurs. Les moins abondants sont les récepteurs GABA-B, de type métabotrope, que nous ne traiterons pas. Les plus abondants sont les récepteurs GABA-A, de type ionotrope : la fixation du GABA au récepteur entraîne l'ouverture d'un canal ionique perméable aux ions  $\text{Cl}^-$  ([Farrant and Nusser, 2005],[Makkar et al., 2010]).

Un récepteurs GABA-A est formé de 5 sous-unités protéiques dont le repliement forme un canal ionique (cf figure 3.6). Il existe sept classes de sous-unités (avec des variants) :  $\alpha$  (1 à 6),  $\beta$  (1 à 3),  $\gamma$  (1 à 3),  $\rho$  (1 à 3),  $\delta$  et  $\epsilon$ . La plupart des récepteurs sont composés de deux sous-unités  $\alpha$ , deux  $\beta$  et une  $\gamma$ .

Le site de fixation du GABA se situe à la jonction entre les sous-unités  $\alpha$  et  $\beta$ . Les récepteurs GABA-A sont la cible de nombreuses molécules couramment utilisées en clinique, dont le site de fixation peut être différent de celui du GABA sur le récepteur. C'est par exemple le cas des benzodiazépines, des barbituriques ou de certains anesthésiques ([Collinson et al., 2002]) comme le propofol ou le chloroforme qui agissent comme modulateurs allostériques positifs. Leur fixation provoque un changement de conformation du récepteur qui devient plus sensible au GABA (augmentation de l'affinité du GABA au récepteur, augmentation de la probabilité d'ouverture du canal ionique en réponse à une liaison du GABA au récepteur) : si une molécule de GABA se fixe sur le récepteur, son effet est alors amplifié.

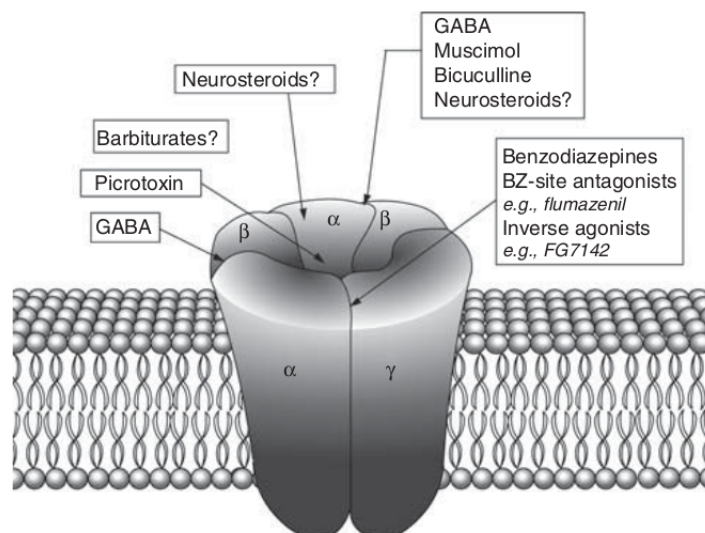


**Figure 3.6** – Schéma d'un récepteur GABA-A, constitué de sous-unités organisées entre elles  $\gamma - \beta - \alpha - \beta - \alpha$  (dans le sens antihoraire quand on regarde depuis la fente synaptique). Différents sites de fixation sont représentés, dont celui du GABA à la jonction entre les sous unités  $\alpha$  et  $\beta$  ([Makkar et al., 2010])

Les combinaisons possibles des sous-unités conduisent à une grande diversité de récepteurs aux propriétés biophysiques différentes. Les propriétés cinétiques des récepteurs peuvent varier avec des différences au niveau de l'affinité neurotransmetteur-récepteur ou des propriétés du canal ionique (cinétique entre changement d'état fermé/ouvert actif/ouvert inactif, variations temporelles et amplitude maximale de la conductance du canal). Ensuite, certaines sous-unités favorisent le regroupement des récepteurs entre eux. C'est par exemple le cas des sous-unités  $\gamma_2$ , contrairement à  $\alpha_5$ . Les récepteurs avec une sous-unité  $\gamma_2$  et sans  $\alpha_5$  se trouvent principalement regroupés dans la fente synaptique alors que ceux contenant du  $\alpha_5$  sans  $\gamma_2$  diffusent sur la surface du neurone pour se retrouver dans la zone extra-synaptique. On parle respectivement de récepteurs synaptiques ou extra-synaptiques. Les récepteurs GABA-A permettent alors deux types d'inhibition : une inhibition "phasique" et une inhibition "tonique" (pour une revue détaillée, voir [Farrant and Nusser, 2005]).

### 3.2.2 Inhibition phasique synaptique

Au niveau de la fente synaptique, le neurone présynaptique libère une concentration élevée de GABA (de l'ordre du millimol par litre) dont la concentration varie rapidement. Le GABA peut d'abord se fixer au niveau de récepteurs GABA-A synaptiques, sur la surface du neurone post-synaptique. Leur proximité au site de largage et la cinétique rapide des récepteurs entre changement d'état (fermé/ouvert actif/ouvert inactif) permet une transmission rapide (de l'ordre de la dizaine de millisecondes) et précise de l'influx nerveux, en phase avec les variations rapides de concentration de GABA. C'est l'inhibition phasique. L'inhibition phasique est principalement assurée par des récepteurs GABA-A synaptiques qui contiennent des sous-unités  $\alpha_1$ ,  $\alpha_2$  ou  $\alpha_3$  (de type  $\alpha_{1|2|3}\beta_{2|3}\gamma_2$ ). Ces sous-unités sont répandues dans tout le cerveau.

### 3.2.3 Inhibition tonique extrasynaptique

Les neurotransmetteurs non recapturés par les transporteurs (GAT1 et GAT3) du neurone pré-synaptique peuvent ensuite diffuser au delà de la fente pour atteindre l'espace extra-synaptique. La concentration de GABA y est beaucoup plus faible que dans la fente (dizaine de nanomoles à quelques micromoles par litre). Les récepteurs extra-synaptiques, plus éloignés du site de largage que les récepteurs synaptiques, sont aussi beaucoup plus dispersés (densité spatiale plus faible). Ils s'étendent sur les soma, les dendrites et les axones du neurone post-synaptique. Ils présentent une plus grande affinité au GABA que les récepteurs synaptiques et une cinétique plus lente de changement d'état. Ceci assure une transmission plus lente (de l'ordre de la dizaine de secondes) et plus bruitée de l'influx nerveux, asynchrone avec les stimulations pré-synaptiques. C'est l'inhibition tonique. Elle permet une hyperpolarisation membranaire sur des durées plus longues que l'inhibition phasique (plusieurs dizaines de secondes contre quelques dizaines de millisecondes). Elle s'étend ainsi sur une durée du même ordre de grandeur que celle pendant laquelle peut s'exercer le contrôle inhibiteur d'un souvenir ou d'une intrusion. L'inhibition tonique dépend de récepteurs GABA-A extra-synaptiques qui contiennent des sous-unités  $\alpha_4$ ,  $\alpha_5$  ou  $\alpha_6$  (de type  $\alpha_{4|6}\beta x\delta$  ou  $\alpha_5\beta x\gamma_2$ ).  $\alpha_5$  est présente uniquement dans l'hippocampe ([Collinson et al., 2002], [Crestani et al., 2002]).

## 3.3 Inhibition GABAergique tonique dans l'hippocampe

### 3.3.1 Effets d'une levée de l'inhibition tonique sur les apprentissages dépendant de l'hippocampe (modèles animaux)

Des récepteurs extra-synaptiques de l'hippocampe contiennent des sous-unités  $\alpha_5$ . Celles-ci sont localisées spécifiquement dans l'hippocampe, principalement dans les structures CA1 et CA3<sup>2</sup>, et absente ailleurs dans le cerveau de mammifères ([Collinson et al., 2002], [Crestani et al., 2002]).

Cette spécificité hippocampique permet aux biologistes de réduire expérimentalement l'inhibition tonique uniquement dans l'hippocampe de souris en supprimant totalement ou partiellement l'expression des sous-unités  $\alpha_5$  (par mutations génétiques). Les animaux sont alors soumis à des tâches d'apprentissages dépendant de l'hippocampe. L'étude de leurs performances permet donc de mieux comprendre le rôle de l'inhibition tonique hippocampique dans ce type de processus cognitif.

[Collinson et al., 2002] mettent en évidence qu'une levée de l'inhibition tonique dans l'hippocampe de souris est associée à de meilleures performances dans une tâche de mémoire spatiale, le test de la piscine de Morris. Les souris mutées trouvent la plateforme immergée un plus grand nombre de fois et plus rapidement que les souris non mutées (i.e. sans levée d'inhibition). Lors de tâches de conditionnement à la peur, un stimulus neutre (appelé stimulus conditionnel) est présenté à l'animal, par exemple une sonnerie, suivi d'un stimulus négatif (appelé stimulus inconditionnel), par exemple un choc électrique. Celui-ci déclenche une réaction physiologique inconditionnelle de peur (typiquement, un freezing ou immobilisation instantanée). La mémorisation de l'association temporelle entre les deux stimuli dépend de l'hippocampe et produit le changement de statut du stimulus neutre. Le retentissement de la sonnerie, même non suivi du choc électrique, produit la réaction physiologique de peur chez l'animal. [Crestani et al., 2002] et [Yee et al., 2004] montrent qu'une levée d'inhibition tonique dans l'hippocampe de souris mutées est associée à une meilleure mémorisation de l'association temporelle entre stimulus conditionnel et inconditionnel. Dans les minutes qui suivent l'application du stimulus conditionnel, les souris mutées présentent un freezing pendant plus longtemps.

Les processus d'apprentissage et de mémoire impliquent au niveau neuronal une augmentation de l'efficacité des synapses à la suite d'un entraînement spécifique et répété. Plus précisément, l'amplitude de la réponse post-synaptique se voit augmenter pour une stimulation de même intensité, après répétitions d'intenses activations pré-synaptiques, généralement courtes et de fréquence élevée (Loi de Hebb). Ce phénomène est appelé potentialisation à long terme (PLT) et fait partie des mécanismes de plasticité synaptique. Il implique le système excitateur glutamatergique, avec ses récepteurs AMPA et surtout NMDA qui permettent le renforcement de la synapse lorsque les dendrites reçoivent d'intenses stimulations (cf section 3.1.2.1).

---

2. Corne d'Ammon (CA) : une des sous-régions de l'hippocampe, elle-même divisible en CA1, CA2 et CA3.



Nous avons vu à la section 3.1.2.3 que l'amplitude de la réponse du neurone post-synaptique dépend de la sommation des courants ioniques positifs et négatifs (phénomène d'intégration somatique). Le système Inhibiteur GABAergique participe à la modulation de la réponse post-synaptique en compensant les effets du système Excitateur glutamatergique. Il peut alors avoir un effet sur les processus d'apprentissage en freinant la PLT.

Dans l'hippocampe, les neurones pyramidaux, notamment de CA1, contiennent des récepteurs GABA-A  $\alpha_5$  et des récepteurs NMDA. Les récepteurs  $\alpha_5$  régulent l'excitabilité des dendrites des neurones pyramidaux ([Crestani et al., 2002]), influençant ainsi la probabilité de PLT des synapses. La balance entre entrées excitatrices et inhibitrices au niveau des synapses est déterminante pour réussir ou non à mémoriser ([Yee et al., 2004]). Pour comprendre le rôle de l'inhibition tonique dans les processus de mémoire dépendant de l'hippocampe, il est donc nécessaire de considérer l'interaction entre poids Excitateur et poids Inhibiteur dans l'hippocampe.

### 3.3.2 Métaphore de la balance Excitation/Inhibition

La métaphore de la balance mécanique de Roberval permet de penser de manière simplifiée la compensation entre Excitation et Inhibition à différentes échelles, celle du neurone ou de la région cérébrale ([Sohal and Rubenstein, 2019], [Sears and Hewett, 2021]). Chaque plateau de la balance contient des poids associés à l'excitation (E) d'un côté et à l'inhibition (I) de l'autre : on parle de balance E/I (cf figure 3.7).

A l'équilibre, les plateaux sont stables et leur position relative permet de penser un ratio entre excitation et inhibition (noté pour la suite ratio E/I) qui caractériserait l'état de la balance. La communauté utilise parfois comme raccourcis de langage l'expression "balance E/I" pour désigner l'état de la balance i.e. le ratio E/I. Il existerait un ratio E/I dans chaque région cérébrale permettant un fonctionnement cérébral non pathologique.

Si le poids Excitateur ou Inhibiteur sur l'un des plateaux est modifié, la balance bascule d'un côté ou de l'autre pour atteindre un nouvel état stationnaire. A l'équilibre, le ratio E/I se trouve modifié ([Sohal and Rubenstein, 2019], [Selten et al., 2018]). Ce changement d'état stationnaire peut être qualifié dans les travaux de la communauté par l'expression "altération de la balance E/I" (par rapport à une situation saine de référence qui est implicite).

L'hypothèse d'un déséquilibre entre Excitation et Inhibition est actuellement étudiée dans différentes pathologies neurologiques (e.g. maladie d'Alzheimer, épilepsie) et psychiatriques (e.g. autisme, schizophrénie, troubles bipolaires, dépression, trouble de stress post-traumatique, [Selten et al., 2018]). Les pistes explorées concernent notamment une modification au niveau du système GABAergique, qui causerait alors une altération de la balance E/I dans certaines régions cérébrales. [Selten et al., 2018] soulignent le rôle vital de l'inhibition dans le maintien de la balance E/I chez des personnes au fonctionnement cérébral sain.

### 3.3.3 La balance Excitation/Inhibition pour penser les apprentissages dépendant de l'hippocampe

Nous poursuivons la métaphore de la balance E/I, illustrée à la figure 3.7, pour nous permettre de faire le lien entre ratio E/I dans l'hippocampe et performances mnésiques. La hauteur du plateau E peut être associée à un niveau de performance mnésique (e.g. mémoire spatiale ou l'association entre un stimulus conditionnel et inconditionnel, cf section 3.3.1). Pour un même poids I, plus le poids E est grand et plus la potentialisation à long terme est favorisée à l'échelle neuronale, ce qui est associé à une augmentation d'activité des neurones de l'hippocampe et à des performances mnésiques accrues. Mécaniquement parlant, l'augmentation du poids E conduit à diminuer la hauteur du plateau E. Nous avons donc orienté l'axe indiquant le sens de croissance d'activité de l'hippocampe et des performances mnésiques vers le bas sur la figure 3.7. Plus le plateau E est bas, c'est-à-dire plus le poids E est lourd, et plus les performances mnésiques augmentent.

Une autre manière de faire basculer la balance du côté Excitateur consiste à diminuer le poids I, sans modifier le poids E. Le poids I "compense" moins le poids E et la hauteur du plateau E diminue. Les processus de PLT à l'échelle neuronale sont alors plus importants et les performances mnésiques seront meilleures.

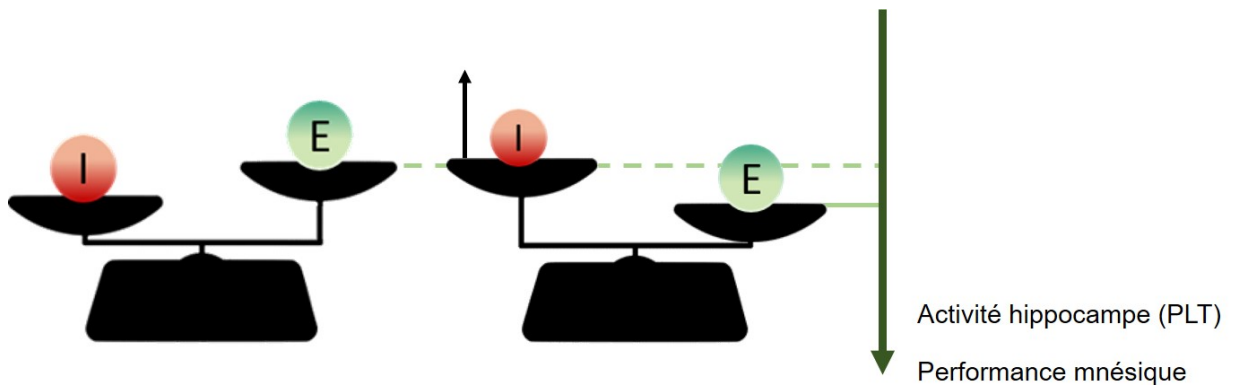


**Figure 3.7** – Schéma de la balance E/I avec une levée de l'inhibition. Le poids Excitateur est associé aux processus de plasticité synaptique de potentialisation à long terme (PLT). Une levée de l'Inhibition est représentée par une diminution du poids I (avec un volume plus petit sur le schéma). Ceci conduit à un bascule de la balance en faveur de l'Excitation : la hauteur du plateau E diminue, ce qui est associé à une activité de l'hippocampe plus grande et à de plus grandes performances mnésiques. Dans cette métaphore, l'axe de la performance mnésique est donc orienté vers le bas, côté E.

Les résultats expérimentaux sur modèles animaux peuvent être réanalysés à la lumière de cette métaphore. Une baisse du poids de l'inhibition tonique dans l'hippocampe chez les animaux mutés par rapport aux souris contrôles est associée à une augmentation des performances mnésiques, qu'il s'agisse de mémoire spatiale ([Collinson et al., 2002]) ou de la mémoire de l'association temporelle entre un son et un choc électrique ([Crestani et al., 2002], [Yee et al., 2004]).



## Quelques altérations possibles du système GABAergique

Les altérations du système GABAergique susceptibles d'impacter l'état de la balance E/I peuvent être diverses ([Selten et al., 2018]) et concerner différentes échelles.

### Echelle synaptique (i.e. échelle neuronale)

A l'échelle de la synapse, nous pouvons citer, de manière non exhaustive, différents éléments pour chaque étape nécessaire à la transmission de l'influx nerveux :

1. Synthèse de neurotransmetteurs : production de GAD (enzyme de production du GABA), quantité de neurotransmetteurs synthétisés
2. Largage ("release") : taille des vésicules, mécanismes de la fusion membranaire des vésicules
3. Diffusion dans la fente synaptique : géométrie de la fente
4. Recapture du GABA : nombre et distribution spatiale des transporteurs (GAT1 et GAT3)

Les points 1 à 4 peuvent modifier la concentration de GABA dans la fente synaptique.

5. Etat des récepteurs : phosphorylation des récepteurs par les protéines kinases ; composition des sous-unités, nombre et distribution/densité spatiale des récepteurs impliqués dans les deux types d'inhibition (phasique ou tonique).
6. Fixation sur le récepteur : affinité ligand-récepteur, production de modulateurs allostériques positifs
7. Courants ioniques : propriété des canaux ioniques (conductance, vitesse de changement d'état), altérations de l'homéostasie du chlorure (modifications de l'expression des transporteurs de  $K^+/Cl^-$  KCC2, [Bergeron et al., 2014] ) <sup>a</sup>

### Echelle de la région (i.e. échelle neurale)

A l'échelle d'une région, le poids inhibiteur de la balance peut être influencé par/dépendre de :

- le nombre d'interneurones inhibiteurs
- la densité de fentes synaptiques fonctionnelles

Le stress chronique, par exemple, impacte le système GABAergique à plusieurs de ces niveaux ([Maguire, 2014], [Czéh et al., 2015], [McEwen et al., 2016]).

---

<sup>a</sup>. Les transporteurs de  $K^+/Cl^-$  (KCC2) permettent de relâcher des ions  $Cl^-$  à l'extérieur du neurone. Des déficit de transporteurs  $K^+/Cl^-$  (KCC2) peuvent augmenter la concentration de  $Cl^-$  intra-cellulaire et donc l'inhibition du neurone

## 3.4 Contrôle inhibiteur des intrusions mnésiques

### 3.4.1 Mesures de concentrations GABA par spectroscopie

[Schmitz et al., 2017] mesurent in vivo la concentration totale (intra et extra neuronale) de GABA dans l'hippocampe de sujets sains au repos, grâce à de la spectroscopie de résonance magnétique nucléaire du proton (Magnetic Resonance Spectroscopy, MRS). Ces sujets réalisent par ailleurs une tâche de contrôle inhibiteur des intrusions mnésiques avec le paradigme Think/No Think. Des données comportementales permettent de mesurer les performances de rappel après la tâche. Des données d'imagerie cérébrale sont obtenues en IRMf pendant la tâche et analysées avec les méthodes que nous présenterons plus en détails au chapitre 5 (sections 5.1.1.2, 5.4.4, 5.5.2). Les concentrations de GABA, dans le cortex préfrontal dorso-latéral (DLPFC) et dans l'hippocampe, sont mises en relation avec les résultats comportementaux et avec les résultats d'imagerie.

Au niveau comportemental, l'étude montre que les participants avec une plus faible concentration de GABA dans l'hippocampe se rappellent ensuite davantage des images No Think (sur lesquelles se sont exercées un contrôle inhibiteur) en comparaison des images Baseline<sup>3</sup>, par rapport aux participants avec une plus grande concentration de GABA dans l'hippocampe. Les premiers oublient moins les images sur lesquelles ils ont exercé un contrôle inhibiteur pendant la tâche, que les participants avec une plus forte concentration de GABA hippocampique. En revanche, la concentration de GABA dans le DLPFC n'est pas associée à des performances de rappel différentes entre participants.

Les analyses d'imagerie montrent une baisse de l'activité de l'hippocampe provoquée par le DLPFC pendant le contrôle inhibiteur des intrusions (No Think). Une plus faible concentration de GABA dans l'hippocampe est associée à une moindre baisse de l'activité de l'hippocampe sous l'effet de la région de contrôle. Par contraste, la baisse d'activité de l'hippocampe ne dépend pas de la concentration GABA dans le DLPFC.

Nous retrouvons ici encore l'idée qu'un moindre "poids inhibiteur" dans l'hippocampe est associée à de plus grandes performances mnésiques (cf figure 3.7), au sens d'un plus fort rappel des images No Think après la tâche. Même après abaissement de l'activité hippocampique par la région de contrôle (DLPFC), le niveau d'excitation resterait suffisamment élevé pour maintenir une activité suffisante de la région cible, permettant une récupération ultérieure des images No Think. De plus, la concentration de GABA dans l'hippocampe semble présenter un effet modulateur sur la baisse d'activité de celui-ci en réponse à l'augmentation d'activité de la région de contrôle. Autrement dit, *l'état de la balance E/I de l'hippocampe semble modifier la sensibilité de celui-ci à une augmentation de l'activité de la région de contrôle.*

---

3. pour lesquelles aucune récupération ou contrôle inhibiteur n'a eu lieu pendant le TNT

### 3.4.2 Hypothèse de la variabilité interindividuelle de l'état des balances E/I

Pour éclairer la question de la variabilité interindividuelle du contrôle inhibiteur, en particulier du contrôle inhibiteur des intrusions mnésiques, nous émettons l'hypothèse d'une variabilité interindividuelle de l'état de la balance E/I pendant la tâche au niveau de :

- l'émission : une activation neurale insuffisante des régions de contrôle.

Dans le cas du contrôle inhibiteur des intrusions mnésiques, [Schmitz et al., 2017] rapportent en début d'introduction l'hypothèse dans la communauté neuroscientifique d'un engagement diminué des régions du cortex préfrontal. Les résultats de cette étude n'indiquent cependant pas que le "poids Inhibiteur" (i.e. la concentration de GABA) dans les régions de contrôle des participants testés puisse participer à la variabilité interindividuelle des mécanismes d'abaissement d'activité de l'hippocampe. Nous explorons alors dans la thèse la variabilité interindividuelle du poids Excitateur dans les régions de contrôle.

- la réception : une modification du système GABAergique dans les régions cibles.

Dans le cas du contrôle inhibiteur des intrusions mnésiques, le « poids Inhibiteur » dans l'hippocampe des sujets pourrait participer à la variabilité interindividuelle.

Pour nous permettre de mieux comprendre les mécanismes qui sous-tendent la variabilité interindividuelle du contrôle inhibiteur dans le continuum sain, et, à plus long terme, les déficits de contrôle inhibiteur dans diverses pathologies psychiatriques, il semble intéressant de :

1. pouvoir caractériser les propriétés neurales (Excitatrice et Inhibitrice) locales au sein de régions cérébrales impliquées dans une tâche de contrôle inhibiteur.  
L'utilisation de l'IRMf permet de mesurer l'activité cérébrale de manière non invasive pendant une tâche de contrôle inhibiteur, mais présente des limites en tant que mesure indirecte de l'activité neurale. C'est ce que nous présentons au chapitre 4.
2. préciser comment l'état des balances E/I locales influence l'effet que peut avoir un groupe de régions de contrôle préfrontales sur la baisse de l'activité d'un groupe de régions cibles.

## Chapitre 4

# Signaux BOLD obtenus avec l'IRMf : des mesures très indirectes de l'activité des neurones sous-jacents

*Pour que des physiciens du non linéaire puissent comprendre la nature de la mesure, je détaille ici la physique du signal BOLD obtenu avec l'Imagerie par Résonance Magnétique fonctionnelle. L'objectif est de montrer qu'il s'agit d'une mesure très indirecte de l'activité neurale sous-jacente. Nous verrons également quelques contraintes liées à l'utilisation de telles mesures.*

L'Imagerie par Résonance Magnétique fonctionnelle (IRMf) a été développée depuis les années 1990 (pour un historique voir [Viard et al., 2021]). C'est la technique d'imagerie non invasive du cerveau entier qui dispose de la meilleure résolution spatiale à ce jour (voir figure 4.1). Sa résolution temporelle, de l'ordre de la seconde, permet de mesurer chez l'humain in vivo et en temps réel des variations du niveau d'oxygène dans le sang<sup>1</sup>, en lien avec les variations locales de l'activité des neurones sous-jacents.

Elle se base sur la résonance magnétique du noyau de l'atome d'hydrogène, contenu en quantité abondante dans les tissus, principalement dans les molécules d'eau. Le proton des atomes d'hydrogène constitue un traceur endogène et la déoxyhémoglobine présente dans le sang est utilisée comme agent de contraste endogène du niveau d'oxygénation des tissus. L'IRMf ne nécessite ainsi aucune injection radioactive, seulement l'exposition de la personne à un fort champ magnétique (aucun effet délétère pour la santé n'a été rapporté jusqu'à aujourd'hui).

---

1. Fondamentalement, nous verrons que l'IRMf mesure les variations de la susceptibilité magnétique qui dépend du niveau d'oxygène dans le sang.

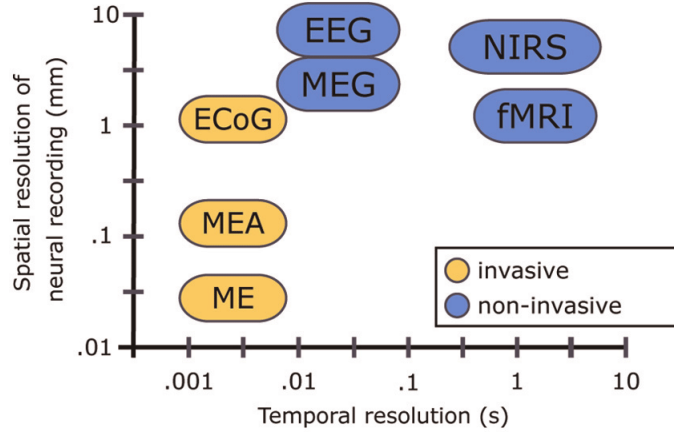


**Figure 4.1** – Résolutions spatiales et temporelles des principales techniques d'imagerie en neurosciences (issu de [Vogel et al., 2015]) : electroencephalography (EEG), magnetoencephalography (MEG), near-infrared spectroscopy (NIRS), functional magnetic resonance imaging (fMRI), electrocorticography (ECOG), microelectrode array (MEA) recordings and single microelectrode (ME) recordings. L'IRMf donne le meilleur compromis entre la résolution spatiale et la résolution temporelle dans l'imagerie non invasive chez l'humain.

## 4.1 Principe physique de l'IRM : obtenir une mesure de la relaxation de l'aimantation macroscopique locale dans les tissus

### 4.1.1 Définition du système, de l'observable macroscopique et des directions

Le spin est une propriété quantique d'une particule qui permet de caractériser son comportement lorsqu'elle est plongée dans un champ magnétique<sup>2</sup>. A chaque spin  $\vec{s}$  est associé un moment magnétique de spin  $\vec{\mu} = \gamma \vec{s}$  ( $\gamma$  est une constante appelée facteur gyromagnétique).  $\vec{\mu}$  représente l'aimantation intrinsèque de la particule, qui se comporte comme un petit aimant susceptible de s'aligner à des lignes de champ magnétique. Le spin du proton (spin nucléaire de l'atome d'hydrogène) est non nul. Le proton possède donc un moment magnétique de spin non nul, ce qui le rend réactif lors de son exposition à un champ magnétique externe.

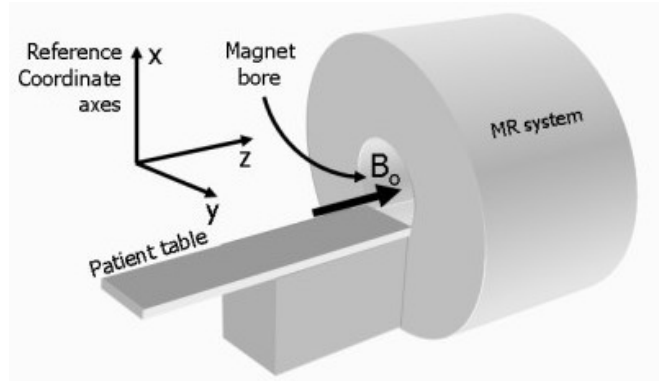
Le système est constitué de l'ensemble des moments magnétiques de spin des atomes de H. Nous distinguerons dans la suite deux niveaux de description. Le niveau microscopique concerne la caractérisation du comportement des moments magnétiques individuels des protons  $\vec{\mu}_i(t)$ . Au niveau macroscopique, la somme des moments magnétiques nucléaires résulte en une aimantation (totale) mesurable, notée  $\vec{M}(t)$ .

2. De la même manière que la charge d'une particule (positive ou négative) permet de caractériser son comportement dans un champ électrique, ou la masse dans un champ gravitationnel

$$\vec{M}(t) = \sum_i \vec{\mu}_i(t) \quad (4.1)$$

En toute rigueur, nous devrions expliciter la dimension spatiale (i.e.  $\vec{M}(\vec{x}, t)$ ) et sommer les  $\mu_i$  dans une petit espace autour de  $\vec{x}$  pour construire le passage d'échelle. Pour faciliter la présentation, nous nous concentrons ici sur une description des phénomènes temporels, dans un point de l'espace  $\vec{x}$  fixé.

L'un des principaux composants d'une IRM est son aimant : une bobine supraconductrice traversée par un courant électrique qui permet de générer en son centre un champ magnétique homogène et constant dans le temps, appelé champ principal et noté  $\vec{B}_0$ . En considérant une personne allongée dans l'IRM, regard vers le haut,  $\vec{B}_0$  est orienté des pieds vers la tête. Chaque vecteur dans cet espace est décomposable dans un repère  $(xyz)$ , de base  $(\vec{e}_x, \vec{e}_y, \vec{e}_z)$ . L'axe  $z$ , appelé axe principal, est dans la même direction et le même sens que  $\vec{B}_0$ . On parle aussi de direction principale. Les axes  $x$  et  $y$  définissent le plan transverse, perpendiculaire à  $\vec{B}_0$ , ainsi qu'illustré ci-après (adapté de [Ridgway, 2010]) :



Chaque vecteur d'aimantation  $\vec{M}$  est ainsi décomposable comme une somme de sa composante longitudinale  $\vec{M}_z$  (dans la direction de  $\vec{B}_0$ ) et sa composante transversale  $\vec{M}_{xy}$  (projection dans le plan transverse).

#### 4.1.2 Condition initiale : application d'un champ magnétique constant $\vec{B}_0$ et description de l'état d'équilibre

Avant l'exposition à  $\vec{B}_0$ , les moments magnétiques individuels  $\vec{\mu}_i$  des atomes d'hydrogène s'orientent dans toutes les directions de manière aléatoire. Une fois soumis à  $\vec{B}_0$ , chaque moment magnétique individuel va adopter un mouvement de précession<sup>3</sup> autour d'un axe orienté dans la direction de  $\vec{B}_0$ . La fréquence de rotation d'un moment magnétique autour de son axe principal est appelée fréquence de Larmor et est proportionnelle à l'intensité du champ magnétique  $\vec{B}_0$  (exprimée en Hz/Tesla)<sup>4</sup>. Plus le champ magnétique dans lequel se

3. mouvement de rotation autour d'un axe, comme une toupie

4. La fréquence de Larmor est caractéristique de chaque type de noyau et dépend de sa composition

trouve un noyau est fort, plus la vitesse de rotation de son moment magnétique est grande. Tous les moments magnétiques  $\vec{\mu}_i$  se mettent à précesser à la même fréquence, mais avec des phases distribuées de manière aléatoire. Au niveau macroscopique, la somme des composantes transversales des moments magnétiques  $\vec{\mu}_i$  annule la composante transversale de l'aimantation totale :  $\vec{M}_{xy} = \vec{0}$ . De plus, la composante longitudinale de chaque moment magnétique  $\vec{\mu}_i$  qui précesse est orientée dans le sens de  $\vec{B}_0$  (spin up ou parallèle) ou dans le sens inverse (spin down ou antiparallèle), tel qu'illustré à la figure 4.2. Une majorité se trouve dans le sens de  $\vec{B}_0$  (corpus de la physique statistique, distribution de Boltzmann), la composante longitudinale de l'aimantation qui résulte de la somme des composantes longitudinales des moments magnétiques individuels est donc positive :  $\vec{M}_z = \underbrace{M_z}_{>0} \vec{e}_z$ .

Le vecteur d'aimantation  $\vec{M} = \vec{M}_{xy} + \vec{M}_z = M_z \vec{e}_z$ , se trouve alors dans la même direction et le même sens que le champ principal  $\vec{B}_0$ . Dans cette configuration, le système se trouve à l'équilibre.

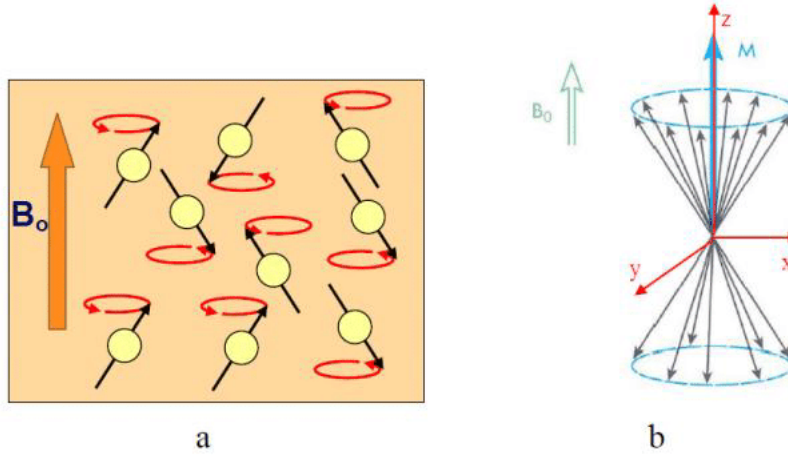


**Figure 4.2** – Précession des moments magnétiques et aimantation principale selon  $\vec{B}_0$   
(issu de [Fatima et al., 2020])

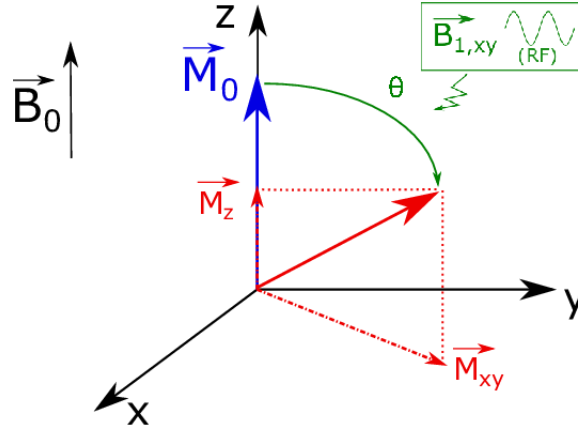
### 4.1.3 Perturbation du système par des impulsions radio fréquences et résonance magnétique

Le système est perturbé sur de courts intervalles de temps par un champ magnétique perpendiculaire à  $\vec{B}_0$  et oscillant à une fréquence dans le domaine des ondes radio, dans la direction d'un axe fixe. Ce champ est noté  $\vec{B}_1(t)$ . Les impulsions radio fréquence sont émises par l'antenne radio-fréquence de l'IRM, deuxième composant principal. Si la fréquence d'oscillation de  $\vec{B}_1(t)$  est exactement égale à la fréquence de précession des moments magnétiques  $\vec{\mu}_i$  (i.e. la fréquence de Larmor) imposée par l'intensité de  $\vec{B}_0$ , le système entre en résonance. Les protons emmagasinent une partie de l'énergie des ondes radio et quittent leur état d'équilibre : c'est la phase d'excitation. La fréquence d'excitation particulière de la perturbation externe  $\vec{B}_1(t)$  est appelée fréquence de résonance du système.

La résonance induit deux phénomènes dans le système. Les spins des protons se mettent à

tourner en phase autour de leur axe principal, toujours à leur fréquence de Larmor, et les angles de précession augmentent avec la durée de l'impulsion. Si elle est suffisante, les spins basculent à  $90^\circ$  du champ magnétique principal  $\vec{B}_0$  (i.e. dans le plan transverse).

Ces deux effets ont pour conséquence de faire croître progressivement la composante transversale  $\vec{M}_{xy}$  de l'aimantation et de faire diminuer progressivement sa composante longitudinale  $\vec{M}_z$ . Si la durée et l'amplitude des impulsions sont suffisantes, la composante longitudinale s'annule ( $\vec{M}_z = 0$ ) et la composante transversale  $\vec{M}_{xy}$  est maximale : l'aimantation  $\vec{M}$  résultante se trouve ainsi complètement projetée dans le plan transverse à  $\vec{B}_0$  et précesse à la fréquence de Larmor autour de la direction principale (donnée par  $\vec{B}_0$ ), tel qu'illustré par ce schéma (tiré de [Lefebvre, 2017]) :



#### 4.1.4 Arrêt des impulsions radio-fréquence : phase de relaxation transverse

La phase d'excitation cesse avec l'arrêt de l'émission des ondes radio. Chaque spin va revenir progressivement à son mouvement de précession d'équilibre en réduisant l'angle formé avec la direction principale. C'est la phase de relaxation du système, qui se traduit par une diminution de la composante transversale de l'aimantation  $\vec{M}_{xy}$  et une augmentation de sa composante longitudinale  $\vec{M}_z$ .

La diminution de la composante transversale de l'aimantation  $\vec{M}_{xy}$  (ou relaxation transverse) est utilisée pour l'IRMf ([Andreelli and Mosbah, 2014]) et nous la présentons dans ce paragraphe. Nous ne détaillons pas ici l'augmentation de la composante longitudinale  $\vec{M}_z$  de l'aimantation (ou relaxation longitudinale), dont la mesure du temps caractéristique de croissance exponentielle ( $T_1$ ) est utilisée en IRM structurelle (cf figure 4.3).



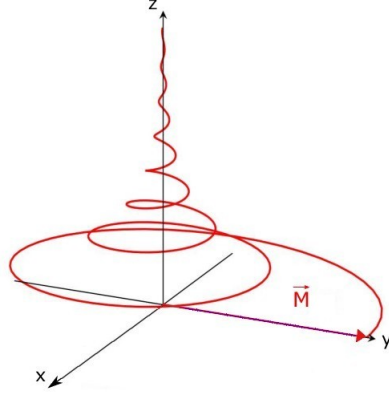


**Figure 4.3** – Précession libre de l'aimantation  $\vec{M}$  pour revenir à la situation d'équilibre. Seule la trajectoire au cours du temps de la pointe du vecteur est représentée (adaptation de [Lefebvre, 2017]). La norme de  $\vec{M}$  n'est pas constante dans le temps, sinon la trajectoire rouge se ferait sur une sphère : les relaxations transversale et longitudinale sont indépendantes et se caractérisent par leur propre temps caractéristique :  $T_2$  pour la relaxation transversale et  $T_1$  pour la relaxation longitudinale.

#### 4.1.4.1 Relaxation $T_2$

Pendant la phase de relaxation du système, les spins des atomes d'hydrogène du tissu interagissent entre eux, ce qui conduit à leur déphasage, d'autant plus rapidement que le tissu est dense. Le déphasage des spins se traduit au niveau macroscopique par une diminution progressive de la composante transversale de l'aimantation  $\vec{M}_{xy}$ , qui tend à s'annuler : on parle aussi de relaxation transverse. La diminution des composantes transversales des moments magnétiques  $\vec{\mu}_i(t)$  contribue de manière négligeable à la relaxation transverse de l'aimantation devant le déphasage.

Dans toute la suite  $M_{xy}$  désignera la composante de  $\vec{M}_{xy}$  selon  $\vec{B}_1$ . Selon l'axe  $\vec{B}_1$ , cette décroissante se traduit par une oscillation avec une enveloppe qui suit une loi exponentielle décroissante (cf figure 4.4). Cette enveloppe est caractérisée par deux paramètres :

- $M_1$  : norme maximale de l'aimantation. Elle dépend de la densité de protons (i.e. atomes de H) dans les tissus et varie d'un tissu à l'autre
- $T_2$  : temps caractéristique de l'exponentielle (i.e. temps écoulé pour obtenir 37 pourcent de  $M_1$ )

$$M_{xy}(t) = M_1 e^{-\frac{t}{T_2}} \quad (4.2)$$

#### 4.1.4.2 Relaxation $T_2^*$

Des variations locales du champ magnétique  $\vec{B}_0$ , aussi appelées inhomogénéités de champ, peuvent être causées par la présence de certaines molécules paramagnétiques dans le milieu comme nous le verrons au paragraphe 4.3. Elles modifient les fréquences de précession des moments magnétiques des protons  $\vec{\mu}_i$ , favorisant ainsi leur déphasage. Au niveau macroscopique, la composante transversale de l'aimantation  $\vec{M}_{xy}$  décroît plus rapidement en présence de ces inhomogénéités. On note  $T_2^*$  le temps caractéristique de la décroissance exponentielle en présence d'inhomogénéités ( $T_2^* < T_2$ , comme représenté sur la figure 4.4). Les vitesses de relaxation dépendent des caractéristiques des tissus qui ont été excités par les ondes radio fréquence. Une valeur de  $T_2^*$  plus élevée correspond à une courbe de relaxation transverse  $\vec{M}_{xy}$  qui décroît plus lentement.

L'IRM fonctionnelle utilise la relaxation en  $T_2^*$  ([Andreelli and Mosbah, 2014]).

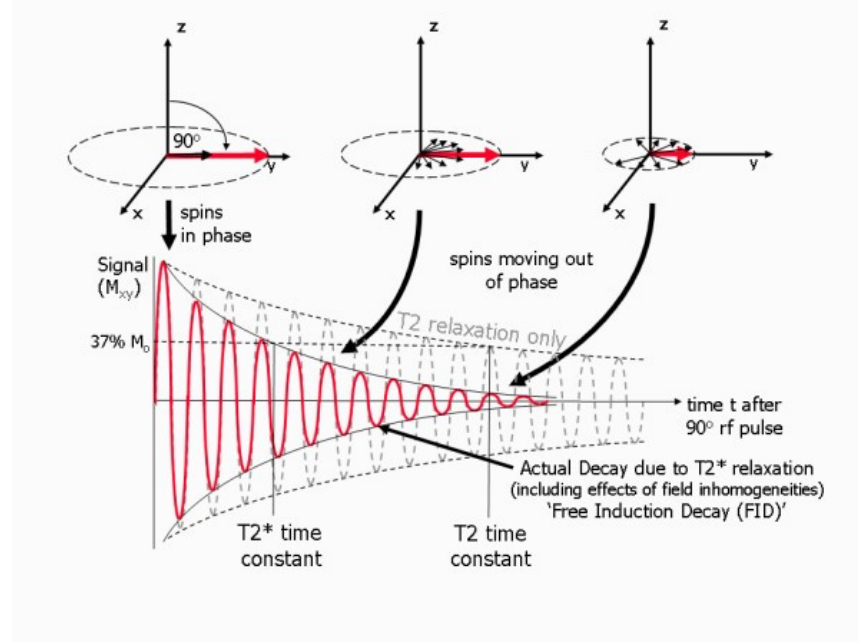


**Figure 4.4** – Comparaison des relaxations  $T_2$  et  $T_2^*$ . La composante transversale de l'aimantation, notée  $\vec{M}_{xy}$  est représentée par un vecteur rouge. Elle tourne autour de l'axe z et sa norme décroît au cours du temps (cercle en pointillés avec un rayon de plus en plus petit). Les petits vecteurs noirs symbolisent les moments magnétiques  $\vec{\mu}_i$  que l'on voit se déphaser de plus en plus au cours du temps.

La projection au cours du temps de  $\vec{M}_{xy}$  selon l'axe y, est représentée par la courbe oscillante rouge sur le graphe du bas, dont l'enveloppe décroît de manière exponentielle en  $T_2^*$ . La courbe grise représente la relaxation, avec un temps caractéristique  $T_2$ , qui aurait eu lieu sans inhomogénéités de champ. En présence de ces inhomogénéités, la composante transversale décroît plus rapidement :  $T_2^* < T_2$  (illustration de [Ridgway, 2010]).

## 4.2 Variation de la concentration locale en déoxyhémoglobine due au couplage neurovasculaire

Les variations du niveau d'activité local des neurones sont suivies de variations hémodynamiques locales pour couvrir les besoins métaboliques ([Andreelli and Mosbah, 2014]) : c'est ce que désigne la notion de *couplage neurovasculaire*.

Les connexions entre neurones, cellules gliales et capillaires, qui forment l'architecture neuro-gliale-vasculaire, sont à l'origine du couplage neurovasculaire. La production de neurotransmetteurs dans la fente synaptique déclenche une suite de réactions chimiques dans les cellules gliales avoisinantes (les astrocytes), enveloppées autour des synapses. Celles-ci présentent des projections jusqu'à la surface des micro capillaires ([Zisis et al., 2021]), comme représenté à la figure 4.5 et vont déclencher une réponse hémodynamique.

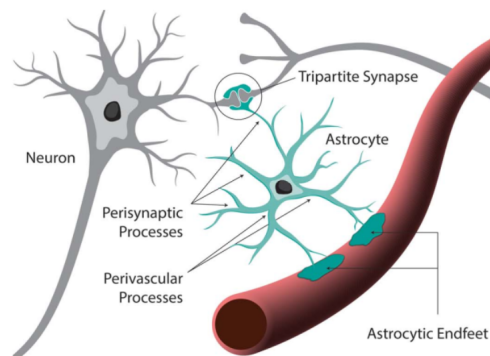


**Figure 4.5** – Les astrocytes s’enroulent autour des synapses et possèdent des projections à la surface des microcapillaires sanguins (issue de [Zisis et al., 2021], fig. 1A)

Les changements hémodynamiques mesurés lors de l'augmentation de l'activité de neurones comprennent une augmentation du volume des compartiments veineux à proximité, du flux sanguin et de la concentration en oxyhémoglobine, associée d'une baisse de la concentration en déoxyhémoglobine ([Buxton, 2012]). Cette dernière baisse semble paradoxale étant donné que les neurones consomment davantage de dioxygène et qu'on s'attendrait alors à une augmentation de la concentration en déoxyhémoglobine. Néanmoins elle s'explique par le fait que l'apport en sang oxygéné est supérieur aux besoins de consommation des neurones ([Ogawa et al., 1992]). Ce sont les variations de la concentration en déoxyhémoglobine qui sont détectables à l'IRMf et qui vont permettre de mesurer de manière indirecte l'activité des neurones. Un modèle du couplage neurovasculaire et de sa mesure avec l'IRMf va être présenté dans le chapitre 5 (section 5.3.2.2).

## 4.3 Principe de l'IRM fonctionnelle : $T_2^*$ dépend des variations locales de la concentration en déoxyhémoglobine

L'IRMf utilise la relaxation en  $T_2^*$  pour quantifier les variations locales de la concentration en déoxyhémoglobine.

### 4.3.1 Oxyhémoglobine et déoxyhémoglobine : des propriétés magnétiques différentes

#### 4.3.1.1 Composition

L'hémoglobine est une protéine contenue dans les globules rouges, qui permet le transport du dioxygène dans le sang, des poumons jusqu'aux tissus. Un globule rouge humain a la forme d'un disque, creusé en son centre, qui mesure environ 7 micromètres de diamètre et contient environ 270 millions de molécules d'hémoglobine ([Pierigè et al., 2008], cf figure 4.6). Chez tous les vertébrés, l'hémoglobine est composée de quatre sous-unités polypeptidiques identiques deux à deux (deux chaînes  $\alpha$  et deux chaînes  $\beta$ ), appelées globines. A l'intérieur de chaque globine se trouve un groupement non protéique, un hème, qui possède en son centre un cation de fer  $Fe^{2+}$  auquel peut se lier le dioxygène de façon réversible (cf figure 4.7). Une liaison covalente se forme entre l'un des atomes d'oxygène et le cation de fer. L'autre atome d'oxygène est lié à l'histidine distale de la globine par une liaison hydrogène.

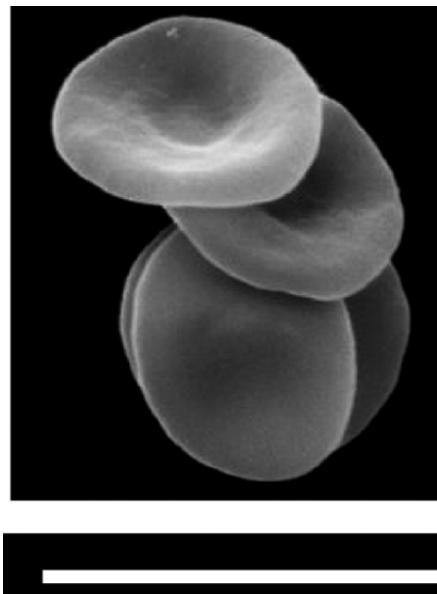


**Figure 4.6** – Images de globules rouges humains obtenues par microscopie électronique. La barre blanche représente  $10\mu m$  ([Pierigè et al., 2008]). Chaque globule contient environ 270 millions de molécules d'hémoglobine, comme celle de la figure 4.7.

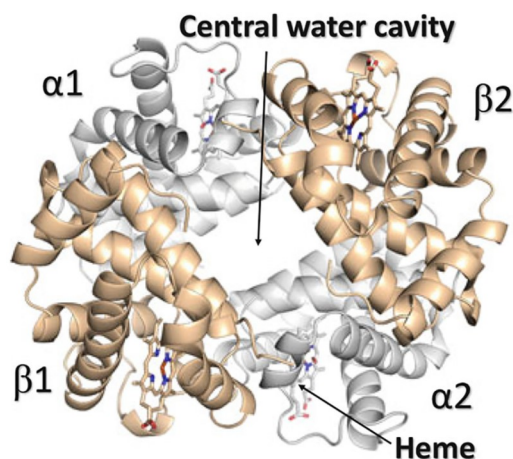


**Figure 4.7** – Structure cristalline de l'hémoglobine sous sa forme déoxygénée, constituée de deux chaînes  $\alpha$  et deux chaînes  $\beta$ , respectivement colorées en gris et doré ([Ahmed et al., 2020]). Chacune contient un hème qui permet de fixer l'oxygène.

Un hème peut ainsi transporter une molécule de dioxygène et chaque hémoglobine, qui contient quatre hèmes, peut transporter au plus quatre molécules de dioxygène. Si l'hémoglobine est liée au moins à une molécule de dioxygène, on parle d'oxyhémoglobine. Lorsque les quatre hèmes sont liés, l'oxyhémoglobine est dite saturée. Au contraire, si l'hémoglobine n'est liée à aucune molécule de dioxygène, on parle de déoxyhémoglobine.

Chaque liaison du dioxygène sur un hème provoque un changement de conformation de la sous unité concernée, qui affecte la conformation des autres sous unités dont l'affinité (ou probabilité de liaison) pour le dioxygène augmente. C'est une liaison allostérique de type coopératif : l'affinité de l'hémoglobine pour le dioxygène croît avec le nombre de molécules de dioxygène liées, et inversement. La déoxyhémoglobine bascule ainsi progressivement de la forme T (plus ouverte) à la forme R (plus compacte) pour former une oxyhémoglobine saturée (quatre  $O_2$ ), et inversement<sup>5</sup>.

#### 4.3.1.2 Propriété magnétique

Du fait des propriétés électroniques de leurs atomes de fer, l'oxyhémoglobine et la déoxyhémoglobine ont des propriétés différentes lorsqu'elles se trouvent dans un champ magnétique statique ([de Marco et al., 2006]).

La configuration électronique du cation  $Fe^{2+}$  d'un hème s'écrit  $[Ar]3d^6$  (corpus de la mécanique quantique). Les électrons de valence se trouvent dans les orbitales atomiques 3d. La répartition de ces électrons de valence dans les orbitales moléculaires qui lient le cation  $Fe^{2+}$  au reste du complexe de l'hème laisse apparaître quatre électrons non appariés ([Wittenberg et al., 1970], [Melville et al., 1975]) dont la somme des spins confère au cation

5. pour visualiser le changement de conformation, nous pourrions nous rapporter à la figure 14.2.b de [Ahmed et al., 2020]

$Fe^{2+}$  (et donc à l'hème sans  $O_2$ ) un spin total non nul<sup>6</sup>. La sous unité de l'hémoglobine, sans ligand  $O_2$ , est paramagnétique : elle possède un moment magnétique de spin qui s'aligne au champ magnétique auquel elle est soumise. La déoxyhémoglobine, constituée des quatre sous unités non liées au dioxygène, est aussi paramagnétique.

La présence de déoxyhémoglobine paramagnétique dans les tissus génère alors des perturbations du champ magnétique local (on parle aussi d'inhomogénéités du champ magnétique). Des gradients de champs magnétiques locaux s'étendent des globules rouges intra-vasculaires jusqu'au milieu périvasculaire ([de Marco et al., 2006]).

Le cation  $Fe^{2+}$  peut former une liaison covalente (de coordination) avec un des atomes d'oxygène d'une molécule de dioxygène. Le spin électronique du complexe lié est nul du fait de l'absence d'électrons non appariés. La sous unité d'hémoglobine liée au dioxygène est diamagnétique : elle possède un moment magnétique de spin total nul et ne s'oriente pas dans la direction d'un champ magnétique lorsqu'elle y est soumise. L'oxyhémoglobine saturée, constituée des quatre sous unités liées au dioxygène, est donc également diamagnétique. Elle ne génère aucune perturbation du champ magnétique local.

### 4.3.2 Principe de la mesure en IRMf : un signal dépend du niveau d'oxygénation du sang (BOLD)

#### 4.3.2.1 Présence de déoxyhémoglobine et variations de $T_2^*$

En présence de déoxyhémoglobine, les protons de H sont soumis au champ magnétique principal  $\vec{B}_0$  auquel se rajoutent des perturbations induites par la déoxyhémoglobine (inhomogénéités). Ces inhomogénéités augmentent localement avec la concentration de déoxyhémoglobine. Au niveau microscopique, les spins des protons se déphasent d'autant plus vite que les inhomogénéités sont grandes et donc la concentration en déoxyhémoglobine élevée. Au niveau macroscopique, la composante transversale du vecteur d'aimantation  $\vec{M}_{xy}$  relaxe alors plus rapidement. Ceci conduit à une diminution du temps caractéristique de relaxation  $T_2^*$ .

La présence de déoxyhémoglobine est ainsi détectable de manière indirecte avec l'IRMf, par une diminution locale de  $T_2^*$ . La déoxyhémoglobine agit comme un agent de contraste endogène du signal  $T_2^*$  mesuré ([Andreelli and Mosbah, 2014]).

#### 4.3.2.2 Mesures du signal BOLD dans un voxel

Nous précisons la nature du signal mesuré dans un petit cube de l'espace dont les côtés mesurent quelques millimètres : un voxel<sup>7</sup>. C'est l'échantillon volumique le plus petit d'une image obtenue avec une IRM.

Pendant toute la durée d'une acquisition d'IRMf, des impulsions radio-fréquences sont émises successivement et vont permettre d'effectuer une série temporelle de mesures dont nous allons préciser la nature. L'antenne radio fréquence est utilisée à la fois comme émetteur d'impulsions et récepteur du signal d'aimantation transverse  $M_{xy}(t)$ .

6. corpus de la mécanique quantique, modèles de la structure électronique des complexes de coordination : théorie du champ cristallin et théorie du champ de ligands

7. cf résolution spatiale de l'IRMf à la figure 4.1. Dans notre cas, un voxel mesurera  $3mm \times 3mm \times 3.75mm$ .

#### 4.3.2.2.1 Signal BOLD pondéré en $T_2^*$

Nous avons vu qu'à chaque instant  $t_{stop}$  où l'émission d'une onde radio fréquence est arrêtée, l'aimantation  $\vec{M}(t)$  relaxe et sa composante transversale  $\vec{M}_{xy}$  suit une loi exponentielle décroissante de temps caractéristique  $T_2^*$ .

Supposons une valeur de  $T_2^*$  fixée. Dans un voxel, il faudrait pouvoir effectuer plusieurs mesures de la courbe de relaxation au cours du temps pour pouvoir estimer  $T_2^*$ .

Puisque la concentration en déoxyhémoglobine varie dans le temps,  $T_2^*$  varie aussi dans le temps. On notera explicitement  $T_2^*(t)$ . En général, on ne mesure qu'un seul point de la courbe de relaxation exponentielle à  $t = t_{stop} + TE$ .  $TE$  est appelé temps d'écho, c'est l'instant précis où s'effectue la mesure après arrêt d'une impulsion radio. Nous notons cette mesure  $S$  (pour signal). C'est une fonction de  $T_2^*(t)$  et de  $S_0$  qui dépend de la densité de protons.

$$S(t) = S_0 e^{-\frac{TE}{T_2^*(t)}} \quad (4.3)$$

Puisqu'on ne mesure qu'un seul point de la courbe, nous ne pouvons estimer ni  $S_0$  et ni  $T_2^*(t)$ . Néanmoins, la densité de protons est considérée comme une propriété constante du tissu et donc  $S_0$  est constante dans le temps.  $S$  est alors une fonction strictement croissante de  $T_2^*$  : plus  $T_2^*$  augmente, plus la courbe décroît lentement et plus le signal  $S$  augmente. On parle d'image *pondérée en  $T_2^*$* . Il ne s'agit pas d'une mesure directe de  $T_2^*$ , mais de la mesure d'une fonction strictement croissante de  $T_2^*$ .

Ce signal, fonction de  $T_2^*$ , dépend donc de la présence de déoxyhémoglobine dans les vaisseaux contenus dans le voxel. On peut aussi dire qu'il dépend du niveau d'oxygène dans le sang : on parle de signal dépendant du niveau d'oxygène dans le sang ou BOLD en anglais pour "Blood Oxygen Level Dependent".

Remarque : En fait, un voxel est constitué de différents tissus, avec notamment une partie contenant du sang et une autre partie non sanguine. Une fraction volumique du voxel est associée à chaque partie. Le signal magnétique mesuré dans un voxel, noté  $S_{tot}$ , peut s'écrire plus précisément comme la somme des signaux de chaque partie, pondéré par leur fraction volumique respective ([Uludağ et al., 2009], [Buxton et al., 1998]).

#### 4.3.2.2.2 Série temporelle de mesures

L'émission successive d'impulsions radio permet de mesurer  $S$  au cours du temps, qui dépend de  $T_2^*$  au cours du temps. Nous obtenons plus précisément un échantillonnage de  $S(t)$ . En notant  $t_i$  l'instant de la  $i$ -ème mesure :

$$S(t_i) = S_0 e^{-\frac{TE}{T_2^*(t_i)}} \quad (4.4)$$

Rappelons que  $T_2^*(t_i)$  n'est pas accessible directement par la mesure. Mais que la mesure du signal BOLD à  $t_i$ , notée  $S(t_i)$ , dépend de  $T_2^*(t_i)$ .

Supposons une deuxième mesure effectuée à un autre instant  $t_j$  ( $j \neq i$ ) pendant la durée d'acquisition. L'intensité du signal mesuré dans ce même voxel vaut :

$$S(t_j) = S_0 e^{-\frac{TE}{T_2^*(t_j)}} \quad (4.5)$$



#### 4.3.2.3 Importance des contrastes

Chaque zone du cerveau consomme constamment du dioxygène et donc peut émettre constamment un signal BOLD. La mesure absolue  $S(t_j)$  ne saurait être interprétée sans la comparer à  $S(t_i)$ , i.e. une mesure de référence à partir de laquelle peut être pensé le contraste  $\Delta S = S(t_j) - S(t_i)$ .

$$\Delta S = S(t_j) - S(t_i) = S_0 \left( e^{-\frac{TE}{T_2^*(t_j)}} - e^{-\frac{TE}{T_2^*(t_i)}} \right) \quad (4.6)$$

Lorsque la concentration en déoxyhémoglobine augmente, nous avons vu que  $T_2^*$  diminue (les inhomogénéités de champ accélèrent le déphasage des spins individuels). Une diminution de  $T_2^*$  par rapport au cas de référence résulte en une diminution du signal BOLD (i.e.  $\Delta S < 0$ ). On parle aussi de contraste BOLD négatif.

Inversement, lorsque la concentration en déoxyhémoglobine diminue, nous avons vu que  $T_2^*$  augmente. Une augmentation de  $T_2^*$  par rapport au cas de référence résulte en une augmentation du signal BOLD (i.e.  $\Delta S > 0$ ). On parle aussi de contraste BOLD positif.

Les expériences menées en IRMf sont construites avec des alternance de durées pendant lesquelles la personne réalise ou pas une tâche particulière (paradigme d'activation en bloc). On peut alors définir des périodes de stimulation (ou d'activation) et des périodes hors stimulation (ou de repos). Le décours temporel du BOLD hors stimulation constitue la ligne de base (ou baseline en anglais).

L'état de référence est l'état du cerveau hors période de stimulation. L'état actif est celui de la personne pendant une période de stimulation, que l'on comparera à l'état de référence par soustraction. L'activité associée à la tâche étudiée correspond à la soustraction du signal BOLD mesuré entre les états actif et de référence. C'est le principe de la soustraction cognitive. Des analyses statistiques (appliquées voxel à voxel) permettent ensuite de vérifier qu'il existe bien des variations du signal à l'état actif par rapport à l'état de référence (cf section 5.1 du prochain chapitre).

#### 4.3.2.4 Signal BOLD et activité neurale sous-jacente

Les variations de l'activité des neurones provoquent des changements hémodynamiques, comme nous l'avons évoqué à la section 4.2. Plus les neurones sont actifs, plus le sang la concentration en oxyhémoglobine augmente et plus la concentration en déoxyhémoglobine diminue.

Les variations du BOLD permettent donc de suivre les variations de la réponse hémodynamique et indirectement les variations de l'activité des neurones sous-jacents. On mesure une augmentation localisée du signal BOLD ( $\Delta S > 0$ ) dans une région où les neurones s'activent. Celle-ci est très faible en amplitude (seulement quelques pourcents par rapport à un état de repos [Uludağ et al., 2009]).



### 4.3.3 Nature des données recueillies en IRMf

L'espace en 3D de l'image est subdivisé en échantillons volumiques (voxels) de dimensions  $3mm \times 3mm \times 3mm$ . Chaque voxel contient un nombre très important de neurones<sup>8</sup>.

Nous mesurons pour chaque voxel de l'image une série temporelle de mesures BOLD<sup>9</sup>. Le signal est discontinu : les mesures sont espacées d'un intervalle de temps, noté TR (dans notre cas  $TR = 2.05s$ ), qui correspond à la durée d'acquisition d'un volume d'image en entier. La série temporelle de mesures est aussi appelée décours temporel.

Le jeu de donnée en IRMf est ainsi une série temporelle de volumes 3D qui se succèdent à une fréquence d'échantillonnage donnée.

### 4.3.4 Description d'une réponse

[Kwong et al., 1992], [Ogawa et al., 1992] mesurent les variations BOLD de sujets sains soumis à des alternances de stimulations visuelles (diodes allumées ou éteintes plongeant la personne dans l'obscurité). Ils montrent qu'en période de stimulation, le signal augmente dans le cortex visuel primaire (V1), comme illustré dans la figure 4.8.

Avec un autre type de séquence IRM, [Kwong et al., 1992] mettent en évidence que la fin de la stimulation est marquée par une baisse du signal en dessous du niveau de base (on parle d'undershoot ou creux post-stimulus). [Menon et al., 1995] proposent la même expérience et détectent dans certaines parties du cortex visuel une baisse initiale du signal BOLD juste avant l'augmentation due à la stimulation (initial dip ou déplétion précoce). Cette baisse correspond à une brève augmentation de la quantité de déoxyhémoglobine au début de la stimulation, comme montré dans des études en imagerie optique intrinsèque (pour un bref historique cf [Buxton, 2012]).

Ces deux phénomènes (initial dip et undershoot illustrés à la figure 4.9) soulignent la complexité des dynamiques métaboliques et fluidiques qui résultent en une mesure BOLD. Dans la section 5.3.2, nous présenterons différents modèles pour faire le lien entre activités neurales sous-jacentes et mesures BOLD, dont le Balloon-Windkessel model qui permet de s'intéresser plus particulièrement aux mécanismes hémodynamiques (5.3.2.2).

### 4.3.5 Limites

#### 4.3.5.1 Mesures indirectes

Les décours temporels BOLD mesurés pendant une tâche de contrôle inhibiteur ne sont pas des mesures directes de l'activité des neurones, ce qui nécessiterait des techniques d'imagerie invasives comme par exemple l'EEG intra-cranien<sup>10</sup> (cf discussion de [Levy and Anderson, 2012], [Cerf et al., 2010]). Ces mesures traduisent les variations des propriétés magnétiques du milieu dues à la réponse hémodynamique, dans des régions de l'espace qui contiennent de plus un très grand nombre de neurones.

---

8. Dans un centimètre cube de cortex, l'ordre de grandeur estimé est de mille milliards ( $10^{12}$ ) de synapses ([Drachman, 2005]). D'autres nombres caractéristiques sont donnés en introduction de 6.2

9. Nous ne détaillerons pas ici les détails techniques pour obtenir de telles mesures (sélection de coupes etc.).

10. MEA ou ME dans la figure 4.1

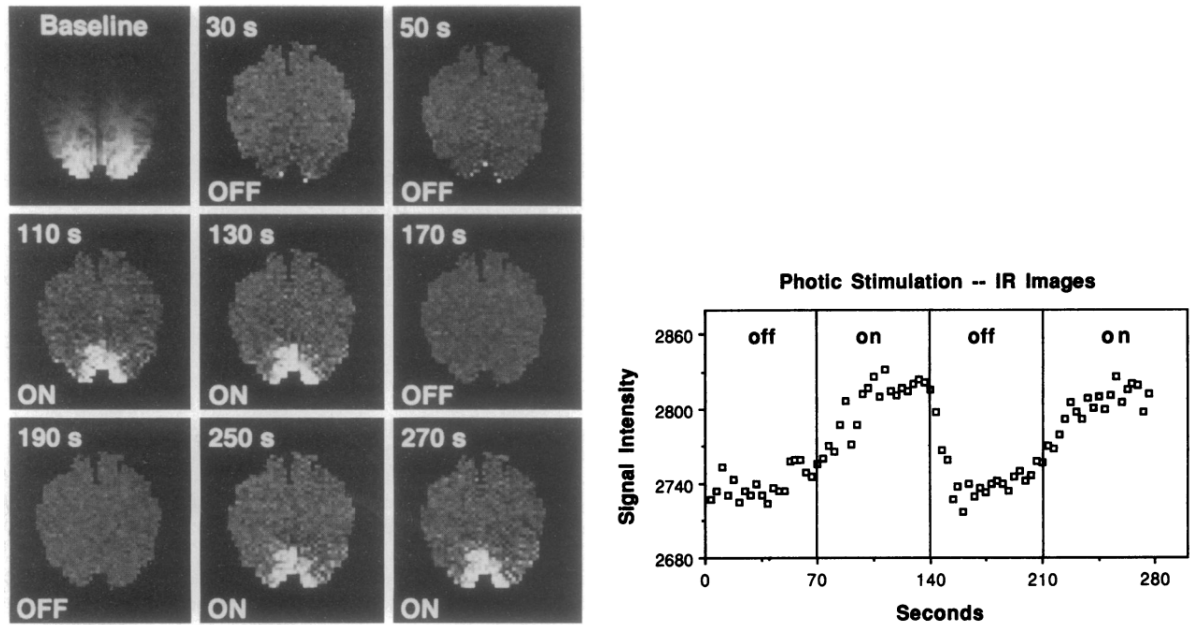


**Figure 4.8** – Activations du cortex visuel mesurées en IRMf pendant une tâche de stimulation visuelle ([Kwong et al., 1992]) Gauche : Mesures BOLD dans une coupe transversale du cerveau. La baseline est l'image de référence lorsque les lumières sont éteintes. Les images suivantes sont le résultat de la soustraction voxel à voxel entre l'image à  $t$  et la baseline. Droite : Décours temporel de l'activité BOLD dans une région du cortex visuel d'un sujet pendant la tâche de stimulation visuelle. Quand les diodes sont allumées (on), l'activité BOLD augmente dans le cortex visuel. Quand elles sont éteintes, le signal BOLD revient à l'état de base.

En outre, certains travaux pointent l'existence d'un décalage temporel entre l'activation neuronale et la réponse BOLD, qui varie en fonction de la variabilité de la réponse hémodynamique ([Andreelli and Mosbah, 2014]). Le niveau de sang oxygéné peut aussi augmenter en préparation de l'activation neuronale, ou pour d'autres raisons non liées directement aux neurones ([Smith, 2012]).

Enfin, les neurones actifs et les variations vasculaires ne sont pas exactement co-localisées (décalage spatial), ce qui peut se traduire par une erreur d'identification de la zone fonctionnellement activée ([Andreelli and Mosbah, 2014], [Uludağ et al., 2009]).

#### 4.3.5.2 Données très bruitées

Les variations de signaux BOLD mesurées pendant une tâche sont très faibles et les mesures sont très bruitées (faible ratio signal/bruit). C'est d'ailleurs pour cela qu'il est nécessaire de répéter la tâche pour améliorer la puissance statistiques des tests menées en routine.

Le bruit dans les mesures BOLD provient de différentes sources. Une partie du bruit est liée

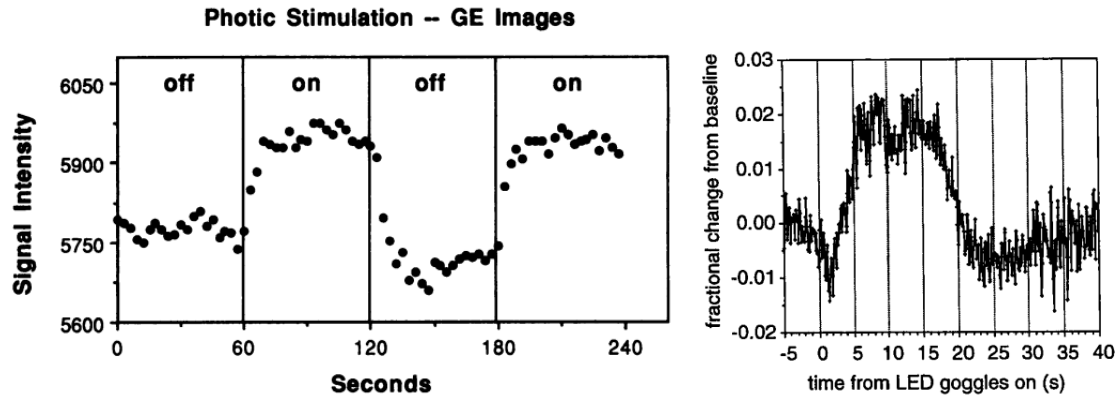


**Figure 4.9** – Initial dip et undershoot caractéristiques de la réponse BOLD.

Gauche : Décours temporel de l'activité BOLD d'un sujet pendant une tâche de stimulation visuelle. L'augmentation de signal BOLD pendant la stimulation est suivie d'une baisse en dessous du niveau de base (undershoot) ([Kwong et al., 1992])

Droite : Décours temporel de l'activité BOLD moyenné sur 5 participants. Initial dip, augmentation de l'activité pendant la stimulation puis undershoot ([Menon et al., 1995])

à la machine (bruits thermiques, force de l'aimant, inhomogénéités des champs magnétiques appliqués, résolution spatiale des voxels etc.). Une autre partie provient de facteurs physiologiques de la personne dans l'IRM (mouvements du corps et de la tête, bruits respiratoires et cardiaques, variabilité de la réponse hémodynamique etc.) et de facteurs comportementaux (variabilité de l'engagement cérébral pendant la tâche).

#### 4.3.5.3 Résolution temporelle

L'IRMf permet de prendre une mesure à intervalles de temps réguliers ( $TR$ ) de l'ordre de la seconde. Dans notre cas,  $TR = 2.05s$ , ce qui ne permet de saisir des phénomènes plus brefs.

Dans le cas du paradigme TNT par exemple, il ne serait pas possible de pouvoir mesurer une augmentation du signal BOLD dans la cible due à une intrusion mnésique avant une baisse résultant de la mise en place des mécanismes cérébraux de contrôle inhibiteur ([Levy and Anderson, 2012]).

# Chapitre 5

## Etudier la communication entre régions cérébrales à partir de signaux BOLD : nécessité de considérer un modèle dynamique du signal BOLD, des variations de l'activité neurale et des couplages inter-régions

*J'ai décrit au chapitre 4 la physique du signal BOLD. J'expose dans ce chapitre ce que je comprends, depuis ma formation de physicienne, de la manière dont il est exploité par la communauté de neuroimagerie.*

*Je situe pour commencer l'étude de l'intégration fonctionnelle par rapport à celle de la ségrégation fonctionnelle. Puis je précise les notions de connectivité fonctionnelle et effective qui permettent d'étudier la communication entre régions cérébrales, contrairement aux approches qui étudient la ségrégation fonctionnelle (section 5.1).*

*Je présente ensuite à la section 5.2 une analyse de connectivité fonctionnelle non basée sur un modèle du signal (coefficients de corrélation de Pearson).*

*Avant de présenter les méthodes de mesures de connectivité fonctionnelle et effective basées sur un modèle du signal, je prendrai le temps de décrire les différents modèles de l'activité neurale et BOLD au cours du temps qui sont couramment utilisés en IRMf. A la différence d'une présentation classique des méthodes d'analyses de signaux en IRMf, je souhaite séparer le travail de modélisation physique de celui de l'analyse de données par estimation de paramètres (formulation d'un modèle statistique de la mesure pour un sujet, technique d'inversion pour un sujet). Cette séparation entre modèle direct et inversion est au coeur du travail proposé dans cette thèse. La section 5.3 permet ainsi d'avoir une vision plus globale des modèles existants en IRMf et découplée des méthodes d'analyses de données. Cela facilitera par la suite le travail de modélisation en permettant des choix éclairés qui partent de l'existant.*

*Nous verrons finalement comment ces modèles dynamiques sont utilisés pour travailler avec*

des données BOLD dans le cadre d'analyses de connectivité fonctionnelle (PPI, cf section 5.4), et effective (DCM, cf section 5.5). Je me limiterai à la présentation des méthodes d'analyses pour un participant, puisque l'objet de mon travail de thèse va consister à construire une procédure d'estimation de paramètres pour un participant. Je ne détaillerai donc pas les méthodes statistiques employées pour analyser les données à l'échelle des groupes de participants.

Cet état de l'art des méthodes classiques d'analyse de connectivités permet de :

1. mieux comprendre les résultats obtenus en IRMf pendant une tâche de contrôle inhibiteur, notamment de contrôle inhibiteur des intrusions mnésiques que nous présenterons au fur et à mesure après chaque méthode
2. comprendre pourquoi elles ne nous suffisent pas et pourquoi nous avons besoin de travailler au développement d'une nouvelle méthode d'analyse des signaux BOLD pour étudier les mécanismes neuraux du contrôle inhibiteur.

## 5.1 Ségrégation et intégration fonctionnelle

Historiquement, les neurosciences se sont d'abord concentrées sur la recherche de la localisation précise de régions cérébrales qui en s'activant assuraient la réalisation de fonctions cognitives particulières ([Stephan et al., 2004], pour un historique voir [Friston, 2011]). On parle aussi de localisationnisme : une fonction peut être localisée dans une région cérébrale précise.

Aujourd'hui, il est admis par la communauté que la réalisation d'une fonction cognitive serait permise par la communication d'informations entre régions cérébrales, qui fonctionnent de manière couplée. On parle aussi de processus d'intégration fonctionnelle. Des régions réparties dans le cerveau peuvent néanmoins être spécialisées dans certaines opérations élémentaires qui décomposent la fonction cognitive. On parle de ségrégation fonctionnelle. Le fonctionnement en réseau de ces régions spécialisées permettrait à la fonction cognitive d'émerger ([Friston, 2011]).

### 5.1.1 Etude de la ségrégation fonctionnelle

#### 5.1.1.1 Principe

L'étude de la ségrégation fonctionnelle cherche à identifier les régions qui sont activées quand le sujet réalise une tâche particulière. Autrement dit, pour chaque voxel, on se demande si le niveau BOLD est plus élevé pendant la tâche d'intérêt par rapport à une autre tâche de référence. Les voxels sont ici pensés de manière découplée : la série temporelle d'un voxel ne dépend pas de celle d'autres voxels. Deux types d'analyses statistiques différentes peuvent être effectuées chez un sujet : des analyses de soustraction et des modèles statistiques de régressions linéaires.

Les analyses de soustraction ont recours à des calculs de contrastes (par exemple la différence de moyennes hors période de stimulation et en période de stimulation) pour chaque voxel/région du cerveau.

Pour les régressions linéaires, il est d'abord nécessaire de construire des régresseurs qui modélisent respectivement les signaux BOLD prédits pour chaque condition expérimentale. Ils sont obtenus à partir d'un modèle de convolution (par la Fonction de Réponse Hémodynamique ou HRF), que nous reverrons au paragraphe 5.3.2.1). Un modèle statistique linéaire de la mesure dans chaque voxel est ensuite construit à partir de ces regresseurs. Le signal BOLD prédit dans un voxel est une combinaison linéaire des signaux BOLD prédits pour chaque condition expérimentale, à laquelle peuvent s'ajouter d'autres facteurs confondants. L'objectif est d'estimer les amplitudes (pentes) des réponses BOLD prédites pour chaque condition expérimentale, en minimisant l'écart entre les signaux BOLD prédit et mesuré (par la méthode des moindres carrés). Ce sont des modèles statistiques univariés (i.e. indépendants les uns des autres). Les régressions sont effectuées voxels par voxels de manière indépendante et permettent d'obtenir une carte paramétrique pour un sujet.

Des statistiques univariées de groupe permettent ensuite de déterminer si les variations du signal pendant la tâche sont statistiquement significatives à l'échelle d'un groupe. Ces tests se font sur les contrastes pour chaque voxel ou sur les paramètres estimés des régressions linéaires pour chaque voxel. Cela permet de générer des cartes statistiques d'activation avec dans chaque voxel le résultat de significativité du test statistique de groupe. Quand ces cartes d'activation sont superposées à une IRM structurelle, nous obtenons une illustration des régions cérébrales qui ont été significativement activées pendant la tâche (e.g. figure 15.1 dans l'annexe 15.1). C'est l'image "classique" d'une IRM fonctionnelle souvent présentée dans les travaux de médiation scientifique.

Néanmoins ces approches ne permettent pas d'étudier l'interaction de différentes régions cérébrales. D'autres types d'analyses permettent d'étudier le fonctionnement en réseau (i.e. couplé) de différentes parties du cerveau. Elles éclairent sur deux types de connectivité différentes : la connectivité fonctionnelle et la connectivité effective.

#### 5.1.1.2 Exemple du contrôle inhibiteur des intrusions mnésiques

L'objectif de ce chapitre n'est pas de présenter en détails les méthodes utilisées pour étudier la ségrégation fonctionnelle, ni les résultats obtenus lorsqu'elles sont employées pour étudier le contrôle inhibiteur des intrusions mnésiques. Nous trouverons dans l'annexe 15.1 les principaux résultats des études qui s'intéressent aux contrastes BOLD pendant une tâche de TNT.

Nous pouvons seulement retenir que pendant le contrôle inhibiteur des intrusions mnésiques, l'activité BOLD de régions de contrôle (DLPFC et VLPFC droits) est augmentée et que celle de l'hippocampe droit est abaissée. Ces variations sont encore plus grande quand le participant fait l'expérience d'une intrusion mnésique. Le DLPFC pourrait ainsi exercer un contrôle inhibiteur sur l'hippocampe, mais ces mesures de contrastes ne permettent pas d'étudier la communication entre ces régions.

Nous notons de manière intéressante que ces contrastes dépendent de la concentration de GABA dans l'hippocampe mais pas dans le DLPFC : plus la concentration de GABA hippocampique est grande et plus l'activité de l'hippocampe est abaissée ([Schmitz et al., 2017]).

Pour étudier les mécanismes cérébraux du contrôle inhibiteur, la question n'est pas seule-

ment de connaître les variations d'activité des régions de manière indépendante. Nous nous intéressons surtout aux influences réciproques. Pour montrer qu'une baisse d'activité dans l'hippocampe est liée à une augmentation d'activité dans les régions préfrontales, d'autres méthodes plus élaborées permettent d'analyser les signaux BOLD pour étudier la communication entre régions cérébrales.

## 5.1.2 Etude de l'intégration fonctionnelle

### 5.1.2.1 Connectivité fonctionnelle

La connectivité fonctionnelle entre deux régions est une mesure de la synchronie des séries temporelles BOLD de ces deux régions<sup>1</sup>. C'est une mesure de la dépendance statistique des décours temporels.

En postulant que les variations BOLD suivent les variations neurales sous-jacentes, deux régions synchrones au niveau BOLD fonctionneraient ensemble au niveau neural lors de la réalisation d'une tâche cognitive.

Même si deux régions sont fonctionnellement connectées, nous ne pouvons rien dire des mécanismes par lesquels les deux régions travaillent ensemble ([Friston, 2011]). La figure 5.1 représente différents scénarios possibles. Ces deux régions peuvent être directement connectées par un faisceau de substance blanche (première ligne) ou non (deuxième ligne), d'autres régions servant alors de relais.

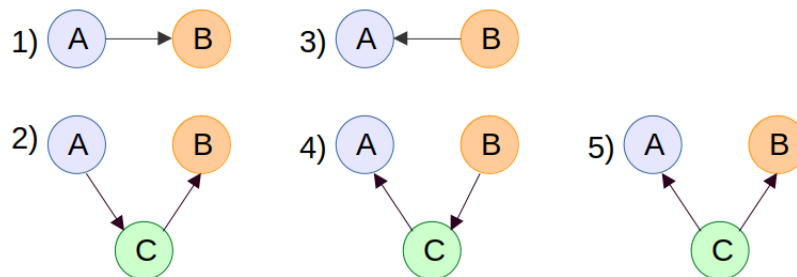


**Figure 5.1** – Scénarios possibles expliquant l'existence d'une connectivité fonctionnelle entre deux régions. Elles peuvent recevoir du signal l'une de l'autre de manière directe (1 et 3) ou par l'intermédiaire d'un relais (2 et 4), ou bien être reliées à une autre troisième région qui envoie un signal commun de manière parallèle (5).

Plusieurs techniques permettent d'étudier la connectivité fonctionnelle. Elles reposent ou non sur un modèle du signal ("model free and model-based approach" [Friston, 2011]) pour construire un modèle statistique de la dépendance des décours temporels. Nous en présentons une de chaque : coefficients de corrélation de Pearson (5.2) et analyses PPI (5.4).

1. Pour chaque région, on peut extraire l'activité moyenne des voxels qui composent la région. Ce qui compte c'est de travailler avec des paires de séries temporelles.



### 5.1.2.2 Connectivité effective

La connectivité effective permet d'étudier l'influence causale que l'activité neurale d'une région A exerce sur une autre région B ([Friston, 2011]). Ces deux régions peuvent être connectées directement ou indirectement par des fibres de substance blanche<sup>2</sup>.

Certaines approches modélisent uniquement les variations du signal BOLD au cours du temps, en supposant qu'elles suivent les variations neurales sous-jacentes (e.g. [Gu et al., 2017]). Les couplages entre régions sont directement exprimés au niveau BOLD. D'autres approches modélisent l'activité neurale et BOLD au cours du temps (cf 5.5). Les couplages entre régions sont alors pensés au niveau neural (e.g. modèle de DCM présenté à la section 5.3.1.2). La mesure de connectivité effective est la mesure du couplage de A vers B ([Gilson et al., 2019]). Un modèle statistique de la mesure basé sur ces modèles dynamiques permet ensuite d'estimer à partir des données les paramètres de couplages d'intérêt.

Nous détaillerons à la section 5.5, l'approche DCM (pour Dynamic Causal Modeling) couramment utilisée en IRM fonctionnelle.

## 5.2 Analyses de connectivité fonctionnelle non basée sur un modèle dynamique du signal

### 5.2.1 Connectivité fonctionnelle entre deux régions

Le degré de synchronie de l'activité de deux régions est mesuré par le coefficient de corrélation de Pearson ( $r$ ) directement calculé entre les séries temporelles BOLD des deux régions (cf [Gilson et al., 2019], en particulier leur appendix B). Par construction, la corrélation entre 2 séries temporelles est comprise entre -1 et 1.

- $r = 1$  : les deux séries sont identiques (à leur moyenne ou variance près). Plus  $r$  est proche de 1, plus les 2 régions sont impliquées dans des processus cognitifs similaires.
- $r = 0$  : les deux séries sont statistiquement indépendantes. Plus  $r$  est proche de 0 (en valeur absolue), plus les processus cognitifs sont indépendants.
- $r = -1$  : les deux séries sont en opposition de phase. Plus  $r$  est négatif (proche de -1), plus les processus cognitifs sont mutuellement exclusifs.

### 5.2.2 Carte de connectivité fonctionnelle

Une carte de connectivité fonctionnelle associée à une région, dite région cible (ou target), est le vecteur des mesures de corrélations entre le décours temporel de cette région cible et tour à tour chacune des régions du cerveau. Ce sont les mesures de connectivité fonctionnelle entre la région cible et toutes les autres régions du cerveau.

L'ensemble des régions fonctionnellement connectées à la région cible forme un réseau fonctionnel.

---

2. comme rapporté par [Benoit et al., 2015] et [Schmitz et al., 2017] pour DCM, section 5.5



### 5.2.3 Connectome et réseau fonctionnel

Le connectome fonctionnel est la matrice qui regroupe les cartes de connectivité pour chaque région du cerveau. A chaque ligne et à chaque colonne correspond une région. Un élément de la matrice, que nous pouvons noter  $r_{ij}$ , est égal au coefficient de corrélation de Pearson entre les séries temporelles des régions  $i$  et  $j$ .

A partir du connectome, des algorithmes de clusterisation peuvent permettre d'identifier des groupes de régions qui présentent des coefficients de corrélation élevés au sein du groupe et faibles avec le reste du cerveau. Un tel groupe forme un réseau fonctionnel : un ensemble de régions fortement connectées fonctionnellement entre elles et peu avec les autres régions. Les régions d'un réseau fonctionnel seraient impliquées dans des processus cognitifs similaires ([Sporns, 2013])

### 5.2.4 Limites

L'étude des interactions entre régions à partir de mesures de connectivités fonctionnelles est limitée :

1. Même si deux régions sont fonctionnellement connectées, nous ne pouvons rien dire des mécanismes par lesquels les deux régions travaillent ensemble ([Friston, 2011]). Les deux régions peuvent être connectées par des fibres de substance blanche de manière directe ou indirecte (d'autres régions servant de relais). Elles peuvent recevoir du signal l'une de l'autre ou bien être reliées à une autre troisième région qui envoie un signal commun de manière parallèle (cf figure 5.1).
2. Ces analyses se basent uniquement sur les mesures BOLD et postulent que les variations neurales sous-jacentes, qui sont celles d'intérêt, suivent les variations du BOLD. Comme nous l'avons vu au chapitre 4, le signal BOLD est une mesure très indirecte de l'activité neurale sous-jacente. Les réponses hémodynamiques peuvent aussi être variables d'une région à l'autre ([Buckner et al., 1996]). Les mesures de connectivité fonctionnelle reflètent donc de manière partielle la synchronie de l'activité neurale.
3. En fonction des techniques de pré-traitements d'imagerie utilisées, il peut rester dans les signaux BOLD plus ou moins de bruit physiologique (cardiaque, respiratoire) ou lié au mouvement qui modifient les mesures de corrélation ([Power et al., 2018], [Parkes et al., 2018]).
4. Les mesures de connectivités fonctionnelles entre deux régions varient au cours du temps lorsqu'on isole des fenêtres d'activité pour un même sujet (et également entre individus). L'étude de la variabilité intra et inter-individuelle des cartes de connectivité fonctionnelle reste un domaine de recherche actif ([Mueller et al., 2013]).

Pour améliorer notre compréhension des interactions entre régions cérébrales, il semble nécessaire de disposer de modèles statistiques plus précis de la mesure et des échanges inter-régions. Ces modèles statistiques vont être construits à partir de modèles dynamiques du signal BOLD.

## 5.3 Les modèles dynamiques des variations du signal BOLD

Il existe des modèles dynamiques qui se basent uniquement sur les variations du BOLD au cours du temps, sans modèle explicite de la dynamique neurale sous-jacente. Ce sont notamment des modèles phénoménologiques issus de la théorie du contrôle linéaire (e.g. [Gu et al., 2017], multivariate Ornstein-Uhlenbeck process dans [Gilson et al., 2019]). Ces modèles se basent sur le postulat que les variations BOLD reflètent les variations neurales. Les échanges d'activité neurale entre régions sont directement traduits au niveau du BOLD sous la forme de couplages entre les activités BOLD des régions. Cependant, les liens entre activité neurale et signal BOLD sont complexes et encore partiellement incompris ([Stephan et al., 2007]).

Pour améliorer notre compréhension des mécanismes neuraux sous-jacents, il est pertinent de travailler avec des modèles explicites des variations de l'activité neurale pendant une tâche, pour arriver jusqu'à un modèle des variations du signal BOLD. Dans ce paragraphe, nous séparons les différents niveaux de modélisation dans un ordre direct :

- un modèle de stimulation neurale pendant une tâche
- un modèle de variations de l'activité neurale au cours du temps en fonction d'interactions entre les régions cérébrales
- un modèle de signal BOLD avec ou sans explicitation du couplage neurovasculaire

Nous essaierons de séparer à chaque fois l'énoncé des hypothèses de modélisation de leur traduction en langage mathématique.

L'observable reste le signal BOLD. Toutes les autres variables d'états sont non accessibles directement par la mesure (on pourra aussi parler d'états cachés ou hidden states en anglais).

### 5.3.1 De la stimulation expérimentale aux variations de l'activité neurale

#### 5.3.1.1 Un modèle des variations de l'activité neurale pendant une tâche : la Stickfunction

Une condition expérimentale consiste en l'exécution d'une tâche particulière par un sujet dans l'IRM (pendant des périodes dites de stimulation). Elle doit toujours être comparée à une situation de référence où le sujet n'exécute pas la tâche (périodes hors stimulation).

Les variations de l'activité neurale d'une région pendant une tâche peuvent être modélisées par l'hypothèse de modélisation suivante : augmentation d'activité neurale seulement en période de stimulation.

Notons  $\Delta t_{stim}$  (resp.  $\Delta t_{\overline{stim}}$ ) l'intervalle qui regroupe l'ensemble des périodes de stimulation (resp. hors stimulation). Cette hypothèse peut se traduire mathématiquement par une fonction créneau (aussi appelée "stickfunction" en anglais), notée  $U$  telle que :

$$U(t) = \begin{cases} 1 & \text{si } t \in \Delta t_{stim} \text{ (période de stimulation)} \\ 0 & \text{si } t \in \Delta t_{\overline{stim}} \text{ (hors période de stimulation)} \end{cases} \quad (5.1)$$

### 5.3.1.2 Modèle phénoménologique linéaire de l'activité neurale (DCM)

Ce modèle a été introduit par [Friston et al., 2003] et est utilisé couramment dans une méthode d'analyse appelée Dynamic Causal Modeling (DCM), qui sera présentée plus en détails à la section 5.5. Il s'agit d'un modèle phénoménologique de l'activité neurale i.e. dont les hypothèses de modélisation ne portent pas sur des mécanismes biologiques d'émergence de l'activité neurale. Il est linéaire et donne l'évolution au cours du temps de l'activité neurale excitatrice d'une région cérébrale, en fonction des couplages inter-régions (qui évoluent au cours du temps) et de forçages appliqués dans certaines régions du réseau pendant une période de stimulation. L'activité neurale est pensée comme le résultat global de processus excitateurs et inhibiteurs non explicites. Les couplages inter-régions ne représentent pas nécessairement des connexions structurelles, mais plutôt des influences causales d'une région vers une autre qui peuvent passer par des relais ([Benoit and Anderson, 2012],[Schmitz et al., 2017]). Dans la suite du paragraphe, nous nous redonnons pas à pas les différentes hypothèses de modélisation et leur traduction mathématique (le modèle sans stimulation, puis avec une et plusieurs conditions expérimentales).

#### 5.3.1.2.1 Modèle sans stimulation

Soient  $N$  régions cérébrales. Chaque région  $i$  est caractérisée par un descripteur : son activité neurale au cours du temps notée  $s_i(t)$ . Nous notons :

- $\vec{s}(t)$  : le vecteur des activités neuronales, de dimension  $N \times 1$
- $\overline{\overline{A}}$  : la matrice de connectivité intrinsèque ("intrinsic connectivity matrix") de dimensions  $N \times N$ , qui donne les couplages entre régions et les auto-couplages("intrinsic/self connexion") dans une situation de référence, i.e. en l'absence de toute manipulation expérimentale. Ceux-ci peuvent être positifs ou négatifs.

L'évolution des activités neuronales au cours du temps dans la situation de référence est donnée par l'équation différentielle homogène suivante :

$$\dot{\vec{s}}(t) = \overline{\overline{A}}\vec{s}(t) \tag{5.2}$$

#### 5.3.1.2.2 Modèle de l'effet produit par une condition expérimentale

Une condition expérimentale produit 2 types d'effets :

- Ajout d'un forçage positif ou négatif dans certaines régions
- Modification de la constante de couplage entre certaines régions

Pour traduire mathématiquement ces hypothèses de modélisation, nous avons besoin de noter :

- $u(t)$  : une fonction créneau qui vaut 1 pendant la condition expérimentale et 0 sinon
- $\vec{C}$  : vecteur de dimension dimension  $N \times 1$  qui donne l'intensité des forçages ("inputs") appliqués sur chaque région en période de stimulation

- $\overline{\overline{B}}$  : matrice des effets modulateurs, de dimension  $N \times N$  ("modulatory effects"). Chaque élément  $B_{mn}$  est égal à l'intensité (positive ou négative) du couplage supplémentaire de la région  $n$  vers  $m$  appliqué pendant la stimulation. Celui-ci s'ajoute au couplage  $A_{mn}$  hors stimulation.

L'évolution des activités neurales au cours du temps est donnée par l'équation différentielle suivante :

$$\dot{\vec{s}}(t) = (\overline{\overline{A}} + u(t)\overline{\overline{B}})\vec{s}(t) + \vec{C}u(t) \quad (5.3)$$

Le vocabulaire qui désigne les matrices  $\overline{\overline{A}}$  et  $\overline{\overline{B}}$  peut varier entre les articles et nous recommandons d'y faire attention à la lecture pour savoir précisément de quels couplages parlent les auteurs. Par exemple dans [Stephan et al., 2004], les éléments de la matrice  $\overline{\overline{A}}$  sont les valeurs de "connectivité effective" entre régions en l'absence de stimulation. La matrice  $\overline{\overline{B}}$  donne les "changements de connectivité effective" pendant la tâche. Dans [Benoit and Anderson, 2012], la matrice  $\overline{\overline{B}}$  donne les valeurs de "connectivité effective" et la somme de  $\overline{\overline{A}} + \overline{\overline{B}}$  donne celles de "connectivité effective absolue".

### 5.3.1.2.3 Modèle des effets produits par plusieurs conditions expérimentales

Soit  $\mathcal{N}_E$  le nombre de conditions expérimentales prises en compte dans le modèle.

Chaque condition expérimentale peut produire les 2 types d'effets précédemment cités, indépendamment des autres conditions. Pour une condition expérimentale  $j$ , nous notons :

- $u_j(t)$  : une fonction créneau qui vaut 1 en période de stimulation et 0 hors stimulation pour la condition expérimentale  $j$ .  
La fonction de stimulation  $u_j(t)$  donne la  $j$ -ème ligne du vecteur  $\vec{u}(t)$ . En notant,  $\vec{e}_j$  un vecteur de 0 avec 1 à la  $j$ -ème composante, on peut écrire :  $\vec{u}(t) = \sum_{j=1}^{\mathcal{N}_E} u_j(t)\vec{e}_j$
- $\vec{C}^j$  : intensité des forçages appliqués sur les régions pendant la condition expérimentale  $j$ .  
Nous notons  $\overline{\overline{C}}$  la matrice des forçages, de dimension  $N \times \mathcal{N}_E$ . Chaque colonne  $j$  est égale à  $\vec{C}^j$ .
- $\overline{\overline{B}}^j$  : matrice des effets modulateurs associée à la condition expérimentale  $j$ , de dimension  $N \times N$  ("modulatory effects")  
Chaque élément  $B_{mn}^j$  est égal à l'intensité (positive ou négative) du couplage supplémentaire de  $n$  vers  $m$  appliqué pendant la condition expérimentale  $j$ .

L'évolution des activités neurales au cours du temps est donnée par :

$$\dot{\vec{s}}(t) = \left(\overline{\overline{A}} + \sum_{j=1}^{\mathcal{N}_E} u_j(t)\overline{\overline{B}}^j\right)\vec{s}(t) + \sum_{j=1}^{\mathcal{N}_E} \vec{C}^j u_j(t) = \left(\overline{\overline{A}} + \sum_{j=1}^{\mathcal{N}_E} u_j(t)\overline{\overline{B}}^j\right)\vec{s}(t) + \overline{\overline{C}}\vec{u}(t) \quad (5.4)$$

Le terme  $\overline{\overline{\kappa}}(t) = \left(\overline{\overline{A}} + \sum_{j=1}^{\mathcal{N}_E} u_j(t)\overline{\overline{B}}^j\right)$  contient tous les couplages inter-régions au cours du temps (en fonction des différentes conditions expérimentales). Dans certains articles,

les représentations graphiques de ce modèle neural ne font parfois pas apparaître tous les couplages, mais uniquement les variations de couplages contenues dans les matrices  $\overline{\overline{B^j}}$ , ainsi que les forçages  $\vec{C^j}$  pour chaque condition expérimentale  $j$  (e.g. [Apšvalka et al., 2022]). Il peut exister un biais de représentation à prendre en compte à la lecture d’articles (nous y reviendrons en 13.2.1.2.2).

## 5.3.2 De l’activité neurale au signal BOLD

### 5.3.2.1 Sans modèle explicite du couplage neurovasculaire

Ce modèle phénoménologique est utilisé de manière standard par la communauté de neuroimagerie ([Friston, 2011]) dans des analyses statistiques univariées<sup>3</sup> ([Worsley and Friston, 1995]) ou multivariées (que nous détaillerons dans la partie 5.4).

Il se base sur un modèle linéaire, causal et invariant de la mesure BOLD suite à des variations de l’activité neurale. Plus précisément, cela signifie que :

- modèle causal : les variations du BOLD (sortie) sont dues aux variations de l’activité neurale (entrée). Elles ne peuvent les précéder dans le temps.
- modèle invariant dans le temps : la réponse du système à une même entrée reste toujours identique, peu importe l’instant de la sollicitation.
- modèle linéaire : la réponse à une combinaison linéaire de plusieurs entrées est égale à la combinaison linéaire des réponses de chaque entrée individuelle.

Les paradigmes événementiels en IRMf permettent de mesurer la réponse BOLD suite à une stimulation de courte durée et d’estimer une fonction de réponse hémodynamique ([Rosen et al., 1998], [Buckner, 1998], [Boynton et al., 1996]). [Dale and Buckner, 1997] ont montré que les hypothèses de linéarité, causalité et invariance étaient assez bien respectées dans le contexte de leur étude (cf figure 5.2).

Ces trois hypothèses de modélisation permettent d’élaborer un signal temporel de sortie (BOLD) en fonction d’un signal temporel en entrée (activité neurale). Ce sont les deux seules variables d’état du modèle. En effet, dans la théorie des systèmes, un système linéaire, causal et invariant est entièrement décrit par sa réponse impulsionnelle (réponse à une impulsion représentée par un delta de Dirac). Toute réponse à un signal d’entrée plus complexe pourra ensuite être calculée à partir de la réponse impulsionnelle en le décomposant comme une combinaison linéaire d’impulsions (cf paragraphe 5.3.2.1.2).

#### 5.3.2.1.1 Fonction de réponse hémodynamique ou HRF

La fonction dite de réponse hémodynamique (ou HRF pour Hemodynamic Response Function en anglais) est la réponse impulsionnelle du système. Elle donne les variations de la mesure BOLD (i.e. réponse hémodynamique mesurée) suite à une activation ponctuelle dans le temps (ou impulsion) des neurones sous-jacents, représentée par une distribution delta de Dirac.

Cette fonction canonique, représentée à la figure 5.3, s’écrit comme la somme pondérée de

---

3. Nous y avons fait référence à la section 5.1

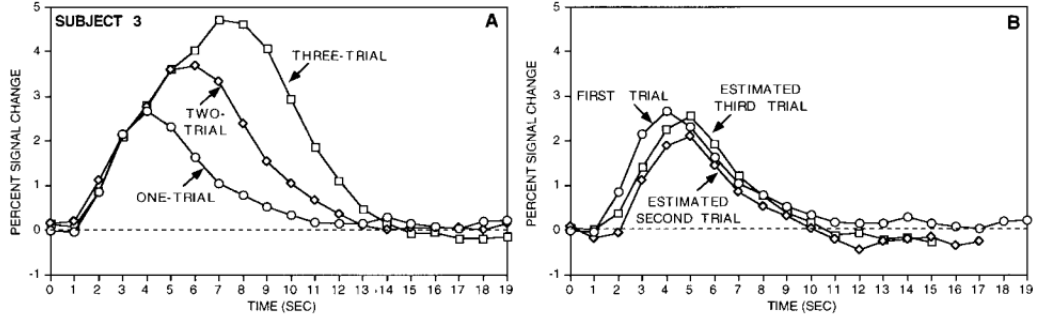


**Figure 5.2** – Linéarité de la réponse BOLD. A gauche : Mesures BOLD obtenues après une, deux ou trois stimulations visuelles brèves (notées 1 trial, 2 trials et 3 trials), de durée 1 seconde, séparées de 1 seconde. La réponse augmente avec le nombre de stimulations. A droite : soustraction des réponses (3 trials - 2 trials, et 2 trials - 1 trial) pour obtenir la contribution d'une stimulation. La réponse mesurée pour 1 trial, et celles calculées par soustraction sont quasiment identiques. La réponse BOLD pour 2 et 3 trials est une somme des réponses pour 1 trial, ce qui montre la linéarité ([Dale and Buckner, 1997]).

deux distributions Gamma ([Friston et al., 1998]), de paramètres  $\alpha_1 = 6$ ,  $\alpha_2 = 16$ ,  $\beta_1 = \beta_2 = 1$  ([Lindquist et al., 2009] appendix A). Le deuxième terme est pondéré par  $c = \frac{1}{6}$ . L'expression analytique de la distribution Gamma se trouve en annexe 17.1.1.

$$h(t) = G(t, \alpha_1, \beta_1) - c G(t, \alpha_2, \beta_2) = \frac{t^{\alpha_1-1} \beta_1^{\alpha_1} e^{-\beta_1 t}}{(\alpha_1 - 1)!} - c \frac{t^{\alpha_2-1} \beta_2^{\alpha_2} e^{-\beta_2 t}}{(\alpha_2 - 1)!} \quad (5.5)$$

Une impulsion neurale (à  $t = 0s$ ) produit la réponse canonique illustrée par la courbe rouge sur le graphe de la figure 5.3. Ce signal BOLD correspond aux variations par rapport au niveau précédent la stimulation. Elle part de 0 (ligne de base) et augmente à partir d'environ 1 seconde pour atteindre un maximum vers 5 secondes après l'impulsion. Notons que la variation de signal est assez faible. La fonction décroît ensuite et intercepte la ligne de base au bout de 12 secondes environ, avant de passer en dessous. La valeur minimale de l'undershoot est atteinte aux alentours de 16 secondes. Le retour à la ligne de base marque la fin de la réponse BOLD a lieu au bout de 25 secondes à 30 secondes environ. La réponse BOLD est donc lente (i.e. sa durée est grande devant celle de l'impulsion) et décalée dans le temps par rapport à la réponse neurale.

### 5.3.2.1.2 Du neural au BOLD par convolution

Soit  $n(t)$  un signal neural en entrée et  $y(t)$  le signal BOLD correspondant en sortie.  $n(t)$  est décomposable comme une combinaison linéaire d'impulsions de Dirac :

$$\forall t > 0, n(t) = \int_{-\infty}^{+\infty} n(\tau) \delta(t - \tau) d\tau \quad (5.6)$$

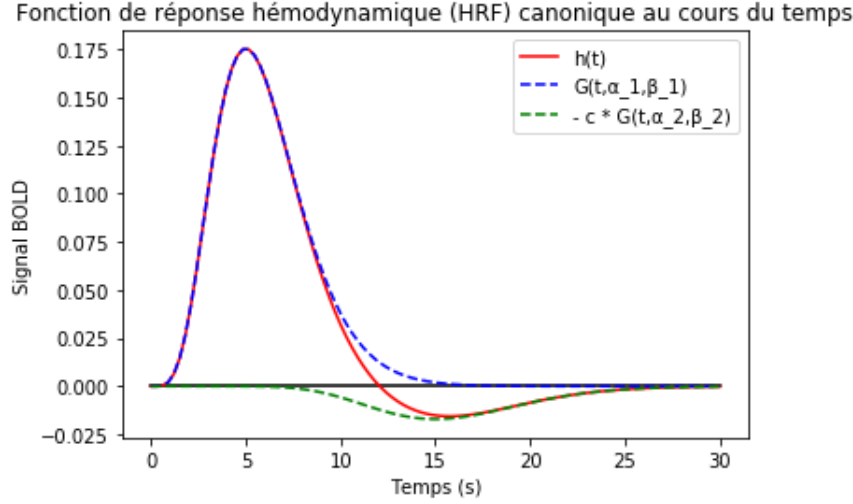


**Figure 5.3** – Représentation graphique de la Fonction de Réponse Hémodynamique. La HRF est une somme de deux fonctions Gamma. La deuxième fonction Gamma permet de modéliser l’undershoot de la réponse hémodynamique.

La réponse impulsionnelle correspond à la réponse d’un filtre linéaire et indépendant du temps, noté  $f$ , à une impulsion de Dirac :  $\forall \hat{t} > 0, f(\delta(\hat{t})) = h(\hat{t})$ . En appliquant l’opération de filtrage sur  $n$ , le signal BOLD  $y$  s’écrit :

$$\forall t > 0, y(t) = f(n(t)) = \int_{-\infty}^{+\infty} f(n(\tau)\delta(t - \tau))d\tau \quad (5.7)$$

$$= \int_{-\infty}^{+\infty} n(\tau)f(\delta(t - \tau))d\tau \text{ (linéarité)} \quad (5.8)$$

$$= \int_{-\infty}^{+\infty} n(\tau)(f \circ \delta)(t - \tau))d\tau \text{ (indépendance du temps)} \quad (5.9)$$

$$= \int_{-\infty}^{+\infty} n(\tau)h(t - \tau))d\tau := (n * h)(t) \quad (5.10)$$

On dit que la réponse BOLD est égale à l’activité neurale<sup>4</sup> filtrée par la réponse hémodynamique ([Buxton, 2012]). C’est une combinaison linéaire des réponses à chaque impulsion individuelle, qui s’exprime mathématiquement sous la forme d’un produit de convolution de  $n$  par la réponse impulsionnelle du système. La déconvolution de  $y$  par la HRF permet de passer des variations du signal BOLD  $y$  aux variations de l’activité neurale  $n$ .

#### Limites :

Avec ce modèle, l’ensemble des variations vasculaires causées par les variations de l’activité des neurones sont implicites. Elles constituent une boîte noire qui conduit à la réponse BOLD mesurée. Ce modèle ne permet donc pas d’étudier les mécanismes hémodynamiques de la réponse BOLD ([Stephan et al., 2004], [Buxton, 2012]).

4. "Exactly what was meant by neural activity was somewhat vague and I think it was often assumed to be local spiking activity" ([Buxton, 2012]).



### 5.3.2.2 Avec un modèle explicite du couplage neurovasculaire

Ce modèle non linéaire est un modèle biophysique qui donne les variations de signaux BOLD au cours du temps en fonction de grandeurs hémodynamiques. Ce modèle sera utilisé dans la suite de la thèse et nous prenons donc le temps de l'expliquer en détails.

#### 5.3.2.2.1 Modèle hémodynamique

##### 5.3.2.2.1.1 Motivations et objectifs

La mesure de la réponse BOLD qui suit une stimulation présente des caractéristiques non linéaires qui diffèrent de la forme de la stimulation ([Friston et al., 2000], [Obata et al., 2004]). Celles-ci comprennent par exemple une baisse du signal par rapport au niveau de base ("early dip"), un pic de réponse avant stabilisation à un niveau actif supérieur au niveau de base ("initial overshoot") puis à l'arrêt de la stimulation une baisse du signal en dessous du niveau de base ("post-stimulus undershoot").

Certaines expériences avec deux stimulations consécutives brèves montrent que la réponse BOLD mesurée n'est pas égale à la somme des deux réponses pour chaque stimulation individuelle ([Friston et al., 2000]). Cette modification de la réponse BOLD à une stimulation lorsque celle-ci est précédée d'une autre stimulation ("hemodynamic refractoriness") est un autre exemple de non linéarité.

Les variations de concentration en déoxyhémoglobine, à l'origine des variations du signal BOLD, sont liées à différents mécanismes hémodynamiques (variations locales au cours du temps du flux et volumes sanguins) ainsi que métaboliques (extraction de dioxygène au cours du temps par les tissus nerveux)([Buxton, 2012]). Un modèle explicite et quantitatif du couplage neurovasculaire permet de préciser ces mécanismes et leurs interactions pour mieux comprendre d'une part comment la variation de l'activité neurale conduit à la réponse BOLD mesurée ([Stephan et al., 2004], [Buxton, 2012]) et d'autre part d'où proviennent les aspects non linéaires de cette réponse BOLD ([Buxton et al., 1998], [Obata et al., 2004]).

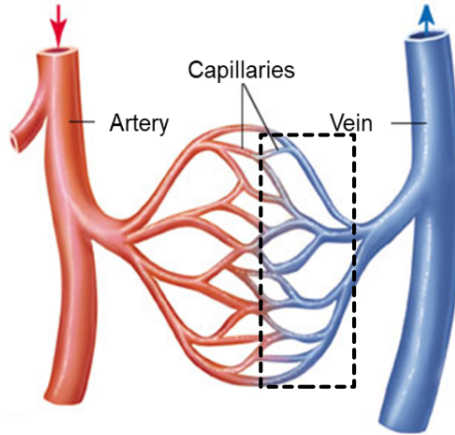
Ceux-ci seraient dû aux décalages temporels entre les variations de flux, volume et d'extraction de dioxygène ([Mandeville et al., 1999], [Obata et al., 2004]), bien plus problemement qu'au décours temporel de l'activité neurale sous jacente.

[Friston et al., 2000] notent que le modèle que nous présentons dans cette partie permet de rendre compte des non linéairités les plus importantes de la réponse BOLD.

##### 5.3.2.2.1.2 Présentation du modèle et des variables d'état

Ce modèle du couplage neurovasculaire se situe au niveau du compartiment veineux (veinules après les artériolles), représenté sur le schéma ci-dessous, où se trouve la majorité de la quantité de déoxyhémoglobine. Les volumes des artériolles sont fixes, mais le volume du compartiment veineux peut changer ([Buxton et al., 1998]).





L'augmentation de l'activité des neurones entraîne une augmentation du flux sanguin au niveau des artériolles, i.e. en entrée du compartiment veineux. Celui-ci se comporte comme un ballon extensible ([Obata et al., 2004]) traversé par le sang. Il se dilate avec l'augmentation du flux de sang oxygéné en entrée, qui a par ailleurs pour effet de diluer la déoxyhémoglobine contenue dans le ballon ([Stephan et al., 2004]). Il se vide en expulsant le sang du fait de sa résistance à la déformation. Le sang oxygéné qui y transite est extrait du ballon par les neurones à proximité, faisant également varier la quantité de déoxyhémoglobine contenue dans le ballon. Les variations de volume et de quantité de déoxyhémoglobine induisent des changements de concentration en déoxyhémoglobine et donc des variations du BOLD.

Le modèle du couplage neurovasculaire et de sa mesure se base donc sur l'évolution temporelle de quatre variables d'état pour chaque voxel/par région, dites variables hémodynamiques : le signal de vasodilatation des artériolles sera noté  $Z(t)$ , le flux  $F(t)$ , le volume  $V(t)$  et la quantité de déoxyhémoglobine  $Q(t)$ . Comme nous le verrons plus loin, toutes les grandeurs seront adimensionnalisées par rapport à leur valeur hors période stimulation (ou au repos). Nous adopterons pour la suite la convention de [Stephan et al., 2007], les lettres majuscules sont les grandeurs dimensionnées et les minuscules représentent les grandeurs normalisées par rapport au repos.

### 5.3.2.2.1.3 Structure du modèle

C'est un modèle composé de 3 blocs qui s'enchaînent. Pour chaque région/voxel de l'image, nous avons :

1. un modèle dynamique linéaire du lien entre l'activité neurale qui sert de forçage, le signal de vasodilatation des artériolles et le flux sanguin en entrée du compartiment veineux (proposé par [Friston et al., 2000])

Ce bloc constitue une extension des deux suivants<sup>5</sup>, qui ont été proposés avant par [Buxton et al., 1998] (cf [Buxton, 2012] pour un historique du modèle).

2. un modèle dynamique non linéaire du lien entre le flux sanguin en entrée du compartiment veineux, qui sert de forçage, et les variations du volume et de la quantité en

5. on parle aussi de "extended Balloon-Windkessel model"

déoxyhémoglobine. Ce bloc est aussi appelé le "Balloon-Windkessel model" ou modèle du ballon-chambre à air ([Buxton et al., 1998], [Mandeville et al., 1999])

Les blocs 1 et 2 constituent le modèle de la réponse hémodynamique (ou couplage neurovasculaire). On parle aussi de modèle de ballon-chambre à air étendu ("extended Balloon-Windkessel Model"). Ils se traduisent mathématiquement sous la forme de quatre équations différentielles non linéaires qui constituent les équations d'état hémodynamiques ([Stephan et al., 2007]).

3. un modèle non linéaire de la mesure BOLD (plus précisément des variations de signal BOLD par rapport au repos), comme une fonction du volume et de la quantité en déoxyhémoglobine (proposé par [Buxton et al., 1998]).

L'expression des paramètres de ce modèle a évolué ([Obata et al., 2004], [Stephan et al., 2007]) et leur valeur dépend de la puissance de l'aimant de l'IRM utilisée.

#### 5.3.2.2.1.4 Propriétés

Le modèle complet présente de plus les 2 propriétés suivantes.

- Modèle hiérarchique : Les variables d'état d'un bloc  $n$  peuvent agir sur le suivant  $n + 1$ . Cependant, les variables d'état du bloc  $n + 1$  ne rétroagissent pas sur le bloc  $n$ . En ce sens, le modèle est hiérarchique et les variables d'état du bloc  $n$  peuvent être considérées comme des forçages pour le bloc  $n + 1$ . Leur l'évolution au cours du temps est régie de manière indépendante par le bloc  $n$ .
- Modèle découplé : Les réponses hémodynamiques des voxels/régions ne sont pas couplées entre elles. Autrement dit, la réponse hémodynamique dans un voxel/une région ne modifie pas celle d'un autre voxel/d'une autre région. Il en est de même pour les mesures BOLD de chaque voxel.

#### 5.3.2.2.1.5 1er bloc : Comment les variations de l'activité neurale provoquent une variation de flux en entrée du compartiment veineux ?

Ce modèle, proposé par [Friston et al., 2000], lie l'activité neurale<sup>6</sup> ( $n_i$ ) au flux en entrée du compartiment veineux ( $f_i$ ) par l'intermédiaire d'un signal de vasodilatation des artériolles ( $z_i$ ). La variation de l'activité neurale produit une dilatation ou compression (i.e. vasodilatation) des artériolles en amont du compartiment veineux, pour réguler le flux en entrée de ce dernier. Des mesures expérimentales ont mis en évidence une relation linéaire entre activité neurale et flux ([Stephan et al., 2004], [Stephan et al., 2007]).

Le modèle se traduit mathématiquement sous la forme de deux équations différentielles linéaires, couplées entre elles.  $z_i$  est un réel,  $f_i$  un réel positif.

$$\begin{cases} \dot{z}_i(t) = \epsilon n_i(t) - \kappa z_i(t) - \gamma(f_i(t) - 1) \\ \dot{f}_i(t) = z_i(t) \end{cases} \quad (5.11)$$

---

6. [Stephan et al., 2004] soulignent que la nature de cette activité neurale n'est pas précisée dans le modèle, notamment parce que la manière dont l'activation des neurones déclenchent des changements hémodynamiques est une question ouverte.

La première équation, de la forme de celle d'un oscillateur amorti, décrit l'évolution temporelle du signal de vasodilatation des artériolles  $z_i(t)$ . La décroissance exponentielle spontanée du signal de vasodilatation (de constante de temps  $\kappa$ ) est compensée par l'activité neurale.  $\epsilon$  est l'efficacité avec laquelle l'activité neurale cause une augmentation du signal de vasodilatation. La rétroaction exercée par le flux permet de freiner la vasodilatation.  $\gamma$  est la constante de temps de la régulation par feedback du flux.  $\kappa = 0.65s^{-1}$ ,  $\epsilon = 1$  et  $\gamma = 0.41s^{-1}$  d'après [Wang et al., 2019].

La deuxième équation donne l'évolution du flux en entrée du compartiment veineux en fonction du signal de vasodilatation.

Remarque (cf [Stephan et al., 2007]) :

$f_i$  représente la valeur du flux normalisé par rapport à sa valeur de repos  $F_0$ . Hors période de stimulation,  $f_i = 1$  et le terme  $f - 1$  s'annule : il n'y a pas de régulation par feedback au repos (i.e. la rétroaction du flux sur le signal de vasodilatation est nulle).

### 5.3.2.2.1.6 Balloon-Windkessel model

#### 1) Description

Ce modèle donne les variations de volume et de quantité de déoxyhémoglobine en fonction du flux sanguin en entrée du compartiment veineux.

Le compartiment se comporte comme un ballon extensible et résistant aux déformations, en accord avec des mesures expérimentales sur le rat . Il est traversé par du sang incompressible. Les variations de volumes suivent passivement les variations de flux en entrée : quand le flux en entrée augmente, le ballon se gonfle avec un certain décalage dans le temps (réponse visco-élastique)<sup>7</sup>, et inversement ([Buxton, 2012]).

L'augmentation de volume du compartiment veineux dilue la concentration de déoxyhémoglobine. Par ailleurs, le ballon se dégonfle en expulsant d'autant plus de sang qu'il a été déformé : le flux en sortie augmente de manière non linéaire avec le volume du ballon. En plus de ces effets mécaniques, la quantité de déoxyhémoglobine contenue dans le ballon varie pour des raisons métaboliques : le dioxygène est extrait du ballon par les neurones à proximité, faisant ainsi augmenter la quantité de déoxyhémoglobine.

#### 2) Traduction mathématique

Les deux variables d'état sont le volume du compartiment veineux ( $V$ ) et la quantité de déoxyhémoglobine ( $Q$ ). Ce sont deux réels positifs. Le flux en entrée du compartiment veineux, noté  $F$  agit comme un forçage. Les étapes de construction du modèle se trouvent dans [Buxton et al., 1998].

Il se traduit mathématiquement sous la forme de deux équations différentielles non linéaires, couplées entre elles. Chacune de ces équations repose sur une hypothèse de conservation de la masse ([Obata et al., 2004]) : conservation de la quantité totale de sang (fluide incompressible) et de la quantité de déoxyhémoglobine. Dans la suite, les différences de grandeurs

---

7. L'undershoot post-stimulus du BOLD trouverait une explication mécanique dans ce retard des variations volumiques par rapport aux variations de flux ([Obata et al., 2004]), et ne serait pas dû à un effet métabolique

(flux et quantité de déoxyhémoglobine) entre l'entrée et la sortie du compartiment veineux permettront alors de construire la variation de la grandeur pendant  $dt$ .

## 2) a) Conservation du volume sanguin

A tout instant, la variation de volume du compartiment veineux est égale à la différence entre les flux entrant  $F_{in}$  et sortant  $F_{out}$  :

$$\dot{V}_i = F_{in}(t) - F_{out}(t) \quad (5.12)$$

Soient  $F_0$ ,  $V_0$  et  $Q_0$  les valeurs au repos (hors stimulation). Elles permettent de définir de nouvelles variables d'état sans dimensions :  $f_{in}(t) = \frac{F_{in}(t)}{F_0}$ ,  $f_{out}(t) = \frac{F_{out}(t)}{F_0}$ ,  $v(t) = \frac{V(t)}{V_0}$  et  $q(t) = \frac{Q(t)}{Q_0}$  (que nous utiliserons un peu plus tard). Par exemple, si  $v(t) = 1$  alors  $V(t) = V_0$ .  $v(t)$  représente le taux de variation de volume du ballon par rapport au volume au repos. Par construction, avant toute activation  $q = v = f_{in} = f_{out} = 1$ . Nous pouvons alors écrire :

$$\dot{v}_i = \frac{1}{\tau_0} (f_{in}(t) - f_{out}(t)) \quad (5.13)$$

avec  $\tau_0 = \frac{V_0}{F_0}$  qui représente le temps de transit moyen nécessaire pour traverser le ballon au repos ([Buxton et al., 1998], [Friston et al., 2000], [Stephan et al., 2007]).

## 2) b) Conservation de la quantité de déoxyhémoglobine

A tout instant, la variation de la quantité en déoxyhémoglobine présente dans le compartiment veineux est égale à la différence entre la quantité entrante  $Q_{in}$  (liée à la consommation de dioxygène présente sur l'oxyhémoglobine) et sortant  $Q_{out}$  (liée au flux de sang sortant du ballon).

$$\dot{Q}_i = Q_{in}(t) - Q_{out}(t) \quad (5.14)$$

En posant,  $C_a$  la concentration de dioxygène dans les artériolles et  $E(t)$  la fraction de dioxygène extraite à  $t$  en entrée du compartiment (i.e. dans le sang entrant),  $Q_{in}$  peut s'écrire :

$$Q_{in}(t) = F_{in}(t)C_aE(t) \quad (5.15)$$

D'autre part,  $Q_{out}$  est égal au produit du flux sortant par la concentration volumique en déoxyhémoglobine présente dans le compartiment veineux :

$$Q_{out} = F_{out} \frac{Q_i}{V_i} \quad (5.16)$$

Nous obtenons :

$$\dot{Q}_i = F_{in}(t)C_aE(t) - F_{out} \frac{Q_i}{V_i} \quad (5.17)$$

En notant  $E_0$  la fraction de dioxygène extraite au repos dans le sang entrant, nous obtenons l'équation adimensionnalisée suivante. Elle donne l'évolution dans le temps du taux de variation de la quantité de déoxyhémoglobine  $q$  par rapport à la valeur au repos.

$$\dot{q}_i = \frac{1}{\tau_0} \left( f_{in}(t) \frac{E(t)}{E_0} - f_{out} \frac{q(t)}{v(t)} \right) \quad (5.18)$$

### 3) Deux hypothèses supplémentaires

#### 3) a) Modèle du flux sortant qui dépend uniquement du volume du ballon :

$$f_{out}(t) \equiv f_{out}(v)$$

Soit  $\alpha$  la résistance du ballon veineux (aussi appelé "Grubb's exponent" d'après [Grubb et al., 1974]). La relation non linéaire adimensionnalisée entre le flux éjecté du ballon et son volume s'écrit ([Buxton et al., 1998]) :

$$f_{out}(v) = v^{\frac{1}{\alpha}} \quad (5.19)$$

Plus le volume du compartiment veineux augmente par rapport au repos, et plus le flux expulsé par rapport au repos est grand.

#### 3) b) Modèle métabolique d'extraction du dioxygène dans le sang entrant, $E(t) \equiv E(f_{in})$

[Buxton et al., 1998] proposent un modèle métabolique d'extraction de l'oxygène couplé au modèle hémodynamique, sous l'hypothèse que l'oxygène extrait est immédiatement métabolisé par les neurones. Les tissus ne font pas de réserve de dioxygène, la concentration tissulaire de dioxygène est toujours nulle. La quantité d'oxygène extraite dépend alors uniquement de la quantité disponible en entrée du compartiment et donc du flux ([Stephan et al., 2004], [Stephan et al., 2007]).

Sous ces hypothèses,  $E(t)$  devient une fonction du flux entrant :  $E(f_{in})$ . Elle s'écrit :

$$E(f_{in}) = 1 - (1 - E_0)^{\frac{1}{f_{in}}} \quad (5.20)$$

Pour augmenter la quantité d'oxygène extraite, il faut nécessairement augmenter le flux sanguin en entrée ([Buxton, 2012]).

### 4) Bilan pour le bloc :

Pour chaque région  $i$ , les équations qui régissent l'évolution temporelle du volume  $v_i$  et de la quantité de déoxyhémoglobine  $q_i$ , adimensionnalisés par rapport au repos, dépendent de trois paramètres uniquement :  $\tau_0$  (mean transit time),  $E_0$  (resting oxygen extraction) et  $\alpha$  (stiffness exponent). Elles s'écrivent :

$$\begin{cases} \dot{v}_i = \frac{1}{\tau_0} (f_i - v_i^{\frac{1}{\alpha}}) \\ \dot{q}_i = \frac{f_i}{\tau_0 E_0} (1 - (1 - E_0)^{\frac{1}{f_i}}) - \frac{1}{\tau_0} q_i v_i^{\frac{1}{\alpha}-1} \end{cases} \quad (5.21)$$

avec  $\tau_0 = 0.98s$ ,  $E_0 = 0.34$ ,  $\alpha = 0.32$  ([Wang et al., 2019]). Dans certaines publications,  $E_0$  et  $\tau_0$  sont aussi respectivement notés  $\rho$  et  $\tau$  (e.g. [Friston et al., 2003], [Wang et al., 2019]). Des simulations sont présentées à la section 5.3.2.2.3.

### 5.3.2.2.2 Modèle BOLD non linéaire

Pour une région, on note  $S_0$  la mesure du signal BOLD au repos (hors stimulation) et  $S$  celle en période d'activation.  $\Delta S = S - S_0$  est la variation du signal BOLD entre l'activation et le repos.

Le modèle du BOLD proposé par [Buxton et al., 1998], est un modèle qui donne au cours du temps la différence d'activité par rapport au repos, normalisée par rapport à la valeur du signal au repos i.e.  $\frac{\Delta S}{S_0}$ .

$\frac{\Delta S}{S_0}$  y est proposé comme une fonction non linéaire du volume  $v(t)$  et de la quantité de déoxyhémoglobine  $q(t)$  au cours du temps, normalisés par rapport à leurs valeurs au repos (respectement  $V_0$  et  $Q_0$ ) :

$$\frac{\Delta S}{S_0} \equiv y(t) = \lambda(v, q) \quad (5.22)$$

#### 5.3.2.2.2.1 En fonction des signaux intra et extra-vasculaires

##### 1) Somme pondérée des signaux intra et extra-vasculaires

Un voxel est l'échantillon volumique le plus petit d'une image obtenue avec une IRM. Il est constitué d'une partie contenant du sang et d'une autre partie non sanguine. Nous pouvons écrire  $S_0$  (au repos) et  $S$  (en stimulation) comme une somme pondérée de signaux intra et extra-vasculaires ([Buxton et al., 1998], [Obata et al., 2004] appendix A).

— Au repos (hors période de stimulation) :

Nous notons  $V_0$  la fraction de volume du voxel occupée par du sang veineux au repos. Comme le sang est un fluide incompressible,  $1 - V_0$  est la fraction de volume du voxel extra-vasculaire (i.e. non occupée par du sang veineux).

Le signal magnétique mesuré au repos dans un voxel, noté  $S_0$ , s'écrit comme la somme des signaux intra-vasculaire  $S_I$  et extra-vasculaire  $S_E$ , pondérée respectivement par leur fraction volumique respective.

$$S_0 = (1 - V_0)S_{E0} + V_0S_{I0} \quad (5.23)$$

Dans chacun des milieux intra et extra-vasculaire, la courbe de relaxation exponentielle transverse de l'aimantation est mesurée à un temps précis,  $t = TE$ , appelé temps d'écho. Les signaux intra et extra-vasculaires correspondent à ce point de mesure précis à  $t = TE$  (cf chapitre 4).

Nous notons  $T_{2E}^*$  le temps caractéristique de la relaxation transverse dans le milieu extra-vasculaire, et  $T_{2I}^*$  dans le milieu intra-vasculaire (au repos). Leurs inverses respectifs sont notés  $R_{2E}^*$  et  $R_{2I}^*$ .  $\tilde{S}_{E0}$  et  $\tilde{S}_{I0}$  sont les densités de protons respectives dans

les milieux intra et extra-vasculaires au repos. Nous avons alors :

$$S_{E0} = \tilde{S}_{E0} e^{-\frac{TE}{T_{2E}^*}} = \tilde{S}_{E0} e^{-TE \times R_{2E}^*} \quad (5.24)$$

$$S_{I0} = \tilde{S}_{I0} e^{-\frac{TE}{T_{2I}^*}} = \tilde{S}_{I0} e^{-TE \times R_{2I}^*} \quad (5.25)$$

Nous pouvons définir  $\epsilon$  comme le rapport des signaux intra-vasculaire et extra-vasculaire au repos ([Obata et al., 2004], repris par [Stephan et al., 2007]) :

$$\epsilon = \frac{S_{I0}}{S_{E0}} \quad (5.26)$$

— En période d'activation :

Nous notons  $V$  la fraction de volume du voxel occupée par du sang veineux en période de stimulation. Le signal magnétique mesuré dans un voxel, noté  $S$ , s'écrit comme la somme des signaux intra-vasculaire  $S_I$  et extra-vasculaire  $S_E$ , pondérée respectivement par leur fraction volumique respective.

$$S = (1 - V)S_E + VS_I \quad (5.27)$$

Les temps caractéristiques varient respectivement de  $\Delta T_{2E}^*$  et  $\Delta T_{2I}^*$ .

$$S_E = \tilde{S}_{E0} e^{-\frac{TE}{T_{2E}^* + \Delta T_{2E}^*}} \quad (5.28)$$

$$S_I = \tilde{S}_{I0} e^{-\frac{TE}{T_{2I}^* + \Delta T_{2I}^*}} \quad (5.29)$$

En posant :

$$\frac{1}{T_{2E}^* + \Delta T_{2E}^*} = R_{2E}^* + \Delta R_{2E}^* \quad (5.30)$$

$$\frac{1}{T_{2I}^* + \Delta T_{2I}^*} = R_{2I}^* + \Delta R_{2I}^* \quad (5.31)$$

Nous obtenons :

$$S_E = \tilde{S}_{E0} e^{-TE \times (R_{2E}^* + \Delta R_{2E}^*)} \quad (5.32)$$

$$S_I = \tilde{S}_{I0} e^{-TE \times (R_{2I}^* + \Delta R_{2I}^*)} \quad (5.33)$$

Finalement le signal mesuré dans le voxel en période de stimulation ([Stephan et al., 2007]) peut être écrit directement en fonction des signaux intra et extra-vasculaires au repos  $S_{E0}$  et  $S_{I0}$  :

$$S = (1 - V)S_{E0} e^{-TE \times \Delta R_{2E}^*} + VS_{I0} e^{-TE \times \Delta R_{2I}^*} \quad (5.34)$$

Notons  $v = \frac{V}{V_0}$  le rapport des fractions volumiques du voxel occupées par le sang entre l'activation et le repos. Ce qui donne :

$$S = (1 - vV_0)S_{E0} e^{-TE \times \Delta R_{2E}^*} + vV_0 S_{I0} e^{-TE \times \Delta R_{2I}^*} \quad (5.35)$$

Il s'agit de l'expression exacte du signal BOLD  $S$  pendant une activation. Lorsque  $TE$ ,  $V_0$  sont connus, ce signal dépend de trois paramètres  $\Delta R_{2E}^*$ ,  $\Delta R_{2I}^*$  et  $v$  ([Stephan et al., 2007]) qui représentent respectivement les variations par rapport au repos de  $T_2^*$  intra et extra-vasculaires et de fraction de volume du voxel occupé par du sang veineux.

## 2) Expression de $\frac{\Delta S}{S_0}$

L'expression exacte de  $\Delta S$  est :

$$\Delta S = S - S_0 \quad (5.36)$$

$$= (1 - vV_0)S_{E0}e^{-TE \times \Delta R_{2E}^*} + vV_0S_{I0}e^{-TE \times \Delta R_{2I}^*} - (1 - V_0)S_{E0} - V_0S_{I0} \quad (5.37)$$

[Stephan et al., 2007] (appendix B), donne une approximation de  $\Delta S$  pour des petites variations autour du repos (i.e.  $\Delta R_{2E}^*$  et  $\Delta R_{2I}^*$  proche de 0 et  $v$  proche de 1.)

$$\frac{\Delta S}{S_0} \approx -\Delta R_{2E}^*TE - V_0\epsilon\Delta R_{2I}^*TE + V_0(1 - v)(1 - \epsilon) \quad (5.38)$$

Les variations du signal BOLD dans le voxel sont dues à trois composantes ([Buxton et al., 1998]) : des variations du signal extra-vasculaire (1er terme), intra-vasculaire (2e terme), des changements de volumes intra et extra-vasculaires (3e terme) qui ont pour effet de déplacer les moments magnétiques  $\vec{\mu}_i$  qui forment l'aimantation totale  $\vec{M}$  (cf chapitre 4).

### 5.3.2.2.2 En fonction du volume et de la quantité de déoxyhémoglobine

Les variations extra et intra-vasculaires  $\Delta R_{2E}^*$  et  $\Delta R_{2I}^*$  peuvent ensuite s'exprimer en fonction du volume  $v$  et de la quantité de déoxyhémoglobine  $q$ .

$\Delta R_{2E}^*$  est exprimé en fonction de la quantité de déoxyhémoglobine par rapport à la valeur de repos,  $q$ . Cette expression est proposée par [Obata et al., 2004] à partir des simulations numériques de [Ogawa et al., 1993].

$$\Delta R_{2E}^* = 4.3\theta_0V_0E_0(q - 1) \quad (5.39)$$

[Obata et al., 2004] utilisent une approximation linéaire sur des données obtenues in vivo chez des animaux (cochons) par [Li et al., 1998], pour proposer une expression de  $\Delta R_{2I}^*$  en fonction de la concentration en déoxyhémoglobine contenue dans le compartiment intravasculaire,  $\frac{q}{v}$ .

$$\Delta R_{2I}^* = r_0E_0\left(\frac{q}{v} - 1\right) \quad (5.40)$$

Nous avons donc :

$$\frac{\Delta S}{S_0} \approx -4.3\theta_0V_0E_0(q - 1)TE - V_0\epsilon r_0E_0\left(\frac{q}{v} - 1\right)TE + V_0(1 - v)(1 - \epsilon) \quad (5.41)$$

Avec les coefficients sans dimension :

$$\begin{cases} k1 = 4.3\theta_0E_0TE \\ k2 = \epsilon r_0E_0TE \\ k3 = 1 - \epsilon \end{cases} \quad (5.42)$$



$\frac{\Delta S}{S_0}$  devient :

$$\frac{\Delta S}{S_0} \approx V_0 \left( k_1(1 - q) + k_2(1 - \frac{q}{v}) + k_3(1 - v) \right) \equiv y \quad (5.43)$$

Le second terme introduit une non linearité.

Des variations de volume et de quantité en déoxyhémoglobine induisent des modification d'aimantation dans le milieu à l'origine du signal BOLD ([Stephan et al., 2007]). Supposons uniquement une variation du niveau d'oxygénation (i.e. de la quantité de déoxyhémoglobine  $q$ ), sans changement de volume (i.e.  $v = 1$ ). La contribution du 3e terme s'annule. La variation de signal  $y$  est proportionnelle à  $k_1 + k_2$  et n'est due qu'aux variations de signal dans les régions intra et extra-vasculaires. [Obata et al., 2004] propose d'autres expériences de pensées pour manipuler et comprendre les différentes contributions aux variations du signal.

Les valeurs numériques des paramètres dépendent de la puissance de l'aimant ([Heinzle et al., 2016]). Pour des IRM avec un champ magnétique à  $B_0 = 3T$ , ([Wang et al., 2019]) :  $V_0 = 0.02$ ,  $\theta_0 = 28.265B_0$ ,  $E_0 \equiv \rho = 0.34$ <sup>8</sup>,  $\epsilon = 0.47$ ,  $r_0 = 110Hz$ .  $TE$  doit être adapté à la séquence IRMf utilisée. Dans notre cas,  $TE = 0.030s$  (cf [Mary et al., 2020]). Pour un historique de la construction de ce modèle (forme de la fonction et expression des paramètres) , nous pourrons nous reporter à [Stephan et al., 2007]<sup>9</sup>

### 5.3.2.2.3 Modèle complet et simulations numériques

En adoptant les notations de [Wang et al., 2019] pour  $E_0 \equiv \rho$  et  $\tau_0 \equiv \tau$ , pour chaque région  $i$ , nous avons quatre équations d'état hémodynamiques :

$$\begin{cases} \dot{z}_i = \epsilon n_i - \kappa z_i - \gamma(f_i - 1) \\ \dot{f}_i = z_i \\ \dot{v}_i = \frac{1}{\tau} \left( f_i - v_i^{\frac{1}{\alpha}} \right) \\ \dot{q}_i = \frac{f_i}{\tau \rho} \left( 1 - (1 - \rho)^{\frac{1}{f_i}} \right) - \frac{1}{\tau} q_i v_i^{\frac{1}{\alpha} - 1} \end{cases} \quad (5.44)$$

8. la même valeur que pour le Balloon-Windkessel Model, section 5.3.2.2.1.6

9. Il est important de noter que lors de la première proposition de modèle par [Buxton et al., 1998], les coefficients  $k_i$  sont estimés pour des IRM à  $B_0 = 1.5T$  avec une mesure à  $TE = 40ms$ . Leur expression est un peu différente et est aujourd'hui encore utilisée dans certains articles.  $k_1$  et  $k_3$  y sont des fonctions linéaires de  $E_0$  et  $k_2$  est une constante :  $k_1 \approx 7E_0$ ,  $k_2 \approx 2$ ,  $k_3 \approx 2E_0 - 0.2$ .

et une fonction qui donne le signal BOLD <sup>10</sup> :

$$y_i = V_0 \left( k_1(1 - q_i) + k_2(1 - q_i v_i^{-1}) + k_3(1 - v_i) \right) \quad (5.45)$$

$$k_1 = 4.3\theta_0\rho TE \quad (5.46)$$

$$k_2 = \epsilon r_0\rho TE \quad (5.47)$$

$$k_3 = 1 - \epsilon \quad (5.48)$$

La figure 5.4 illustre les différentes couches du modèle complet et montre les dépendances entre variables d'état pour une région  $i$ .

Les quatre variables d'état hémodynamiques (signal de vasodilatation, flux, volume et quantité de déoxyhémoglobine) sont généralement non accessibles à la mesure et on parle d'états cachés. Le signal BOLD constitue l'unique mesure par région.

Notons que beaucoup d'autres propositions ont été faites par la communauté pour modéliser le couplage neurovasculaire (au niveau du modèle du ballon, du flux sortant  $f_{in}(v)$  ou du modèle métabolique d'extraction  $E(f_{in})$  par exemple) et la mesure de la réponse BOLD associée (cf [Stephan et al., 2004], "Growing sophistication of the models" dans l'article de [Buxton, 2012]).

---

10. Modèle de variations par rapport au repos (i.e. période sans stimulation)

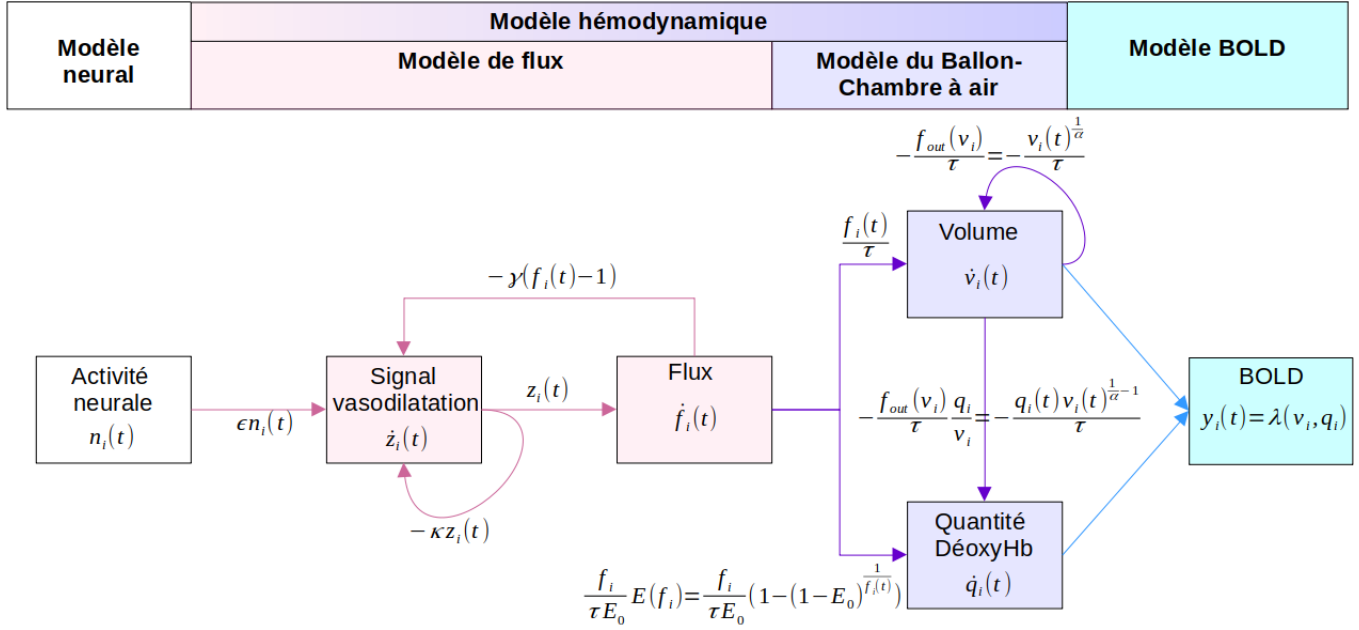


**Figure 5.4** – Synthèse du modèle du couplage neurovasculaire et de sa mesure BOLD : un modèle hiérarchique et découplé entre régions.

L'activité neurale dans une région cause des variations hémodynamiques, d'abord de vasodilation des artériolles et de flux en entrée du compartiment veineux (modèle de flux), puis de volume et quantité de déoxyhémoglobine dans le compartiment veineux (modèle du Ballon-Chambre à air). Ces deux couches forment le "Balloon-Windkessel extended model" ou modèle du Ballon-Chambre à air étendu.

Les variations de volume et quantité de déoxyhémoglobine provoquent alors une variation du signal BOLD dans la région.

La réponse hémodynamique et sa mesure BOLD dans une région ne modifie pas celle d'une autre région. En effet, les couplages entre variables hémodynamiques sont uniquement présents au sein d'une même région et absence entre les régions. On peut parler de modèle découplé entre les régions.

Synthèse graphique inspirée de [Friston et al., 2000].

Les graphes suivants montrent les résultats de quelques simulations numériques, d'une durée totale de 100 secondes. Nous prenons comme conditions initiales  $z(t=0) = 0$ ,  $f(t=0) = 1$ ,  $v(t=0) = 1$ ,  $q(t=0) = 1$  ([Deco et al., 2018], [Wang et al., 2019]).

Pour une région, nous imposons comme input neural  $n(t)$  une fonction créneau qui vaut 1 en période de stimulation et 0 au repos (cf 5.3.1.1). Nous stimulons la région deux fois pendant ces 100 secondes. La durée d'une stimulation, i.e. d'une porte, est égale soit à 20 secondes (lignes A et B), soit à 3 secondes (lignes C et D). La durée entre les deux stimulations vaut soit 20 secondes (lignes A et C), soit 3 secondes (lignes B et D).

Sur les colonnes de gauche (1), nous avons représenté la fonction neurale  $n(t)$  en poin-

tillés, qui sert de forçage au modèle de flux, ainsi que les deux variables d'état de cette couche du modèle complet (signal de vasodilatation  $z(t)$  et flux  $f(t)$ ).

Sur les colonnes du milieu (2), nous avons représenté le flux  $f(t)$  en pointillés, qui sert de forçage au modèle du Ballon-Chambre à air, ainsi que les deux variables d'état de cette couche du modèle complet (volume et quantité de déoxyhémoglobine).

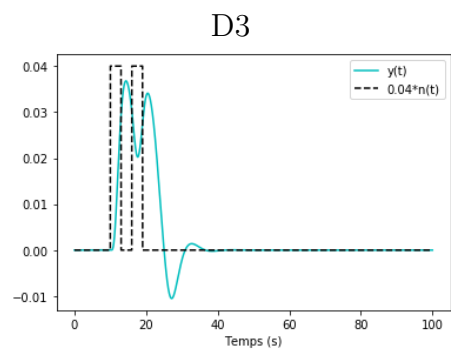
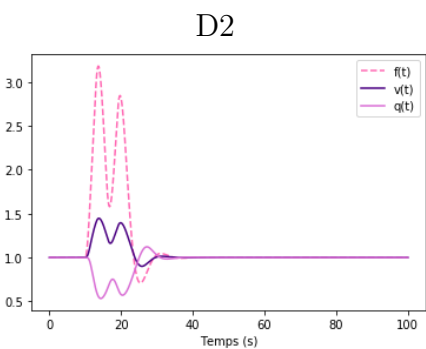
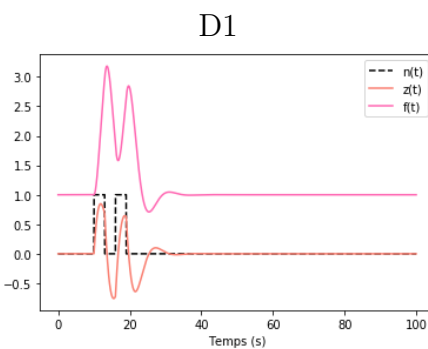
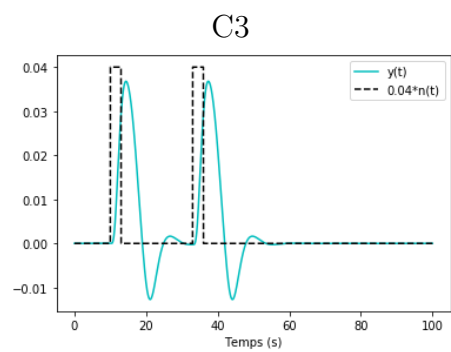
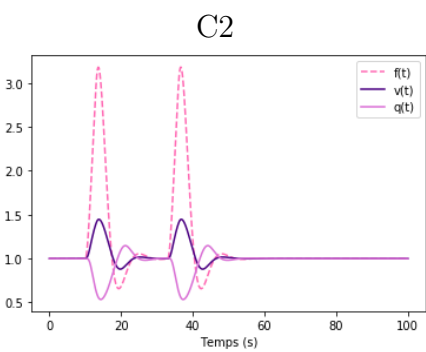
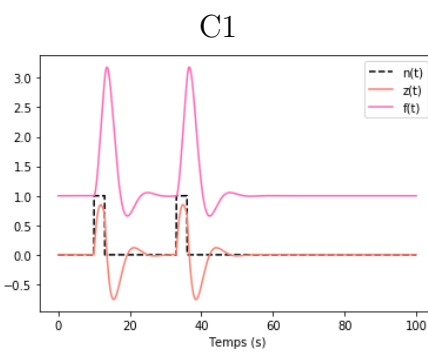
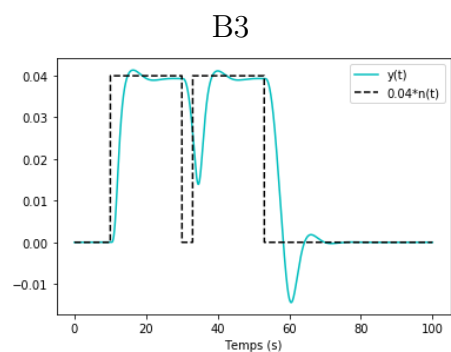
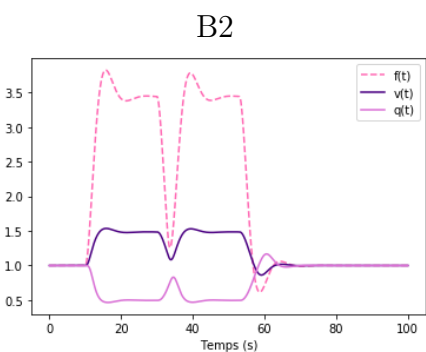
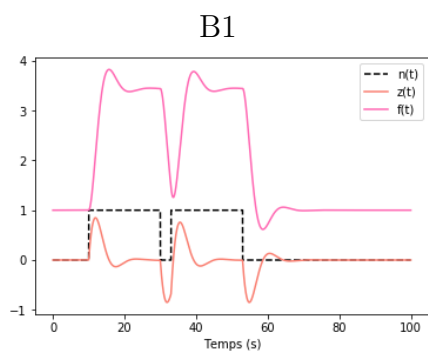
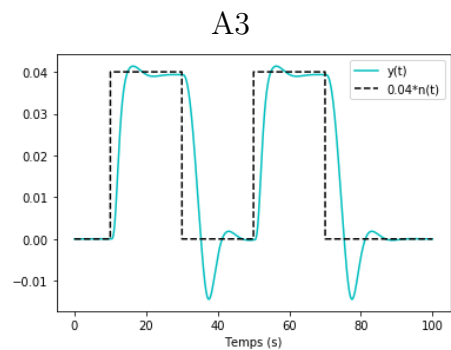
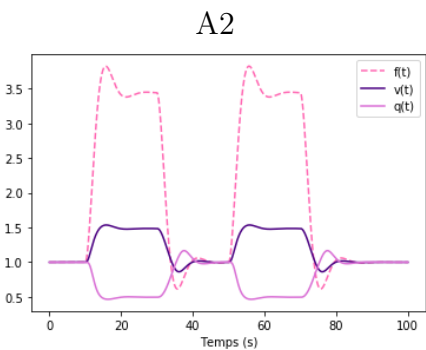
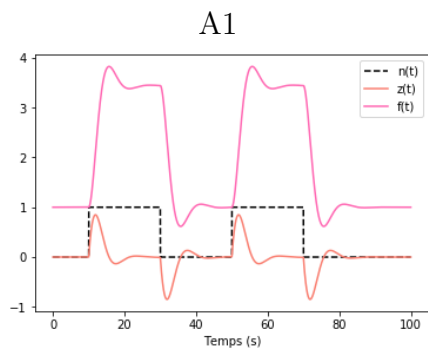
Sur les colonnes de droite (3), nous avons représenté le signal BOLD résultant et nous avons reporté la fonction neurale  $n(t)$ , mise à l'échelle.

Nous décrivons en priorité les variations du signal BOLD au cours du temps. Les observations rapportées sont également valables pour les variables d'état hémodynamiques, dont le niveau basal est donné pour les conditions initiales qui ont été données ci-dessus.

Sans stimulation, le signal BOLD est nul (niveau basal). Nous observons un décalage temporel entre le début de la stimulation et l'augmentation du signal BOLD. Sur les figures A3 et B3, nous observons que la durée de la stimulation est suffisamment grande pour que le signal BOLD se stabilise autour d'une valeur proche de 0.04 (un point fixe), en passant juste avant par un overshoot. Au contraire, sur les figures C3 et D3, la durée de la stimulation est trop courte pour que le point fixe puisse être atteint. L'overshoot n'a pas eu non plus le temps de se produire.

Le signal BOLD redescend après l'arrêt de la stimulation, avec un certain décalage dans le temps. Si la durée au repos est suffisamment grande (figures A3 et C3), le signal BOLD passe par un creux post-stimulation (undershoot) avant de tendre vers le niveau basal nul. Si durée au repos est trop courte (figures B3 et D3), le signal BOLD n'a pas le temps de redescendre à son niveau basal avant de remonter à cause de la stimulation suivante.

La fin de la simulation est marquée par un undershoot et le retour au niveau basal.



*Dans la suite, nous présentons deux méthodes d'analyses de données qui permettent d'étudier la connectivité fonctionnelle (PPI) et effective (DCM) entre régions cérébrales.*

*Nous présentons les étapes de l'analyse des données d'un participant (i.e. analyses individuelles), en explicitant à chaque fois :*

- *le modèle dynamique considéré*
- *le modèle statistique de la mesure*
- *les paramètres d'intérêt à estimer*
- *la méthode d'inversion de ces paramètres*

*Nous reprendrons ces idées dans la suite de la thèse. Le chapitre 10 permettra de construire une méthode d'analyse des données d'un participant pendant une tâche de contrôle inhibiteur, pour laquelle il sera nécessaire de préciser chacune de ces étapes.*

*La méthode d'analyse des données d'un participant est à distinguer des analyses à l'échelle d'un groupe de sujets. Pour chaque participant, les séries temporelles BOLD sont analysées séparément. Une fois les paramètres d'intérêt estimés pour chaque participant, des modèles statistiques de groupe permettent ensuite d'étudier les résultats de connectivité fonctionnelle ou effective de plusieurs sujets. Nous ne les présenterons pas dans ce texte.*

## 5.4 Analyse de connectivité fonctionnelle basée sur un modèle dynamique : PPI (Psychophysiological Interactions)

### 5.4.1 Objectif de l'analyse

Nous disposons pour chaque région des décours temporels BOLD mesurés pendant une tâche (i.e. un point tous les  $TR$ , dans notre laboratoire  $TR = 2.05s$ ).

La tâche peut être divisée en  $\mathcal{N}_E$  conditions expérimentales. Par exemple pendant une tâche Think/No Think de contrôle inhibiteur mnésique, nous pouvons citer la condition Think, No Think non intrusive et No Think intrusive ( $\mathcal{N}_E = 3$ ). Nous connaissons les moments de présentations des stimuli associés à chaque condition expérimentale. Ceci nous permet de construire une fonction de stimulation pour chaque condition expérimentale, notée  $U^{E_j}(t)$  ( $j \in [1; \mathcal{N}_E]$ ). Elle vaut 1 en période de stimulation et 0 sinon.

L'analyse gPPI (pour "generalized psychophysiological interaction", [McLaren et al., 2012]) nécessite de choisir un couple de régions : une région de contrôle (aussi appelée "seed" en anglais) et une région cible ("target").

La région de contrôle et la région cible présentent une variation d'activité BOLD due à la mise en oeuvre de mécanismes cérébraux spécifiques impliqués pendant une condition expérimentale. Nous souhaitons étudier les variations d'activité BOLD dans la cible liées aux variations de l'activité BOLD dans la région de contrôle spécifiquement mises en oeuvre pendant une condition expérimentale ([O'Reilly et al., 2012]).

Nous supposons que les variations du signal BOLD mesurées dans la cible, même après pré-traitements, sont un mélange de plusieurs sources principales :

1. les variations d'activité BOLD de la région de contrôle due à chaque condition expérimentale  $E_j : y_{RC}^{E_j}(t)$  ( $j \in [1; \mathcal{N}_E]$ )
2. les variations d'activité BOLD de la cible due à chaque condition expérimentale  $E_j : y_C^{E_j}(t)$  ( $j \in [1; \mathcal{N}_E]$ )
3. les variations d'activité BOLD de la région de contrôle qui ne sont pas forcément dues à la tâche. En effet, la région de contrôle et la cible peuvent présenter des variations en phase du fait de mécanismes physiologiques communs (comme la respiration) ou de connexions avec d'autres régions qui envoient un signal identique (cf figure 5.1 schéma 5).

Les décours temporels dus aux deux premières sources de variations sont modélisés à partir du modèle linéaire de convolution pour passer de l'échelle neurale à l'échelle BOLD (cf section 5.3.2.1). Nous en donnons le détail dans le paragraphe 5.4.2. La troisième source de variation est directement égale au décours temporel BOLD mesuré dans la région de contrôle, noté  $y_{RC}(t)$ .

Ces décours temporels serviront ensuite de régresseurs dans un modèle statistique linéaire général (GLM) de la mesure d'activité BOLD dans la cible pour un sujet, que nous expliciterons en 5.4.3. Nous préciserons les paramètres de ce modèle statistique à estimer, ainsi que la méthode d'inversion des paramètres.

## 5.4.2 Les régresseurs construits à partir du modèle de convolution linéaire

### 5.4.2.1 Régresseur PPI au niveau BOLD

Le régresseur qui modélise les variations de l'activité BOLD d'une région de contrôle spécifiquement dues à la condition expérimentale  $E_j$  est appelée régresseur PPI (associé à la condition expérimentale  $E_j$ ). Nous le notons  $y_{RC}^{E_j}(t)$  ( $j \in [1; \mathcal{N}_E]$ ). Il est obtenu en suivant ces étapes :

- Nous partons du signal BOLD de la région de contrôle :  $y_{RC}(t)$
- Déconvolution du signal BOLD de la région de contrôle par la Fonction de Réponse Hémodynamique (HRF) pour obtenir son signal neural "théorique" :  $N_{RC}(t)$  <sup>11</sup>
- Multiplication de ce signal neural avec la fonction de stimulation  $U^{E_j}(t)$  pour obtenir un modèle de signal neural dû à la condition expérimentale  $E_j$  :  $N_{RC}^{E_j}(t) = U^{E_j}(t) \times N_{RC}(t)$ . Ceci revient à ne garder le signal neural qu'en période de stimulation (i.e. quand  $U^{E_j}(t) = 1$ ) et à annuler toutes les variations du signal neural hors stimulation (i.e. quand  $U^{E_j}(t) = 0$ ).  
 $N_{RC}^{E_j}(t)$  est un régresseur PPI au niveau neural.
- Convolution de  $N_{RC}^{E_j}(t)$  par la HRF, notée  $h(t)$ , pour trouver le signal BOLD dû à la condition expérimentale  $E_j$ . Il sera noté  $y_{RC}^{E_j}(t) : y_{RC}^{E_j}(t) = N_{RC}^{E_j}(t) \otimes h(t)$ .

---

11. La même déconvolution pour chaque condition expérimentale  $E_j$ , cf [McLaren et al., 2012]

Il est alors possible de construire autant de régresseurs PPI BOLD que de conditions expérimentales :  $y_{RC}^{E_j}(t)$ .

Nous avons présenté la méthode de déconvolution utilisée par [McLaren et al., 2012]. Cette méthode permet d'extraire un signal théorique pendant la tâche au niveau neural. La convolution permet ensuite de déduire le signal BOLD théorique pendant la tâche. Il existe également une autre méthode non équivalente pour générer un régresseur PPI ([Di et al., 2021], [Gitelman et al., 2003]), sans passer par la déconvolution. Le signal BOLD théorique pendant la tâche est directement calculé au niveau BOLD ([O'Reilly et al., 2012]). Il s'agit d'abord de calculer un signal neural théorique pendant la tâche :  $N_{RC}^{E_j}(t) = U^{E_j}(t) \otimes h(t)$ , puis de le multiplier par le décours temporel BOLD mesuré dans la région de contrôle :  $y_{RC}^{E_j}(t) = N_{RC}^{E_j}(t)y_{RC}(t)$ .

#### 5.4.2.2 Réponse BOLD théorique de la cible due uniquement à sa stimulation pendant la tâche

$y_C^{E_j}$  modélise les variations de l'activité BOLD d'une région cible dues spécifiquement à la condition expérimentale  $E_j$ .

Le signal neural dû à la condition expérimentale  $E_j$  est modélisé directement par la fonction de stimulation :  $N_C^{E_j}(t) = U^{E_j}(t)$ . Le signal BOLD associé  $y_C^{E_j}(t)$  est obtenu en convoluant  $N_C^{E_j}(t)$  par la HRF  $h(t)$  :  $y_C^{E_j}(t) = N_C^{E_j}(t) \otimes h(t)$ .

#### 5.4.3 Modèle statistique linéaire général (ou GLM)

Un modèle statistique linéaire des mesures de variations d'activité BOLD dans la cible est formulé à partir des régresseurs précédents, pour un sujet.

Les mesures prédites par le modèle sont notées  $\hat{y}_C(t)$  et peuvent s'écrire :

$$\hat{y}_C(t) = \alpha + \sum_{j=1}^{\mathcal{N}_E} \beta_{RC}^{E_j} y_{RC}^{E_j}(t) + \sum_{j=1}^{\mathcal{N}_E} \beta_C^{E_j} y_C^{E_j}(t) + \beta_{RC} y_{RC}(t) \quad (5.49)$$

$\alpha$  désigne l'ordonnée à l'origine et les paramètres  $\beta$  sont les amplitudes des différents régresseurs.

Notons  $\epsilon(t)$  est la réalisation à  $t$  d'une variable aléatoire qui suit une distribution gaussienne. D'après le modèle linéaire général, les mesures  $y_C(t)$  sont égales à :

$$y_C(t) = \hat{y}_C(t) + \epsilon(t) \quad (5.50)$$

Les paramètres  $\alpha$ ,  $\beta$  et  $\epsilon$  sont les paramètres du modèle statistique qu'il s'agit d'estimer. Supposons  $N$  points de mesures sur l'intervalle de temps considéré. Nous notons  $t_i$  avec  $i \in [1; N]$  un instant de mesure. Une méthode de régression linéaire multiple par minimisation de la somme des différences des carrés entre valeurs mesurées et prédites par le modèle (i.e.  $\sum_{i=1}^N (y_C(t_i) - \hat{y}_C(t_i))^2$ ) est couramment utilisée à cet effet.

Nous nous intéressons particulièrement aux paramètres  $\beta_{RC}^{E_j}$  (aussi appelés  $\beta_{PPI}$ ). Pour chaque condition expérimentale  $E_j$ ,  $\beta_{RC}^{E_j}$  est égal à la variation d'activité dans la cible lorsque



la région de contrôle augmente son activité d'une unité (cf revue [Friston, 2011]). C'est une mesure de la connectivité fonctionnelle entre la région de contrôle et la cible pendant la condition expérimentale  $E_j$ , mais qui ne donne pas d'information sur un lien causal entre les deux variations ([O'Reilly et al., 2012]).

Pour pouvoir interpréter les résultats, il est important de comparer une amplitude dans une situation d'intérêt par rapport à une situation de référence. C'est donc la différence de  $\beta_{RC}^{E_j}$  entre deux conditions expérimentales  $E_j$  différentes qui est importante.

Nous avons décrit ici un GLM pour un couple de régions de (contrôle, cible). Il est plus largement possible d'obtenir pour chaque couple de régions (contrôle,cible) les amplitudes  $\beta_{RC}^{E_j}$  de la variation d'activité de la cible due à la variation d'activité dans la région de contrôle pour la condition expérimentale  $E_j$ . Il y a donc autant de GLM que de couples de régions.

#### PPI

- le modèle dynamique considéré : modèle de convolution linéaire (5.3.2.1)
- le modèle statistique de la mesure : modèle linéaire général (GLM dont les régresseurs sont construits à partir du modèle ci-dessus)
- les paramètres d'intérêt du modèle statistique : pentes des GLM, notamment  $\beta_{RC}^{E_j} \equiv \beta_{PPI}$
- la méthode d'inversion de ces paramètres : régression linéaire par moindres carrés

### 5.4.4 Exemple du contrôle inhibiteur des intrusions mnésiques

#### 5.4.4.1 Résultats d'imagerie

[Schmitz et al., 2017] procède à une analyse PPI avec des données de contrôle inhibiteur d'intrusions mnésiques (paradigme TNT). L'étude met en évidence l'existence d'une connectivité fonctionnelle négative entre le DLPFC droit et l'hippocampe droit pendant la condition No Think (par rapport à la condition Think). Ceci permet de conclure que ces deux régions sont impliquées pendant le contrôle inhibiteur d'une intrusion et que l'augmentation d'activité du DLPFC est associée à une baisse d'activité de l'hippocampe. Néanmoins, nous ne pouvons pas en déduire qu'un signal se propage du DLPFC jusqu'à l'hippocampe pour abaisser son activité.

#### 5.4.4.2 Liens avec la concentration GABA

[Schmitz et al., 2017] mesure par spectroscopie la concentration de GABA au repos, dans l'hippocampe et le DLPFC.

##### 5.4.4.2.1 Dans les régions de contrôle

La concentration de GABA dans le DLPFC n'est pas corrélée significativement à la connectivité fonctionnelle PPI entre le DLPFC et l'hippocampe. L'augmentation d'activité du

DLPFC associée à la baisse d'activité de l'hippocampe pendant le contrôle inhibiteur mnésique semble indépendante de la concentration de GABA dans le DLPFC.

#### 5.4.4.2.2 Dans les régions cibles

Les participants avec une forte concentration de GABA dans l'hippocampe présentent une connectivité fonctionnelle significativement négative entre le DLPFC et l'hippocampe, et plus grande en valeur absolue que chez les participants avec une faible concentration de GABA (cf figure 5.5). Chez les participants avec une plus forte concentration de GABA hippocampique, l'augmentation d'activité du DLPFC pendant le contrôle inhibiteur mnésique est donc associée à une baisse d'activité de l'hippocampe plus importante que chez les participants avec une faible concentration de GABA.

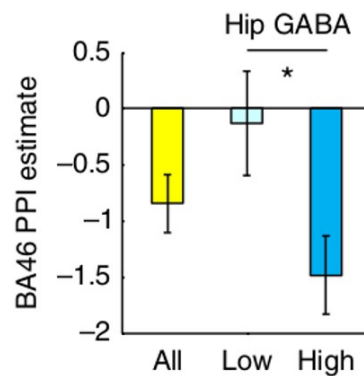


**Figure 5.5** – Couplages PPI négatifs entre le DLPFC et l'hippocampe mesurés pendant les suppressions mnésiques chez des participants ayant une faible ou une forte concentration de GABA dans l'hippocampe (t-test,  $*p < 0.05$ ). Les barres d'erreur indiquent l'erreur standard par rapport à la moyenne ([Schmitz et al., 2017]).

## 5.5 Analyse de connectivité effective basée sur un modèle dynamique : DCM

### 5.5.1 Présentation de la méthode d'analyse

Dynamic Causal Modeling (DCM) est une approche d'analyse des données en neurosciences qui permet d'investiguer la communication neurale entre régions cérébrales pendant une stimulation. Elle peut être employée avec différentes modalités d'imagerie : imagerie par résonance magnétique fonctionnelle, électroencéphalographie et magnétoencéphalographie ([Daunizeau et al., 2011]).

Elle résulte du travail de différentes communautés sur deux axes principaux ([Daunizeau et al., 2011]) :

- le développement de modèles dynamiques de l'activité neurale non mesurée (nous parlerons d'états cachés ou "hidden states") et de la mesure résultante (BOLD dans le cas de l'IRMf, aussi appelée observable)
- le développement de méthodes statistiques d'inversion pour estimer les paramètres des modèles à partir des données et comparer différents modèles entre eux

Dans cette partie, nous détaillerons la méthode d'analyse la plus couramment utilisée avec des données d'IRMf pour étudier la connectivité effective entre régions cérébrales (proposée par [Friston et al., 2003]). Elle permet d'étudier l'influence qu'un système neural exerce sur un autre pendant une tâche, qui se traduit par l'intensité d'un paramètre de couplage dans le modèle dynamique neural.

Nous prendrons soin d'explicitier les 4 points suivants qui permettent d'analyser les données d'un sujet : 1) le modèle dynamique considéré, 2) le modèle statistique de la mesure, 3) les paramètres d'intérêt à estimer et enfin 4) la méthode d'inversion de ces paramètres. Nous ne les détaillons pas tous, puisque les points 2) et 4) présentent des similitudes avec les éléments qui seront développés dans le chapitre 10 de la thèse, à la section 10.3.

#### 5.5.1.1 Modèle dynamique

Le modèle dynamique utilisé est un modèle hiérarchique. Les variations de l'activité neurale dans le temps sont modélisées par le modèle neural phénoménologique présenté à la section 5.3.1.2. Le modèle qui permet le passage entre le niveau neural et la mesure BOLD est le modèle hémodynamique non linéaire développé en 5.3.2.2 (extended Balloon-Windkessel Model et BOLD Model). Les mesures BOLD dépendent ainsi d'autres variables d'état au cours du temps (neurales  $\vec{s}(t)$  et hémodynamiques  $\vec{z}(t), \vec{f}(t), \vec{v}(t), \vec{q}(t)$ ). Ces-dernières sont qualifiées d'états cachés.

#### 5.5.1.2 Estimation des paramètres de ce modèle à partir de données

A partir de ce modèle dynamique, il est possible de construire un modèle statistique sous forme d'un graphe probabiliste bayésien<sup>12</sup> qui décrit la série temporelle des mesures. On parle aussi de modèles génératifs puisqu'ils permettent de décrire la probabilité d'occurrence d'une série de mesures générées à partir d'un modèle dynamique des états cachés sous-jacents ([Friston, 2011]). Nous reviendrons plus en détails sur cet aspect dans le chapitre 10.

Les paramètres que l'on cherche à estimer sont des paramètres du modèle neural ( $\bar{A}$ ,  $\bar{B}^j$  et  $\bar{C}$  dans l'équation (5.4)), des paramètres hémodynamiques et BOLD (équations (5.44) et (5.48)), ainsi que des paramètres de bruit sur l'observable ([Friston et al., 2003], [Penny et al., 2004]). Pour l'étude de la connectivité effective entre régions pendant une tâche, nous nous intéressons particulièrement aux variations des paramètres de couplages neuraux que porte la matrice  $\bar{B}^j$  (cf présentation du modèle neural de DCM en 5.3.1.2). Une méthode d'inversion bayésienne variationnelle permet d'estimer ces paramètres à partir d'une série temporelle de mesures BOLD ([Stephan et al., 2007], [Friston, 2011], [Frässle et al., 2017]). En spécifiant les densités de probabilité a priori sur les paramètres

---

12. Réseau bayésien dynamique

à estimer, elle permet de calculer les densités de probabilité a posteriori et de comparer plusieurs modèles entre eux. Nous y reviendrons dans le chapitre 10.

#### DCM

- le modèle dynamique considéré : modèle neural phénoménologique (section 5.3.1.2), modèle du Ballon-Chambre à air étendu et modèle BOLD (5.3.2.2)
- le modèle statistique de la mesure : réseau bayésien dynamique
- les paramètres d'intérêt du modèle statistique : paramètres neuraux (couplages intrinsèques, variations de couplages et forçages), hémodynamiques et de bruit sur l'observable
- la méthode d'inversion de ces paramètres : méthode d'inversion bayésienne variationnelle

## 5.5.2 Exemple du contrôle inhibiteur des intrusions mnésiques

DCM peut être utilisé pour étudier le contrôle inhibiteur des intrusions mnésiques, et notamment pour tester une influence négative du DLPFC sur l'activité de l'hippocampe ([Benoit and Anderson, 2012]).

### 5.5.2.1 Quel modèle direct ?

[Benoit and Anderson, 2012], [Benoit et al., 2015] et [Schmitz et al., 2017] utilisent un modèle à deux régions : le DLPFC droit et l'hippocampe droit<sup>13</sup>. En l'absence de toute stimulation, chaque région possède un autocouplage inhibiteur (i.e. négatif) et les connexions intrinsèques entre les deux régions sont symétriques<sup>14</sup>. Celles-ci modélisent une interaction fonctionnelle qui peut être polysynaptique.

En No Think, il est possible de moduler des coefficients de couplage entre les régions et d'appliquer des forçages sur les régions.

Différentes hypothèses de modélisation proposent que ces modulations aient lieu soit :

1. nul part
2. de l'hippocampe vers le DLPFC (bottom-up)
3. du DLPFC vers l'hippocampe (top-down)
4. du DLPFC vers l'hippocampe, et de l'hippocampe vers le DLPFC (modulation bidirectionnelle)

Ces quatre hypothèses forment quatre "familles de modèles", représentées à la figure 5.6. Les propositions 3) et 4) sont dites congruentes avec l'hypothèse d'une modulation "top-down",

13. Nous faisons le choix de présenter en priorité des études DCM qui portent sur deux régions, parce que ce sera la configuration de référence étudiée dans la thèse (cf section 6.2). D'autres études DCM portent sur plus de régions (e.g. une région de contrôle et trois cibles dans [Gagnepain et al., 2014] et [Gagnepain et al., 2017]).

14. "bidirectional intrinsic connections" ([Benoit and Anderson, 2012])

contrairement à 1) et 2).

Dans chacun des cas, d'autres hypothèses de modélisation permettent de placer l'input pendant la condition No Think soit :

- a sur le DLPFC
- b sur l'hippocampe
- c sur les deux

On parle aussi de "sous-familles de modèles".

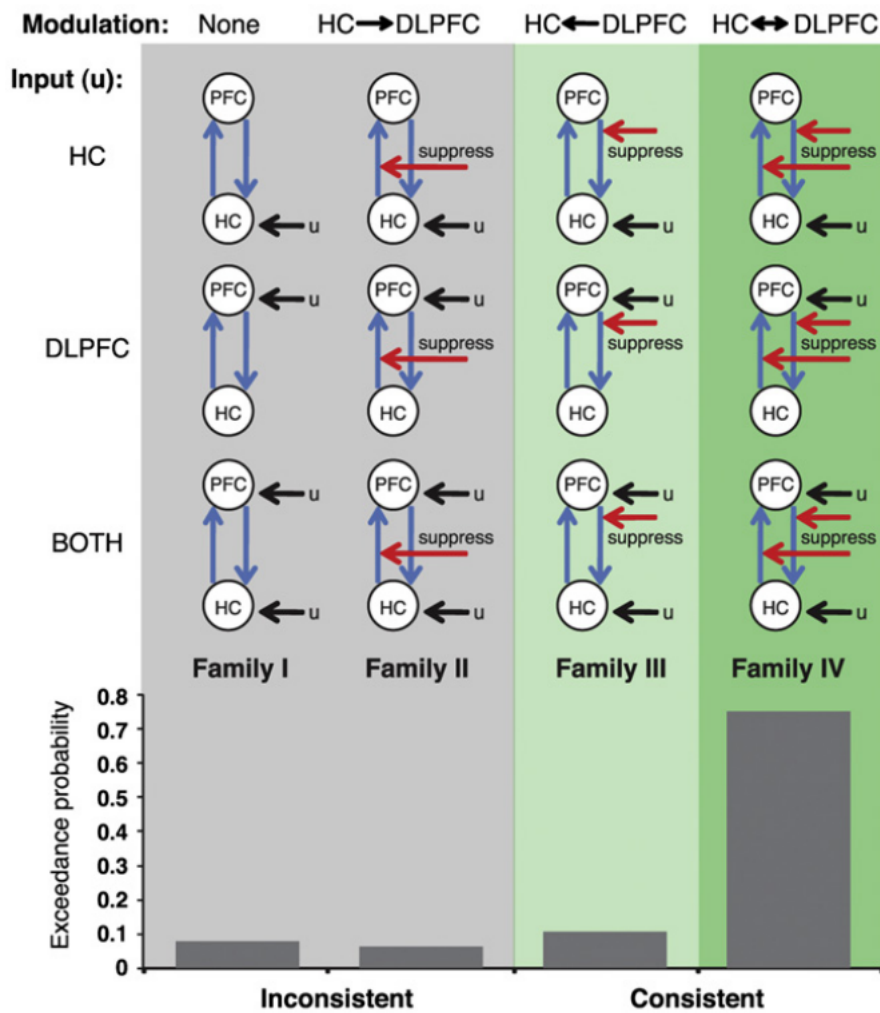


**Figure 5.6** – [Benoit and Anderson, 2012] consistent/inconsistent avec l'hypo top down

Pour chaque participant, il est possible d'inverser les paramètres de chacun de ces "modèles" (i.e. les paramètres intrinsèques, de modulation et les inputs). Les paramètres de modulation sont les mesures de connectivité effective qui représentent l'influence causale qu'une région exerce sur une autre pendant une condition expérimentale ([Friston, 2011]).

Dans la suite, nous présentons les résultats obtenus à l'échelle d'un groupe de participants en ce qu'ils permettent de mieux comprendre les mécanismes cérébraux du contrôle inhibiteur des intrusions mnésiques. Un des objectifs de la thèse étant de mener une analyse de premier niveau, i.e. sur un participant, nous ne détaillons donc pas ici les méthodes statistiques de groupe qui permettent d'obtenir ces résultats. Nous pourrions cependant nous reporter directement aux études citées dans le texte, ainsi qu'au travail de [Stephan et al., 2010] et [Penny et al., 2010].

L'objectif de l'analyse de groupe est double :

1. sélectionner les hypothèses de modélisation les plus probables. On parle de comparaison ou de sélection de modèles.<sup>15</sup>
2. estimer les paramètres intrinsèques, de modulation et les inputs de ces modèles<sup>16</sup>

### 5.5.2.2 Sélection de modèles à l'échelle d'un groupe de participants

Les résultats montrent d'abord que la famille 4 est la plus probable : au niveau de la connectivité, l'hypothèse de modélisation la plus probable est celle d'une modulation bidirectionnelle. Le modèle direct pour inverser les données comprend donc une modulation du DLPFC vers l'hippocampe, suggérant que le DLPFC exerce une influence sur l'activation de l'hippocampe pendant l'exercice d'un contrôle inhibiteur ([Benoit and Anderson, 2012], [Benoit et al., 2015], [Schmitz et al., 2017]).

Pour les inputs, les résultats sont plus contrastés : il est appliqué sur le DLPFC dans [Schmitz et al., 2017]<sup>17</sup>, sur le DLPFC et l'hippocampe dans [Benoit et al., 2015] et aucune des trois possibilités n'est plus probable qu'une autre dans [Benoit and Anderson, 2012].

### 5.5.2.3 Modulation de couplage du DLPFC vers l'hippocampe en lien avec les données comportementales

#### 5.5.2.3.1 Performances de rappel post-TNT

Les tests de rappels post-TNT permettent de comparer les taux de rappels des images associées aux mots verts (Think), rouges (No Think) et non présentés pendant la tâche (Baseline). Dans [Benoit and Anderson, 2012] les participants qui ont oublié le plus d'images associées aux mots No Think, en comparaison des images associées aux mots Baseline, sont qualifiés de "high forgetters", et se distinguent des "low forgetters".

Cette étude montre que seuls les "high forgetters" présentent un paramètre de modulation négatif du DLPFC vers l'hippocampe pendant la condition No Think. Par ailleurs, il n'y a pas de différence entre les participants au niveau du paramètre de couplage intrinsèque du DLPFC vers l'hippocampe. La connectivité effective absolue est mesurée comme la somme entre le paramètre intrinsèque et de modulation. A nouveau, seuls les "high forgetters" ont une connectivité effective absolue négative du DLPFC vers l'hippocampe.

Sous l'hypothèse du modèle direct de DCM, une augmentation d'activité du DLPFC cause

---

15. Bayesian Model Selection (BMS) est une méthode statistique qui indique la probabilité qu'un modèle donné soit plus probable qu'un autre pour générer les données (cf [Stephan et al., 2010]).

16. Bayesian Model Averaging (BMA)

17. uniquement pour les participants avec une concentration élevée de GABA dans l'hippocampe

donc chez ces participants une réduction d'activité de l'hippocampe en No Think, par rapport à la période hors stimulation. Ceci montre une influence top-down négative du DLPFC vers l'hippocampe ([Anderson et al., 2016]) pendant la condition No Think, chez les participants qui parviennent à mettre en place un mécanisme efficace de suppression mnésique. Autrement dit, les participants qui oublient plus sont ceux qui présentent une plus forte inhibition du DLPFC vers l'hippocampe. A noter que cette modulation négative n'est pas mesurée chez des participants à qui il a été demandé de mettre en place des stratégies de substitution de pensées ([Benoit et al., 2015]).

Dans leur discussion, [Benoit and Anderson, 2012] indique que chez les "high forgetters", la combinaison entre une activation plus importante du DLPFC<sup>18</sup> et un paramètre de modulation plus grand entre le DLPFC et l'hippocampe, conduit à un abaissement plus important de l'activité de l'hippocampe.

#### 5.5.2.3.2 Reports d'intrusions pendant le TNT

Dans l'étude de [Benoit et al., 2015], les participants reportent s'ils ont eu ou non des intrusions pendant la présentation du mot rouge. La tâche est divisée en plusieurs blocs. L'inversion des paramètres de DCM sur un ou deux blocs consécutifs permet de mener des estimations de la connectivité effective absolue<sup>19</sup> au cours de la tâche.

Les participants qui ont une plus grande baisse de la fréquence d'intrusions pendant la tâche sont ceux qui présentent un plus fort coefficient de couplage négatif entre le DLPFC et l'hippocampe en début de tâche, ainsi qu'une plus grande variation de connectivité effective absolue au cours de la tâche. Cette dernière devient de plus en plus faible et ceci s'expliquerait par le fait que les participants ont de moins en moins besoin de mettre en place un contrôle inhibiteur réactif pour chasser les intrusions. Ces résultats permettent alors de mettre en évidence un mécanisme top-down adaptatif qui permet de contrôler les intrusions mnésiques.

#### 5.5.2.4 Un mécanisme adaptatif

Comme nous l'avons vu à la section 2.2.4, les participants mettraient d'abord en place un mécanisme de contrôle prédictif ([Benoit et al., 2015])<sup>20</sup>. Si ce contrôle est suffisant, il n'y a pas d'intrusion. S'il n'est pas suffisant, l'intrusion apparaît à l'esprit. Il faudrait alors exercer un contrôle réactif supplémentaire pour chasser activement l'intrusion, ce qui conduirait à une baisse accrue de l'activité de l'hippocampe par rapport à l'exercice d'un seul contrôle prédictif (cf figure 15.4).

Pendant une intrusion, la trace mnésique réactivée deviendrait labile et elle pourrait ainsi être altérée par les mécanismes de contrôle inhibiteur ([Levy and Anderson, 2012], [Leone, 2021]). Ces mécanismes de suppression mnésique, mis en place pendant le contrôle d'une intrusion, permettraient de diminuer les reviviscences ultérieures ([Benoit et al., 2015],

---

18. mesurée par contrastes et non avec DCM puisqu'aucune des trois sous-familles de modèles n'est plus probable qu'une autre dans la famille 4

19. somme des paramètres intrinsèque et de modulation

20. basé sur une estimation personnelle de la probabilité d'avoir une intrusion au prochain essai (cf [Leone et al., 2022])



[Gagnepain et al., 2017]). La diminution des intrusions ne nécessiterait alors pas d'exercer autant de contrôle réactif qu'en début de tâche.

#### 5.5.2.5 Forçage dans le DLPFC et liens avec les concentration de GABA

[Schmitz et al., 2017] mesurent par spectroscopie la concentration de GABA dans l'hippocampe de sujets sains au repos. Chez les participants avec une grande concentration de GABA hippocampique, un forçage dans le DLPFC module l'activité de l'hippocampe pendant le No Think : le DLPFC exerce donc un contrôle inhibiteur top-down vers l'hippocampe. Au contraire, chez les participants avec une faible concentration de GABA hippocampique, un forçage dans le DLPFC ne semble pas moduler l'activité de l'hippocampe.

*Pendant une tâche de contrôle inhibiteur des intrusions mnésiques, la possibilité qu'un signal top-down entrant par le DLPFC abaisse l'activité de l'hippocampe semble donc dépendre de la concentration de GABA dans l'hippocampe.*

## 5.6 Le positionnement de la thèse

### 5.6.1 Limites des études actuelles sur le contrôle inhibiteur des intrusions mnésiques

Nous avons vu que le système GABAergique tonique de l'hippocampe joue un rôle important dans les processus cognitifs qui impliquent l'hippocampe (section 3.3), et notamment dans le contrôle inhibiteur des intrusions mnésiques (sections 3.4, 5.1.1.2, 5.4.4 et 5.5.2). Ce "poids inhibiteur" de l'hippocampe, qui reflète un état de la balance Excitation/Inhibition locale, a notamment été étudié par spectroscopie à travers des mesures de concentration de GABA dans l'hippocampe chez des sujets sains ([Schmitz et al., 2017]). Ces mesures de concentration indiquent cependant une combinaison de GABA intracellulaire (i.e. contenu dans les vésicules) et extracellulaire (i.e. synaptique et extrasynaptique) de tous les inter-neurons GABAergique contenus dans le volume de la région<sup>21</sup>. La concentration de GABA extracellulaire est directement impliquée dans les processus de transmission entre neurones, contrairement à la concentration de GABA intracellulaire qui constitue une réserve. Il serait intéressant de pouvoir préciser un "poids" GABAergique synaptique qui caractériserait directement l'intensité de la transmission entre les neurones contenus dans une région cérébrale. De plus, les mesures de concentration de GABA dans l'étude de [Schmitz et al., 2017] ne sont pas effectuées en même temps que la tâche de contrôle inhibiteur, mais au repos. Il serait intéressant de pouvoir caractériser l'état de la balance Excitation/Inhibition de l'hippocampe pendant la tâche de contrôle inhibiteur, c'est-à-dire directement à partir des données BOLD obtenues pendant la tâche.

Par ailleurs, les méthodes d'analyse classiques de signaux BOLD se basent sur des modèles statistiques de la mesure qui prennent en compte ou non un modèle dynamique de l'activité cérébrale. Parmi les modèles dynamiques standards, certains modélisent l'activité neurale et

---

21. cf discussion de [Schmitz et al., 2017]



les échanges entre régions, puis le couplage neurovasculaire et enfin la mesure BOLD (comme pour DCM, [Friston et al., 2003]). L'activité neurale est pensée comme le résultat global de processus excitateurs et inhibiteurs non explicites et cela ne permet pas d'étudier le rôle des neurones inhibiteurs localement présents dans les régions. La caractérisation des mécanismes neuraux impliqués dans une tâche de contrôle inhibiteur, en particulier de contrôle inhibiteur des intrusions mnésiques, reste donc très limitée en raison de l'absence de modèle neural précis en IRMf, c'est-à-dire de modèle qui explicite les dynamiques neurales excitatrices et inhibitrices dans les régions cérébrales.

### 5.6.2 La question que se pose la communauté

Comment inférer à partir de données d'IRMf des propriétés excitatrices et inhibitrices (E/I) neurales locales au sein de régions cérébrales impliquées dans une tâche de contrôle inhibiteur. Nous prenons comme cas d'exposition le contrôle inhibiteur des intrusions mnésiques.

# Deuxième partie

## Choix de modélisation

*Dans cette partie, je présente au chapitre 6 le modèle qui sera utilisé (i.e. analysé et inversé) dans tout le reste de la thèse. Le travail mené au chapitre 7 me permet d'énoncer la liste des questions très précises que cherche à éclairer la thèse.*

## Chapitre 6

# Modèle du signal BOLD construit à partir des variations de l'activité neurale excitatrice et inhibitrice

*Pour étudier la balance neurale E/I à partir d'un signal BOLD, il est nécessaire de travailler à partir d'un modèle dynamique qui fournit de manière explicite les évolutions temporelles des activités neurales excitatrice et inhibitrice. Je présente à la section 6.2 le modèle de masses neurales proposé par [Naskar et al., 2021] et que nous avons adopté. Celui-ci se base sur une description du fonctionnement neuronal<sup>1</sup> à l'échelle de la synapse, que nous décrivons dans un premier temps à la section 6.1. Nous n'interrogerons pas ce passage d'échelle en champ moyen.*

*Comme modèle du couplage neurovasculaire et de sa mesure BOLD, nous adoptons le modèle du "Balloon-Chambre à air étendu", comme d'autres travaux de la communauté neuroscientifique (e.g. [Wang et al., 2019]), parce qu'il permet de prendre en compte des aspects non linéaires de la réponse BOLD (cf section 5.3.2.2 du chapitre précédent). À partir de cette modélisation du passage d'échelle de l'activation synaptique au signal BOLD, en passant par l'échelle neurale, nous souhaitons proposer un nouveau modèle neural du contrôle inhibiteur. Pour un historique de l'utilisation des modèles de masses neurales, il est possible de se rapporter à l'introduction de l'article (chapitre 9).*

### 6.1 Modélisation de la dynamique d'association-dissociation entre récepteurs et neurotransmetteurs au niveau d'une synapse

Lorsqu'un neurone présynaptique décharge, il libère dans la fente synaptique des neurotransmetteurs. Ces neurotransmetteurs diffusent dans la fente et peuvent se fixer aux récepteurs du neurone post-synaptique (cf chapitre 3). Nous allons présenter un modèle de

---

1. voir ou revoir encadré 2.3.2.3 pour la distinction entre "neural" et "neuronal"

la réaction d'association et de dissociation des récepteurs membranaires post-synaptiques aux neurotransmetteurs.

### 6.1.1 A l'échelle d'un récepteur

#### 6.1.1.1 Modélisation de l'état d'un récepteur par une chaîne de Markov entre $t$ et $t + dt$

Nous notons  $R$  l'état d'un récepteur non lié à un neurotransmetteur (récepteur inactif) et  $RN$  l'état d'un récepteur lié à un neurotransmetteur (récepteur actif). L'évolution de l'état d'un récepteur entre  $t$  et  $t + dt$  peut être modélisée à partir de la chaîne de Markov représentée sur la figure 6.1 :

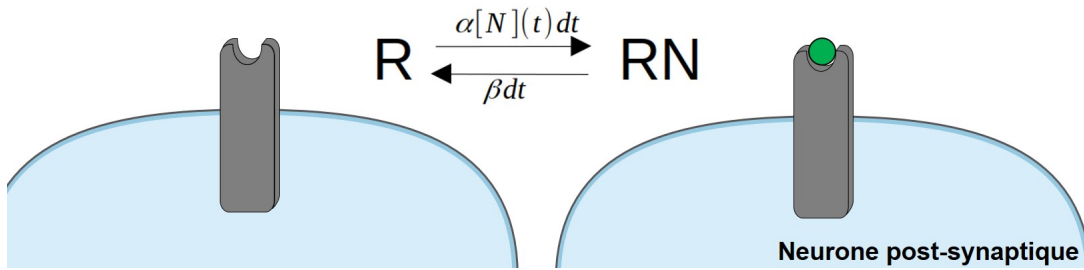


Figure 6.1

Définition des probabilités de transition entre ces deux états :

- $p(R, t + dt | RN, t)$  est la probabilité qu'un récepteur lié ( $RN$ ) à  $t$  perde son neurotransmetteur pendant  $dt$  et se retrouve donc dans l'état non lié ( $R$ ) à  $t + dt$ . Elle est supposée constante au cours du temps :  $p(R, t + dt | RN, t) \equiv \beta dt$ .  $\beta$  est exprimée en  $s^{-1}$ .
- $p(RN, t + dt | R, t)$  est la probabilité qu'un récepteur non lié ( $R$ ) à  $t$  gagne un neurotransmetteur pendant  $dt$  et se retrouve donc dans l'état lié ( $RN$ ) à  $t + dt$ . Elle augmente proportionnellement avec la concentration de neurotransmetteurs libres qui diffusent dans la fente synaptique à  $t$ , notée  $[N](t)$  :  $p(RN, t + dt | R, t) \equiv \alpha [N](t) dt$ .  $[N]$  est une concentration molaire exprimée en  $mol L^{-1} \equiv M$ .  $\alpha$  est exprimée en  $s^{-1} M^{-1}$ .

#### 6.1.1.2 Evolution temporelle de la probabilité de l'état lié du récepteur

Nous notons  $p(\eta, t)$  et  $p(\eta, t + dt)$  les probabilités d'observer un récepteur dans un état  $\eta$  (lié  $RN$  ou non lié  $R$ ) respectivement à  $t$  et  $t + dt$ .

La probabilité d'observer l'état  $RN$  à  $t + dt$  dépend des trois événements suivants :

- un récepteur est non lié ( $R$ ) à  $t$  et accueille un neurotransmetteur pendant  $dt$ . Sa probabilité vaut  $p(RN, t + dt | R, t)p(R, t)$
- un récepteur est lié ( $RN$ ) à  $t$  et perd son neurotransmetteur pendant  $dt$ . Sa probabilité vaut  $p(R, t + dt | RN, t)p(RN, t)$

- un récepteur est lié (RN) à  $t$ , et l'est toujours à  $t + dt$  (il ne se passe rien pendant  $dt$ ).  
Sa probabilité vaut  $p(RN, t)$

Nous pouvons écrire la probabilité d'être dans l'état RN à  $t + dt$  comme la somme des probabilités des trois événements précédents (avec un signe négatif pour signifier que le récepteur perd le neurotransmetteur) :

$$p(RN, t + dt) = p(RN, t) - p(R, t + dt|RN, t)p(RN, t) + p(RN, t + dt|R, t)p(R, t) \quad (6.1)$$

En reprenant les probabilités de transition définies au paragraphe 6.1.1.1, nous avons :

$$p(RN, t + dt) = p(RN, t) - \beta dt p(RN, t) + \alpha[N](t) dt p(R, t) \quad (6.2)$$

En réorganisant les termes, nous obtenons :

$$\frac{p(RN, t + dt) - p(RN, t)}{dt} = -\beta p(RN, t) + \alpha[N](t) p(R, t) \quad (6.3)$$

De plus, le récepteur se trouve soit dans l'état R soit dans l'état RN, à tout instant. Nous pouvons donc écrire :

$$p(RN, t) + p(R, t) = 1 \quad (6.4)$$

L'équation (6.3) devient :

$$\frac{p(RN, t + dt) - p(RN, t)}{dt} = -\beta p(RN, t) + \alpha[N](t)(1 - p(RN, t)) \quad (6.5)$$

Nous définissons  $s(t) \equiv p(RN, t)$  et  $s(t + dt) \equiv p(RN, t + dt)$ , les probabilités d'observer un récepteur dans l'état lié respectivement à  $t$  et  $t + dt$ . Nous pouvons alors écrire :

$$\frac{s(t + dt) - s(t)}{dt} = -\beta s(t) + \alpha[N](t)(1 - s(t)) \quad (6.6)$$

En posant  $ds \equiv s(t + dt) - s(t)$ , nous obtenons :

$$\boxed{\frac{ds}{dt} = -\beta s(t) + \alpha[N](t)(1 - s(t))} \quad (6.7)$$

Cette équation différentielle régit l'évolution temporelle de la probabilité qu'un récepteur soit lié à un neurotransmetteur (état RN)<sup>2</sup>.

---

2. Il est nécessaire de passer par une intégration entre  $t$  et  $t + dt$  pour construire un intuitif, comme nous l'avons présenté.

## 6.1.2 A l'échelle de la synapse

### 6.1.2.1 Evolution temporelle de la fraction post-synaptique de récepteurs liés

Nous notons  $[N](t)$  la concentration molaire à  $t$  (en  $\text{molL}^{-1} \equiv M$ ) de neurotransmetteurs libres dans la fente synaptique.  $[RN](t)$  et  $[R](t)$  sont les concentrations molaires de récepteurs à  $t$  sur la membrane de la synapse, liés et non liés à un neurotransmetteur.

Nous formulons quelques hypothèses de modélisation supplémentaires :

1. Nous considérons que la concentration en neurotransmetteurs  $[N]$  dans la fente est suffisamment grande (devant  $[R](t)$  et  $[RN](t)$ ) pour ne pas diminuer avec les dynamiques d'association-dissociation entre récepteurs et neurotransmetteurs
2. Nous supposons que le nombre total de récepteurs présents sur le neurone post-synaptique est fixe : la somme des récepteurs liés et non liés à un neurotransmetteur est toujours la même et on peut écrire  $[R](t) + [RN](t) = Cte$ .

$\frac{[RN](t)}{[R](t) + [RN](t)}$  est la fraction de récepteurs liés à un neurotransmetteur (RN), parmi tous les récepteurs liés (RN) et non liés (R). L'interprétation fréquentiste de la probabilité nous permet d'écrire une égalité entre cette fraction et la probabilité  $s$  pour un récepteur d'être dans l'état RN (introduite au paragraphe précédent) :

$$\frac{[RN](t)}{[R](t) + [RN](t)} \equiv s(t) \quad (6.8)$$

En considérant les récepteurs non liés (R) nous avons :

$$\frac{[R](t)}{[R](t) + [RN](t)} = \frac{[R](t) + [RN](t) - [RN](t)}{[R](t) + [RN](t)} = 1 - \frac{[RN](t)}{[R](t) + [RN](t)} = 1 - s(t) \quad (6.9)$$

$1 - s(t)$  correspond donc à la fraction de récepteurs non liés, parmi tous les récepteurs de la synapse.

L'équation différentielle (6.7) donne donc l'évolution temporelle de la fraction de récepteurs liés à un neurotransmetteur (RN), parmi tous les récepteurs de la synapse<sup>3</sup>.

### 6.1.2.2 Evolution temporelle de la fraction post-synaptique de canaux ouverts

Chaque récepteur du neurone post-synaptique est couplé à un canal ionique. L'association d'un neurotransmetteur à un récepteur provoque l'ouverture du canal associé, qui laisse entrer des ions dans le neurone post-synaptique (cf chapitre 3). A un état du récepteur correspond ainsi un unique état du canal ionique, depuis lequel nous pourrions décrire le système.

---

3. En chimie des solutions, un ligand (L) peut s'associer à un récepteur pour former un complexe (C). L'évolution temporelle de la concentration en complexes ligands-recepteurs (C) est décrite par l'équation différentielle du premier ordre suivante :  $\frac{d[C]}{dt} = k_+[L][R] - k_-[C]$ .  $k_+$  est la constante cinétique d'association (en  $s^{-1}$ ) et  $k_-$  la constante cinétique de dissociation (en  $M s^{-1} \equiv \text{molL}^{-1} s^{-1}$ ). Dans notre cas, en remplaçant  $s = \frac{[RN]}{[R] + [RN]}$  dans l'équation différentielle (6.7), nous retrouvons bien l'équation d'association-dissociation entre un ligand et un récepteur en chimie des solutions.

Ainsi, l'état  $\underline{RN}$  du récepteur correspond à l'état d'un canal ouvert. L'état  $\underline{R}$  correspond à celui d'un canal fermé.

La variable  $s$  qui est la fraction de récepteurs liés à un neurotransmetteur (parmi tous les récepteurs membranaires post-synaptiques), désigne ainsi également la fraction de canaux ouverts (parmi tous les canaux membranaires post-synaptiques). Elle est désignée en anglais par les expressions "synaptic gating variable" ([Deco et al., 2018]), "fraction of open channels" ([Deco et al., 2014]).

L'équation différentielle (6.7) donne ainsi l'évolution temporelle de la fraction de canaux ouverts  $s(t)$  de la synapse,  $\alpha[N](t)$  et  $\beta$  sont des densités de probabilité de transition entre les états "canal fermé" et "canal ouvert".  $\alpha[N](t)dt$  est la fraction de canaux fermés à  $t$  qui se seront ouverts à  $t + dt$ .  $\beta dt$  est la fraction de canaux ouverts à  $t$  qui se seront fermés à  $t + dt$ .

### 6.1.2.3 Décharge présynaptique et libération de neurotransmetteurs

L'équation (6.7) fait intervenir  $[N](t)$ , la concentration de neurotransmetteurs libres qui diffusent dans la fente synaptique. Celle-ci varie dans le temps en fonction de l'activité du neurone présynaptique. Plus le neurone présynaptique décharge, c'est-à-dire plus la fréquence de ses potentiels d'action est élevée, plus la concentration de neurotransmetteurs libérés est grande.

[Naskar et al., 2021] font une proposition pour prendre en compte cette dépendance entre  $[N](t)$  et la fréquence des potentiels d'action au cours du temps.

Soit  $r(t)$  la fréquence des potentiels d'action présynaptiques à l'instant  $t$  et  $[T]$  la concentration de neurotransmetteurs libérée par hertz (en  $\text{M Hz}^{-1} = \text{M.s}$ ), qui est fixée.

La concentration de neurotransmetteurs libérée quand le neurone présynaptique décharge avec une fréquence  $r(t)$  à l'instant  $t$  s'écrit :

$$[N](t) = [T] r(t) \quad (6.10)$$

La concentration de neurotransmetteurs libérée  $[N]$  dépend du temps et n'est impactée que par la décharge présynaptique (i.e. la fréquence de potentiels d'action  $r(t)$ ). Elle est supposée suffisamment grande pour ne pas varier avec les dynamiques d'association-dissociation entre récepteurs et neurotransmetteurs en post-synaptique, ni avec la recapture de neurotransmetteurs par le neurone pré-synaptique.

L'équation (6.7) devient alors :

$$\frac{ds}{dt} = -\beta s(t) + \alpha [T] r(t) (1 - s(t)) \quad (6.11)$$

### 6.1.2.4 Canaux excitateurs et canaux inhibiteurs

Cette modélisation s'applique pour un récepteur exciteur ou inhbiteur, couplé à un canal exciteur ou inhbiteur<sup>4</sup>.

---

4. i.e qui laisse entrer des ions  $\text{Na}^+$  pour le premier ou  $\text{Cl}^-$  pour le deuxième.



Un neurone post-synaptique peut recevoir du signal de neurones présynaptiques excitateurs et inhibiteurs en même temps (cf figure 6.2). Nous pouvons alors distinguer sur le neurone post-synaptique une fraction de canaux excitateurs ouverts (parmi tous les canaux excitateurs) et une fraction de canaux inhibiteurs ouverts (parmi tous les canaux inhibiteurs).

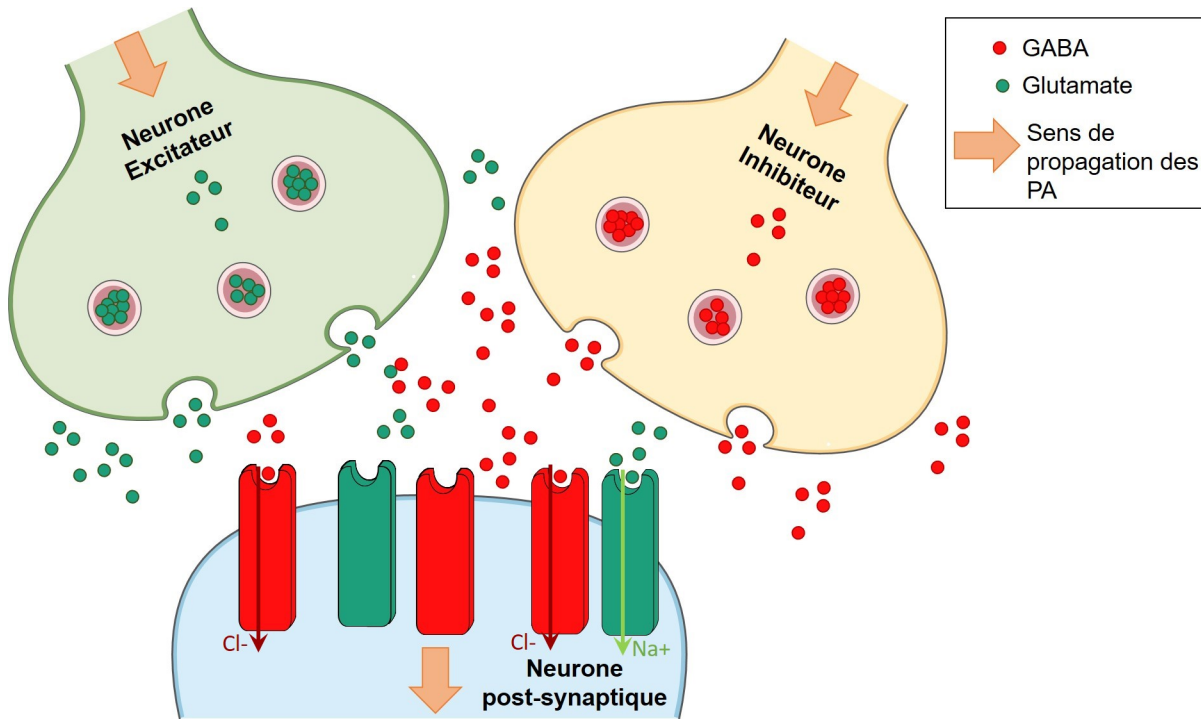


**Figure 6.2** – Schéma d'une synapse avec libération de neurotransmetteurs excitateurs (E) par un neurone présynaptique E et de neurotransmetteurs inhibiteurs (I) par un neurone présynaptique I. Ces neurotransmetteurs E ou I se fixent de manière spécifique sur les récepteurs E ou I du neurone post-synaptique. Lorsqu'un neurotransmetteur se fixe sur un récepteur, un canal ionique s'ouvre (représenté par une flèche). Le neurone post-synaptique possède 2 récepteurs E (dont 1 lié à un neurotransmetteur) et 3 récepteurs I (dont 2 liés à un neurotransmetteur). La fraction de récepteurs E liés, plus couramment décrite comme la fraction de canaux E ouverts vaut :  $sn = 1/2$ . La fraction de canaux I ouverts vaut :  $sg = 2/3$ .

Ce modèle dynamique de l'activité synaptique servira d'hypothèse de modélisation pour le modèle neural en champ moyen décrit à la section 6.2

## 6.2 Modélisation de la dynamique neurale à l'échelle des régions : approche en champ moyen

Le cerveau est constitué d'un très grand nombre de neurones et de synapses. Le néo-cortex humain serait constitué approximativement de l'ordre de  $10^{10}$  neurones, chaque neurone effectuant en moyenne de l'ordre de  $10^4$  connexions synaptiques avec d'autres neurones. Dans un centimètre cube de cortex, l'ordre de grandeur estimé est de  $10^{12}$  synapses ([Drachman, 2005]).

Pour modéliser le fonctionnement cérébral à l'échelle de régions cérébrales (quelques  $cm^3$ ), il est impossible de prendre en compte les dynamiques synaptiques une par une. Des modèles en champs moyen, comme le Dynamic Mean Field Model (DMFM) permettent d'approximer une dynamique synaptique moyenne à l'échelle des régions<sup>5</sup> ([Brunel and Wang, 2001], [Wong and Wang, 2006], [Deco and Jirsa, 2012], [Deco et al., 2013], [Deco et al., 2014]). L'activité synaptique "moyenne" correspond à la fraction moyenne de canaux ouverts, i.e. de récepteurs synaptiques liés à un neurotransmetteur (état RN). Elle dépend de l'activité moyenne des neurones qui déchargent et libèrent des neurotransmetteurs dans les synapses. Le comportement moyen d'un groupe de neurones, en anglais : pool, se base sur le fonctionnement neuronal décrit au chapitre 3. Ainsi, un pool est caractérisé par un courant synaptique moyen et une fréquence moyenne de potentiels d'action, qui influencent les évolutions dans le temps des fractions moyennes de canaux post-synaptiques ouverts.

Certains modèles considèrent uniquement un "pool excitateur" dont l'activité moyenne résulte des interactions implicites avec les neurones inhibiteurs de la région : ce sont des modèles à un pool par région (e.g. [Deco et al., 2013], [Wang et al., 2019]). Au contraire, d'autres modèles prennent explicitement en compte "pool excitateur" et "pool inhibiteur" : ces modèles comprennent deux pools par région (e.g. [Deco and Jirsa, 2012], [Deco et al., 2018], [Naskar et al., 2021]).

[Naskar et al., 2021] différencie explicitement deux types de neurones : les neurones excitateurs et les neurones inhibiteurs, qui vont avoir un effet sur deux types de canaux différents, les canaux excitateurs couplés aux récepteurs excitateurs (NMDA) et les canaux inhibiteurs couplés aux récepteurs inhibiteurs (GABA). Le pool excitateur est caractérisé par un courant excitateur moyen ( $xn$ ) et une fréquence moyenne de potentiel d'action ( $rn$ ), qui fait évoluer la fraction moyenne de canaux excitateur ouverts ( $sn$ ). De même, le pool inhibiteur est décrit par un courant inhibiteur moyen ( $xg$ ) et une fréquence moyenne de potentiel d'action ( $rg$ ), qui fait évoluer la fraction moyenne de canaux inhibiteurs ouverts ( $sg$ ). Les auteurs proposent de décrire la dynamique d'association/dissociation moyenne des neurotransmetteurs sur les récepteurs synaptiques en se basant directement sur un modèle à l'échelle d'une synapse (décrit à la section 6.1). Ils précisent qu'il s'agit d'un "multiscale Dynamic Mean Field Model" pour expliciter le lien entre l'échelle synaptique moyenne et l'échelle synaptique individuelle.

Pour étudier la variabilité interindividuelle du contrôle inhibiteur à partir de données obtenues en IRMf, nous avons exposé dans la partie I, et notamment dans les conclusions du chapitre 5, la nécessité de prendre en compte un modèle neural qui explicite les dyna-

---

5. pour plus de détails, voir l'introduction de l'article du chapitre 9

miques neurales excitatrices et inhibitrices dans les régions cérébrales. En ce sens, le modèle proposé par [Naskar et al., 2021] nous semble pertinent. Par ailleurs, il se base sur des hypothèses biologiques du fonctionnement des neurones à l'échelle de la synapse, contrairement à d'autres modèles phénoménologiques plus couramment employés (DCM), ce qui le rend particulièrement intéressant pour l'étude de la variabilité interindividuelle de l'état des balances E/I dans des régions de contrôle et des régions cibles.

Dans la suite, nous présentons en détails les différents étages de ce modèle hiérarchique. Pour  $N$  régions cérébrale, il se compose de  $2N$  variables d'état (2 activités synaptiques moyennes par région, une excitatrice  $sn$  et une inhibitrice  $sg$ ), et de  $4N$  variables intermédiaires (2 courants moyens par région, un excitateur  $xn$  et un inhibiteur  $xg$ , ainsi que 2 fréquences moyennes de potentiels d'action, un excitateur  $rn$  et un inhibiteur  $rg$ ). Nous donnons également les valeurs des paramètres qui seront utilisées dans nos codes, sans changements.

## 6.2.1 Activités synaptiques moyennes

### 6.2.1.1 Equations d'évolution temporelle des fractions moyennes de canaux ouverts

Pour une région  $i$ ,  $sn_i$  (resp.  $sg_i$ ) est une mesure de l'activité synaptique moyenne du pool excitateur (resp. inhibiteur).

$sn_i$  représente la fraction moyenne de canaux excitateurs ouverts<sup>6</sup>.  $sg_i$  représente la fraction moyenne de canaux inhibiteurs ouverts<sup>7</sup>.  $sn_i$  et  $sg_i$  sont donc des grandeurs sans unité, comprises entre 0 et 1 inclus.

L'équation (6.11) donne l'évolution temporelle de l'activité synaptique moyenne de chaque pool (excitateur ou inhibiteur), et se réécrit en les variables du modèle par :

$$\begin{cases} s\dot{n}_i(t) = -\beta^E sn_i(t) + \alpha^E T_{glu} rn_i(t)(1 - sn_i(t)) \\ s\dot{g}_i(t) = -\beta^I sg_i(t) + \alpha^I T_{gaba} rg_i(t)(1 - sg_i(t)) \end{cases} \quad (6.12)$$

La signification des constantes  $\alpha$  et  $\beta$  a été donnée à la section 6.1.

Les concentrations molaires moyennes de glutamate et GABA libérés s'écrivent respectivement  $T_{glu} rn_i(t)$  et  $T_{gaba} rg_i(t)$ . Elles dépendent de :

- $T_{glu}$  et  $T_{gaba}$  (en  $M \text{ Hz}^{-1}$ ) : les concentrations molaires de glutamate et GABA libérés pour une décharge présynaptique de 1Hz.
- les fréquences de potentiels d'action  $rn_i(t)$  et  $rg_i(t)$  (en Hz).

La propagation des potentiels d'action le long des neurones équivalents moyens va induire la libération de neurotransmetteurs dans la fente synaptique. Leur concentration augmente avec la fréquence moyenne des potentiels d'action. Ceci est en cohérence avec le fonctionnement d'un neurone décrit au chapitre 3).

6. i.e. de récepteurs NMDA liés à un neurotransmetteur glutamate

7. i.e. de récepteurs GABA-A liés à un neurotransmetteur GABA

### 6.2.1.2 Bruit

Dans le modèle de [Naskar et al., 2021], un processus de Wiener de variance  $\sigma^2$  permet d'ajouter un bruit gaussien aux évolutions temporelles de  $sn_i(t)$  et  $sg_i(t)$ , et nous le traduisons par les équations :

$$\begin{cases} dsn_i(t) = \left( -\beta^E sn_i(t) + \alpha^E T_{glu} r n_i(t) (1 - sn_i(t)) \right) dt + \sigma \nu_i(t) \\ dsg_i(t) = \left( -\beta^I sg_i(t) + \alpha^I T_{gaba} r g_i(t) (1 - sg_i(t)) \right) dt + \sigma \nu_i(t) \end{cases} \quad (6.13)$$

où  $\nu_i(t)$  est un processus de Wiener de variance 1. La valeur pour la variance du bruit utilisée dans [Naskar et al., 2021] est très faible ( $\sigma = 10^{-3}$  nA) devant l'intervalle de variation entre 0 et 1 pour  $sn$  et  $sg$ . Il est donc probable que ce bruit ne joue alors aucun rôle. Pour des valeurs de bruit plus significatives, une étude complète de son rôle serait sans doute bienvenue (en particulier pour déterminer si le bruit reste sommatif ou s'il peut devenir multiplicatif, et à partir de quel seuil...). À ce stade, nous laissons ce terme de bruit de côté (il ne servira pas dans l'analyse). Un terme de bruit sera en revanche considéré dans l'inversion, à l'échelle du signal BOLD.

### 6.2.1.3 Valeurs des paramètres

Nous reprenons les valeurs de [Naskar et al., 2021].

Paramètres pour le pool exciteur :

$$\beta^E = 6.6 \text{ s}^{-1}, \alpha^E = 72 \text{ s}^{-1} \text{ mM}^{-1}$$

[Naskar et al., 2021] proposent un intervalle de valeurs<sup>8</sup> de  $T_{glu}$  chez le sujet sains :  $T_{glu} \in [6.1; 10.5]$  mM ms, dont nous prenons la médiane approchée :  $T_{glu} = 8 \cdot 10^{-3}$  mM s

Paramètres pour le pool inhibiteur :

$$\beta^I = 180 \text{ s}^{-1}, \alpha^I = 530 \text{ s}^{-1} \text{ mM}^{-1}.$$

[Naskar et al., 2021] proposent un intervalle de valeurs de  $T_{gaba}$  chez le sujet sains :  $T_{gaba} \in [1.5; 4.2]$  mM ms, dont nous prenons la médiane approchée :  $T_{gaba} = 3 \cdot 10^{-3}$  mM s.

Nous notons que les valeurs des paramètres  $\alpha^E$ ,  $\beta^E$ ,  $\alpha^I$  et  $\beta^I$  utilisées par [Naskar et al., 2021] sont concordantes avec celles de [Destexhe et al., 1994] page 18, pour un récepteur NMDA et un récepteur GABA. [Naskar et al., 2021] décrivent bien une évolution des activités synaptiques moyennes à l'échelle neurale à partir d'un modèle d'association-dissociation à l'échelle d'un récepteur, étendu à une synapse (comme décrit à la section 6.1). Leur modèle sur les fréquences moyennes de potentiels d'action et courants ioniques moyens,

---

8. L'article présente une erreur d'unité : les concentrations  $T_{glu}$  et  $T_{gaba}$  devraient être exprimées en mM ms et non en mM (avec la notation de la concentration molaire  $M \equiv \text{mol/L}$ ). Cet élément est important pour les conversions d'unité lorsque nous souhaitons travailler dans les unités du Système International, comme nous l'avons fait.

que nous présentons ci-dessous, se base sur les travaux antérieurs qui proposent une construction analytique du passage en champ moyen pour passer d'un modèle neuronal à un modèle neural (e.g. [Brunel and Wang, 2001], [Wong and Wang, 2006], [Deco and Jirsa, 2012], [Deco et al., 2013], [Deco et al., 2014]).

## 6.2.2 Fréquences moyennes de potentiels d'action

### 6.2.2.1 Fonctions des courants ioniques

Pour une région  $i$ ,  $rn_i$  (resp.  $rg_i$ ) est la fréquence moyenne des potentiels d'action dans un neurone excitateur (resp. inhibiteur). Elle est exprimée en Hertz ( $s^{-1}$ ) et est positive ( $\mathbb{R}^+$ ).

Elle dépend des courants ioniques moyens. En effet, nous avons vu que la naissance de potentiels d'action est due à la dépolarisation membranaire du neurone provoquée par le passage d'ions au travers de sa membrane (cf chapitre 3).

$rn_i$  (resp.  $rg_i$ ) s'écrit avec une fonction croissante du courant ionique moyen excitateur  $x_n$  (resp. inhibiteur  $x_g$ ) en entrée du pôle excitateur (resp. inhibiteur). Elle se modélise sous la forme d'une sigmoïde, d'après les travaux de [Abbott and Chance, 2005] :

$$\begin{cases} rn_i(t) = \frac{a_E x n_i(t) - b_E}{1 - e^{-d_E(a_E x n_i(t) - b_E)}} \\ rg_i(t) = \frac{a_I x g_i(t) - b_I}{1 - e^{-d_I(a_I x g_i(t) - b_I)}} \end{cases} \quad (6.14)$$

Au-delà d'un certain seuil, la fréquence des potentiels d'action devient proportionnelle au courant (cf figure 6.3). Plus le courant est élevé, plus la fréquence des potentiels d'actions est élevée.

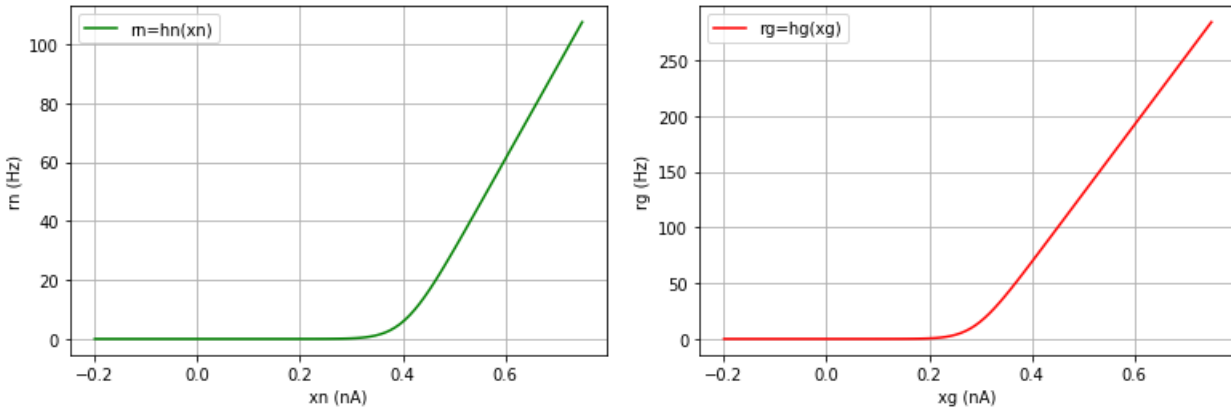


**Figure 6.3** – Fréquence moyenne de potentiels d'action en fonction de l'amplitude du courant ionique moyen.

A gauche : pour le pool excitateur. A droite : pour le pool inhibiteur

### 6.2.2.2 Valeurs des paramètres

Nous reprenons les valeurs de [Naskar et al., 2021].

Paramètres pour le pool exciteur :

$$a_E = 310 \text{ nC}^{-1}, b_E = 125 \text{ Hz}, d_E = 0.16 \text{ s}.$$

Paramètres pour le pool inhibiteur :

$$a_I = 615 \text{ nC}^{-1}, b_I = 177 \text{ Hz}, d_I = 0.087 \text{ s}.$$

### 6.2.3 Courants ioniques moyens

Pour une région  $i$ ,  $xn_i$  (resp.  $xg_i$ ) représente le courant ionique moyen en entrée du pool exciteur (resp. inhibiteur). Ce sont des grandeurs exprimés en nanoampères (nA), définies sur  $\mathbb{R}$ .

Ces variables dépendent des fractions moyennes de canaux ouverts  $sn_i$  et  $sg_i$ . Elles résultent d'une somme de différentes sources de courant, conformément au processus biologique d'intégration neuronale. Nous allons détailler les différents termes qui sont représentés à la figure 6.4.

$$\begin{cases} xn_i(t) = W_+ J_{nmda} sn_i(t) - J_{gaba,i} sg_i(t) + W_E I_0 + G J_{nmda} \sum_{j \neq i} C_{ij} sn_j(t) \\ xg_i(t) = J_{nmda} sn_i(t) - W_- sg_i(t) + W_I I_0 \end{cases} \quad (6.15)$$

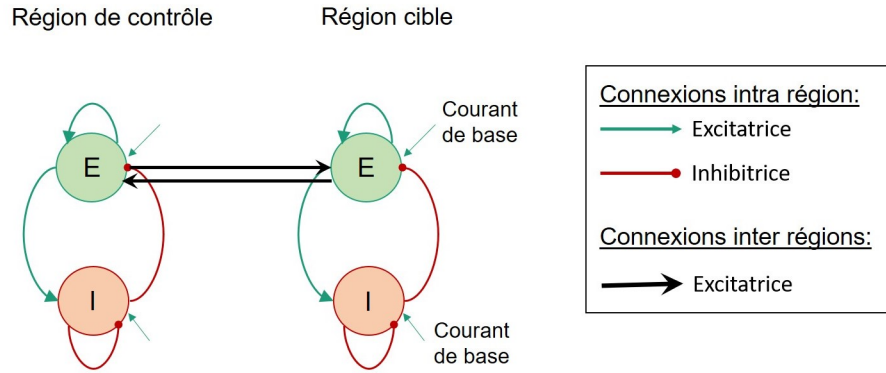


**Figure 6.4** – Illustration des courants ioniques pour deux régions couplées (une région de contrôle et une région cible).

#### 6.2.3.1 Courant ionique moyen en entrée d'un pool exciteur

La courant ionique moyen  $xn_i$  est généré suite à la fixation des neurotransmetteurs glutamate et GABA sur les récepteurs excitateurs et inhibiteurs situés sur les neurones du pool

excitateur.

Ces neurotransmetteurs ont été libérés par :

- les neurones de la même région :
  - neurones excitateurs, dont l'effet est modélisé par une connexion récurrente du pool excitateur sur lui-même :  $W_+ J_{nmda} s n_i$  (représenté sur la figure 6.4 par une flèche verte d'un pool E vers lui-même).
  - neurones inhibiteurs, dont l'effet est modélisé par une connexion du pool inhibiteur vers le pool excitateur :  $-J_{gaba,i} s g_i$  (représenté sur la figure 6.4 par une flèche rouge d'un pool I vers le pool E de la même région).
- les neurones excitateurs (uniquement<sup>9</sup>) des régions voisines. La connexion du pool excitateur d'une région  $j$  vers le pool excitateur d'une région  $i$  est traduite par le terme  $G J_{nmda} \sum_{j \neq i} k_E C_{ij} s n_j$  (représenté sur la figure 6.4 par une flèche noire du pool E d'une région vers le pool E de l'autre région).  
 $C_{ij}$  sont les éléments de la matrice de connectivité structurale  $\overline{\overline{C}}$ . Chaque coefficient  $C_{ij}$  est proportionnel au nombre total de fibres de  $j$  vers  $i$ , une donnée mesurée pour chaque sujet. Ils sont compris entre 0 et 1 inclus (la matrice est normalisée<sup>10</sup>). La matrice est de plus symétrique ( $C_{ij} = C_{ji}$ ) : on suppose que le nombre de fibres qui transporte du signal de  $j$  vers  $i$  est identique à celui de  $i$  vers  $j$ . La constante  $G$  est un paramètre qui régule l'intensité des couplages inter-régions.

Un courant ionique constant permet d'assurer une activité basale dans le pool excitateur :  $W_E I_0$  (représenté sur la figure 6.4 par une flèche verte sur le pool E de la région).

### 6.2.3.2 Courant ionique moyen en entrée d'un pool inhibiteur

Le courant ionique inhibiteur moyen  $xg$  est généré suite à la fixation des neurotransmetteurs glutamate et GABA sur les récepteurs excitateurs et inhibiteurs situés sur les neurones du pool inhibiteur.

Ces neurotransmetteurs ont été libérés par les neurones de la même région :

- neurones excitateurs, dont l'effet est modélisé par une connexion du pool excitateur vers le pool inhibiteur :  $J_{nmda} s n_i$  (représenté sur la figure 6.4 par une flèche verte d'un pool E vers le pool I de la même région).
- neurones inhibiteurs, dont l'effet est modélisé par une connexion récurrente du pool inhibiteur sur lui-même :  $-W_- s g_i$  (représenté sur la figure 6.4 par une flèche rouge d'un pool I vers lui-même).

Un courant ionique constant permet d'assurer une activité basale dans le pool inhibiteur :  $W_I I_0$  (représenté sur la figure 6.4 par une flèche verte sur le pool I de la région).

---

9. Les neurones excitateurs et inhibiteurs seraient connectés uniquement localement dans les régions cérébrales. Au contraire, la communication entre régions cérébrales passerait par des neurones excitateurs uniquement.

10. Division de tous les éléments de la matrice de connectivité structurale brute par l'élément le plus grand : la matrice porte donc l'information de l'importance relative des couplages inter-région

### 6.2.3.3 Valeurs des paramètres

Nous reprenons les valeurs de [Naskar et al., 2021].

Courants intra région :

- dus aux couplages intra région :
  - Autocouplage excitateur :  $W_+ = 1.4$ ,  $J_{nmda} = 0.15$  nA
  - Couplage inhibiteur vers excitateur :  $J_{gaba,i} > 0$   
 Un paramètre par région (en nA), qui sera l'objet d'exploration. Sa valeur de référence est fixée par une procédure d'inversion permettant d'assurer un régime de repos à 3 Hz<sup>11</sup>.
  - Couplage excitateur vers inhibiteur :  $J_{nmda} = 0.15$  nA
  - Autocouplage inhibiteur :  $W_- = 1$  nA
- dus aux forçages intrinsèques :
  - Pour un pool excitateur :  $W_E = 1$ ,  $I_O = 0.382$  nA
  - Pour un pool inhibiteur :  $W_I = 0.7$ ,  $I_O = 0.382$  nA

Courants inter région :

$G = 0.69$ ,  $J_{nmda} = 0.15$  nA,  $C_{ij} \in [0, 1]$ .

### 6.2.4 Modèle de stimulation pendant une tâche de contrôle inhibiteur

Nous avons besoin de modéliser les variations d'activité dues à une tâche de contrôle inhibiteur. En nous basant sur la proposition de [Deco et al., 2014], nous choisissons d'appliquer des forçages au niveau des courants excitateurs  $xn_i$  de chaque région  $i$ . Le modèle des variations pendant la tâche est celui qui a été présenté au chapitre 5 section 5.3.1.1 et qui se traduit par un créneau appliqué en période de stimulation. L'amplitude du créneau est noté  $I_{stim,i}$ .

En définissons  $U$  telle que :

$$U(t) = \begin{cases} 1 & \text{en période de stimulation} \\ 0 & \text{hors période de stimulation} \end{cases} \quad (6.16)$$

Le courant excitateur  $xn_i$  s'écrit alors :

$$xn_i = W_+ J_{nmda} sn_i - J_{gaba,i} sg_i + W_E I_0 + I_{stim,i} U(t) + G J_{nmda} \sum_{j \neq i} C_{ij} sn_j \quad (6.17)$$

---

11. [Naskar et al., 2021] utilise une procédure dynamique pour fixer cette valeur de référence, mais qui s'avère inutile puisque l'inversion peut en fait être conduite analytiquement.



Le contrôle inhibiteur étant pensé comme un processus de régulation top-down (cf chapitre 2.3), nous faisons le choix de ne stimuler que les régions de contrôle. La stimulation est donc modélisée par une variation du forçage, d’une amplitude de  $I_{stim,i}$ , sur le pool excitateur des régions de contrôle.

### 6.2.5 Cas de deux régions : une région de contrôle et une région cible

Dans la thèse, nous nous intéressons particulièrement à un système de référence composé de deux régions : une région de contrôle et une région cible, représentées à la figure 6.5<sup>12</sup>. Il s’agit du cas princeps avant d’envisager toute généralisation à un système contenant deux sous-systèmes de régions (i.e. un ensemble de régions de contrôle et un ensemble de régions cibles). Il servira de cas d’étude quand nous envisagerons d’étendre les possibilités de connectivité inter-régions (voir chapitre 8).

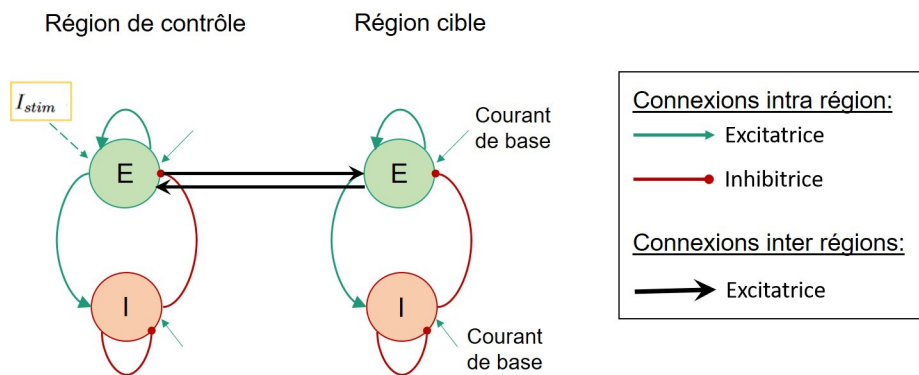


**Figure 6.5** – Illustration des courants ioniques pendant une stimulation pour une région de contrôle et une région cible. Au repos  $I_{stim} = 0$ .

12. Nous avons à cet effet choisi de présenter en priorité à la section 5.5.2 les études DCM qui portent sur deux régions, le DLPFC et l’hippocampe.

## 6.3 Modèle de la mesure : modèles hémodynamique et BOLD

Nous venons de présenter le modèle neural qui donne les variations des activités excitatrices et inhibitrices de régions cérébrales au cours du temps. Au chapitre 4, nous avons décrit la nature de la mesure BOLD qui permet d'étudier les mécanismes cérébraux du contrôle inhibiteur : nous ne mesurons pas directement les activités neurales excitatrices et inhibitrices. Il est donc nécessaire de travailler en plus avec un modèle de la mesure à partir de ces variations neurales. Au chapitre 5, nous avons présenté deux modèles possibles de la mesure BOLD (cf 5.3.2.1 et 5.3.2.2).

Comme [Wang et al., 2019] et [Naskar et al., 2021], nous travaillons avec un modèle biophysique explicite du couplage neurovasculaire (Ballon-Chambre à air étendu) et de la mesure BOLD. Il permet en effet de mieux prendre en compte les effets non linéaires constatés sur des données. Pour plus de détails sur ce modèle et sa structure, nous nous reporterons à la section 5.3.2.2. Ce modèle de la mesure BOLD est hiérarchique et découplé entre les régions. Les seuls couplages entre les régions ont donc lieu au niveau des courants excitateurs  $xn_i$  du modèle neural sous-jacent.

### 6.3.1 Modèle hémodynamique : extended Balloon-Windkessel Model

$sn_i$  est la fraction moyenne de canaux ouvert de la région  $i$ . Cette activité neurale excitatrice agit comme un forçage en entrée du modèle hémodynamique.

$z_i$  représente le signal de vasodilatation de la région  $i$ , adimensionnalisé par rapport à sa valeur au repos.  $f_i$  désigne le flux en entrée du compartiment veineux,  $v_i$  le volume du compartiment veineux et  $q_i$  la quantité de déoxyhémoglobine contenue dans le compartiment veineux de la région  $i$ , tous adimensionnalisés par rapport à leurs valeurs au repos.

$$\begin{cases} \dot{z}_i = sn_i - \kappa z_i - \gamma(f_i - 1) \\ \dot{f}_i = z_i \\ \dot{v}_i = \frac{1}{\tau}(f_i - v_i^{\frac{1}{\alpha}}) \\ \dot{q}_i = \frac{f_i}{\tau\rho}\left(1 - (1 - \rho)^{\frac{1}{f_i}}\right) - \frac{1}{\tau}q_i v_i^{\frac{1}{\alpha}-1} \end{cases} \quad (6.18)$$

Valeurs des paramètres ([Wang et al., 2019]) :

$$\kappa = 0.65 \text{ s}^{-1}, \gamma = 0.41 \text{ s}^{-1}, \tau = 0.98 \text{ s}, \rho = 0.34, \alpha = 0.32.$$

### 6.3.2 Modèle BOLD

Le signal BOLD de la région  $i$  est une fonction du volume  $v_i$  et de la quantité de déoxy-hémoglobine  $q_i$  :

$$y_i = V_0 \left( k_1(1 - q_i) + k_2(1 - q_i v_i^{-1}) + k_3(1 - v_i) \right) \quad (6.19)$$

$$\text{Avec les paramètres : } \begin{cases} k_1 = 4.3\theta_0\rho TE \\ k_2 = \epsilon r_0\rho TE \\ k_3 = 1 - \epsilon \end{cases} \quad (6.20)$$

Valeurs des paramètres ([Wang et al., 2019]) :

$V_0 = 0.02$ ,  $\theta_0 = 28.265B_0$  avec un champ magnétique  $B_0 = 3$  T,  $\rho = 0.34$  (le même que dans le modèle hémodynamique),  $TE = 0.030$  s (adapté à notre séquence d'IRMf, cf [Mary et al., 2020]),  $\epsilon = 0.47$ ,  $r_0 = 110$  Hz.

Dans le chapitre suivant, nous montrerons quelques simulations numériques permettant d'illustrer le comportement du modèle à deux régions, du neural jusqu'au BOLD (cf section 7.1.1.2.1).

## 6.4 Traduction de la question que pose la communauté étant donné ces choix de modélisation

A partir du modèle présenté dans ce chapitre, nous nous intéressons aux questions suivantes :

1. Peut-on obtenir une baisse du signal BOLD dans la région cible quand l'activité BOLD de la région de contrôle augmente ?
2. Peut-on identifier des paramètres qui permettent de modifier l'état de la balance neurale locale E/I ?
3. Comment les propriétés E/I neurales locales<sup>13</sup> influencent-elles les variations de signaux BOLD des régions impliquées dans une tâche de contrôle inhibiteur, en particulier la baisse d'activité des régions cibles due à l'augmentation d'activité des régions de contrôle ?
4. Comment inférer les paramètres neuraux qui représentent l'état des balances E/I locales de chaque région à partir de signaux BOLD recueillis pendant une tâche de contrôle inhibiteur ?

---

13. ou balances E/I locales

5. Nous prenons comme cas d'étude le contrôle inhibiteur des intrusions mnésiques (testé avec le paradigme TNT). Est-il possible de présenter une preuve de concept du point 4, avec la mise en oeuvre numérique de la procédure d'inversion sur les données de participant.

Pour préciser ces questions et pouvoir les traduire dans le modèle (en termes de variables et paramètres d'intérêt), nous en étudions les propriétés dynamiques dans le chapitre 7.



# Chapitre 7

## Problématique de thèse et organisation du travail

*Nous avons fait au chapitre 6 le choix d'un modèle direct (hiérarchique : neural, hémodynamique et BOLD). Nous n'en changeons plus et le travail consiste maintenant à analyser ses propriétés.*

*L'objectif de ce chapitre est de :*

1. *éclairer par l'analyse le passage entre un niveau de description neurale<sup>1</sup> et le signal BOLD. Nous préciserons :*
  - *la comparaison entre les temps d'installation du modèle neural et ceux du modèle hémodynamique/BOLD*
  - *comment le niveau d'activité BOLD  $y_i$  d'une région dépend de son activité neurale excitatrice  $s_n$  ?*
2. *éclairer par l'analyse l'effet de la stimulation des pools d'une région isolée pour faire émerger la question de l'architecture de connectivité*
3. *identifier des paramètres neuraux, représentant des propriétés biologiques, qui pourraient être différents entre sujets pour étudier la variabilité interindividuelle de l'état des balances E/I de régions cérébrales impliquées dans une tâche de contrôle inhibiteur.*

---

1. voir ou revoir encadré 2.3.2.3 pour la distinction entre "neural" et "neuronal"

## 7.1 Etude du passage entre les niveaux de description neurale et BOLD

### 7.1.1 Comparaison entre les temps d'installation du modèle neural et du modèle hémodynamique

#### 7.1.1.1 Temps caractéristiques du modèle hémodynamique

Nous observons que, pour une région donnée, les deux premières équations différentielles du modèle sont découplées des suivantes : nous pouvons donc les résoudre indépendamment, et la dynamique du système complet sera dominée par celle de ces deux équations.

• **hypothèse** : après quelques explorations numériques, et au regard des jeux de paramètres, nous faisons l'hypothèse que le système neural est installé quasi instantanément par rapport au système hémodynamique, c'est-à-dire que nous pouvons remplacer  $sn(t)$  par  $sn^*$  dans la première équation de [6.18]. Cette hypothèse sera totalement justifiée dans la suite par la comparaison des temps caractéristiques des deux modèles.

De plus, les deux premières équations du premier ordre en temps se résument en fait à une seule équation différentielle du 2nd ordre, linéaire, qui, en partant de l'hypothèse que le système neural est installé quasi instantanément par rapport au système hémodynamique, peut se ré-écrire comme :

$$\ddot{f} + \kappa\dot{f} + \gamma f = sn^* + \gamma \quad (7.1)$$

Nous reconnaissons une équation différentielle bien connue qui est celle d'un **oscillateur harmonique amorti**, dont la solution et les caractéristiques sont "classiques"<sup>2</sup>. Les racines du polynôme caractéristique peuvent s'écrire :

$$r_1, r_2 \simeq -0.325 \pm i 0.551$$

Nous en déduisons d'une part le temps caractéristique du modèle hémodynamique :

$$\tau = \text{abs}(Re(r_i))^{-1} \simeq 3.077s$$

et d'autre part, la valeur de la pseudo-période d'amortissement qui est de

$$T = \frac{2\pi}{0.551} \simeq 11.388 s$$

---

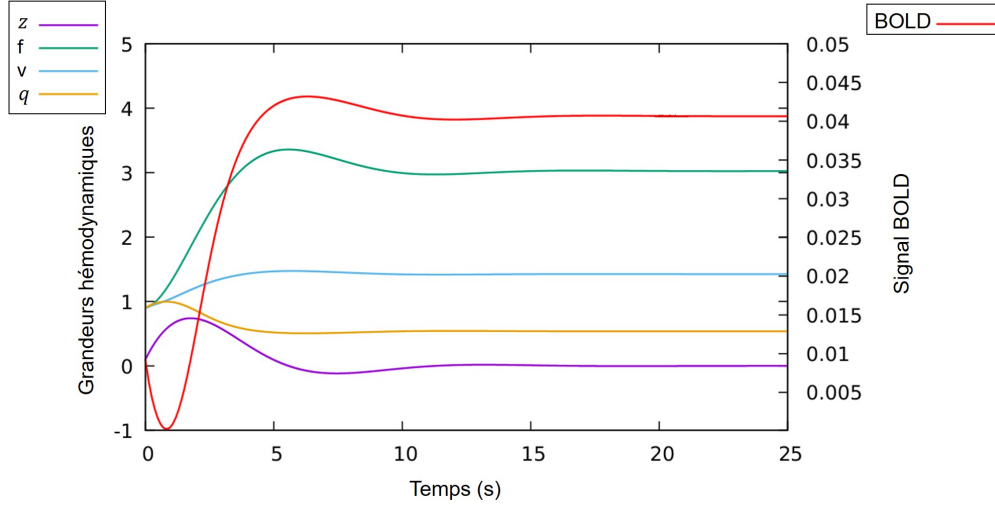
2. La solution de l'équation homogène (second membre nul) s'obtient en recherchant les modes propres, à partir de la solution du polynôme caractéristique obtenu en utilisant une solution d'essai de la forme  $f_0 e^{rt}$  :

$$\mathcal{P}(r) = r^2 + \kappa r + \gamma = 0$$

Le discriminant  $\Delta = \kappa^2 - 4\gamma$  va pouvoir être calculé avec les données des paramètres qui sont fournis par la communauté :  $\kappa = 0.65s^{-1}$ ;  $\gamma = 0.41s^{-1}$  et nous trouvons  $\Delta \simeq -1.2175$  **néga**tif.

Les deux premières équations différentielles (ou l'équivalente du 2nd ordre) engendrent donc des oscillations amorties.

Pour illustration, les évolutions respectives des variables hémodynamiques sont représentées ci-dessous :



### 7.1.1.2 Temps caractéristiques du modèle neural pour un système à deux régions cérébrales

Les équations différentielles décrivant l'évolution au cours du temps des activités synaptiques excitatrices  $sn_i(t)$  et inhibitrices  $sg_i(t)$  concernant plusieurs régions cérébrales indicées  $i$  ( $i = 1, 2$ ) ont été établies ci-dessus [6.12].

$$\begin{cases} \dot{sn}_i(t) = -\beta^E sn_i(t) + \alpha^E T_{glu}(1 - sn_i(t))rn_i(t) \\ \dot{sg}_i(t) = -\beta^I sg_i(t) + \alpha^I T_{gaba}(1 - sg_i(t))rg_i(t) \end{cases} \quad (7.2)$$

où apparaissent les dépendances à des variables secondaires (dites "de transfert")  $rn_i(t)$  et  $rg_i(t)$  fréquences des potentiels d'action, elles-mêmes dépendant de façon univoque des intensités électriques des potentiels  $xn_i(t)$  et  $xg_i(t)$ .

Une caractéristique importante est que ce système d'équations est fortement non-linéaire, notamment du fait des produits de variables tels que  $(1 - sn_i(t))rn_i(t)$ . Il en résulte que les solutions ne peuvent être décrites par des combinaisons linéaires d'exponentielles du type  $\sum_k a_k \exp(\lambda_k t)$  (le principe de superposition ne s'applique pas). Du fait de ces couplages, la dynamique de ce système échappe à tout raisonnement de type cause-effet sur les paramètres et forçages apparaissant dans les équations. Le jeu des rétroactions est complexe au fil du temps et les variables sont fortement couplées. Nous recourons alors à l'analyse linéaire au voisinage du ou des points fixes<sup>3</sup> s'ils existent, qui va donner un certain nombre d'indications sur la stabilité du système dynamique, mais aussi les sensibilités aux paramètres et les rétroactions, et pour ce qui nous concerne ici sur les temps caractéristiques de l'évolution.

Dans le cadre du contrôle inhibiteur, nous allons nous intéresser à une configuration "modèle" constituée de deux régions cérébrales, une région dite de contrôle (dénotée  $rc$  dans

3. aussi appelé états stationnaires si le point fixe est stable, pour lesquels la dynamique est stationnaire dans le temps s'ils sont pris comme condition initiale exacte.



les indices en version française, et  $C$  pour Control dans les indices en version anglaise) où sera appliquée l'activation du contrôle inhibiteur (e.g. une région du cortex préfrontal) et une région dite cible (dénotée  $c$  dans les indices en version française, et  $T$  pour Target dans les versions anglaises), dont l'activité doit baisser sous l'effet de l'activation de la région de contrôle (e.g. une région de l'hippocampe pour le cas de la suppression mnésique).

Ce système de référence à 2 régions voit son évolution (activités synaptiques, fréquences de tir et courants) déterminée par un système de 6 équations algébro-différentielles.

#### 7.1.1.2.1 Etude dynamique par simulations

Au premier chef, il est informatif d'analyser le comportement dynamique de ce modèle algébro-différentiel par voie numérique, et notamment sa réaction comportementale en situation stimulée.

Pour deux régions (une région de contrôle et une région cible), une évolution typique des activités synaptiques et les fréquences de potentiels d'action est représentée par les courbes de la figure 7.1.

Il faut, bien entendu, garder à l'esprit que les résultats de simulation sont fortement dépendants du jeu des paramètres neuro-physiologiques ou cinétiques du modèle. Nous avons utilisé ici les valeurs données au chapitre précédent. En ce qui concerne les valeurs du paramètre lié au neurotransmetteur de l'inhibition ( $J_{gaba,i}$ ), nous lui affectons la valeur qui vérifie la contrainte de fréquence  $rn = 3$  Hz dans les situations de repos (resting state) c'est-à-dire  $J_{gaba,rc} = J_{gaba,c} = 1.758$  nA.

Nous vérifions que les activités synaptiques des pools excitateurs  $sn_{rc}$  et  $sn_c$  des deux régions sont à l'état stationnaire de repos ( $\simeq 0.2$ ) juste avant l'activation de la stimulation à 20 s et nous observons qu'elles vont évoluer très rapidement vers un nouvel état stationnaire après application de la stimulation sur la région de contrôle à l'instant  $t = 20$  s ;  $sn_{rc}$  croît alors jusque vers 0.79 tandis que  $sn_c$  croît jusque vers 0.60<sup>4</sup> ; ensuite, lorsque le stimulus est éteint à l'instant  $t = 40$  s, les activités synaptiques retournent aux valeurs de repos. Dans les deux cas, ces évolutions des niveaux d'activité synaptiques vers l'état stationnaire se font dans un "temps d'installation"<sup>5</sup> de l'ordre de 1 s.

En ce qui concerne les signaux BOLD des deux régions, nous voyons qu'ils partent d'un même état non stimulé voisin de 0.016 (adim, échelle de droite). Lors de l'application du stimulus, le  $BOLD_{rc}$  croît jusqu'à 0.035 (doublement) alors que le  $BOLD_c$ , comme le  $sn_c$  qui le force, croît vers une valeur plus basse que le contrôle : 0.031. Les évolutions des BOLD se font selon des oscillations amorties, et atteignent leurs états stationnaires respectifs au bout d'environ 15 s après l'application de la stimulation, qui est donc le temps d'installation typique du signal BOLD.

---

4. La différence de comportement entre les deux régions provient de la brisure de symétrie introduite par le fait que seule la région  $rc$  est stimulée

5. par convention, nous définissons un "**temps d'installation**"  $t_{ins}$  (ou temps de relaxation) à 99.999% de l'état stationnaire, qui est d'environ 11.5 fois la valeur du temps caractéristique de l'évolution ( $11.5 \simeq \ln(10^5)$ ) résultant de l'étude de la fonction  $e^{-t/\tau}$ ).

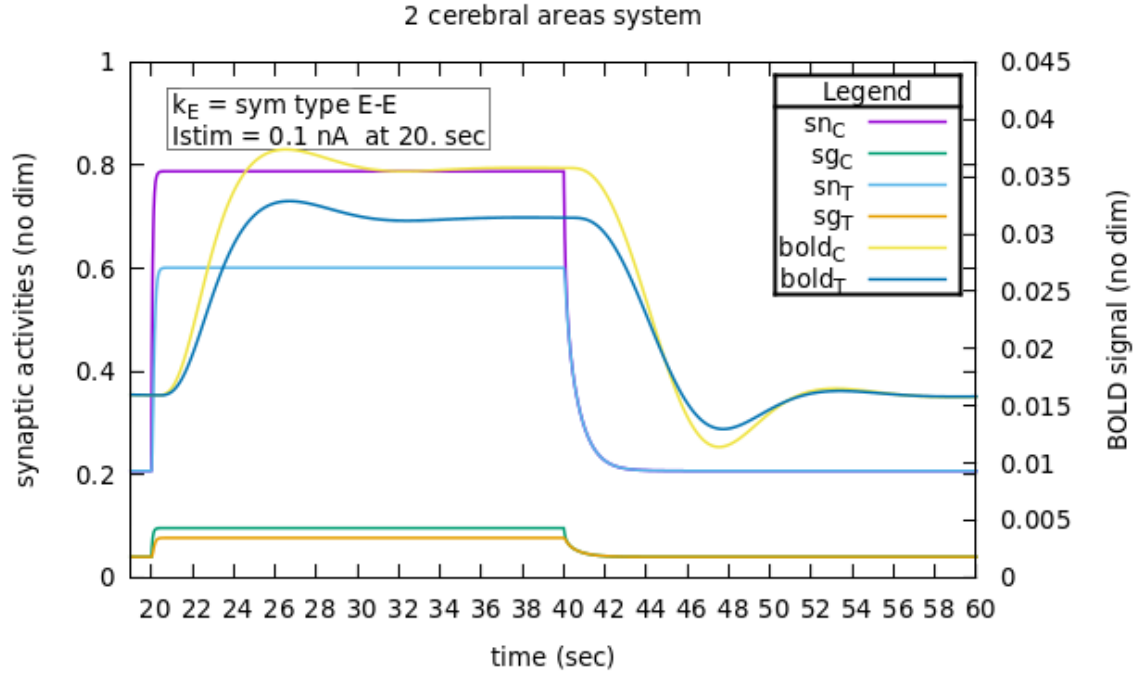


**Figure 7.1** – Evolution typique des activités synaptiques et des signaux BOLD dans le système à deux régions. Partant d'une condition initiale arbitraire, le système est simulé jusqu'à atteindre son état de repos, et un stimulus de 0.1 nA est appliqué entre 20 s et 40 s sur la région de contrôle. Les indices de régions sont donnés en version anglaise : C pour Control, T pour Target.

Par ailleurs, les fréquences de potentiels d'action  $rn_{rc}$  et  $rn_c$  sont bien contraintes à la valeur 3 Hz à l'état de repos, et augmentent respectivement à environ 25 et 10 Hz en état stimulé (non montré).

Nous notons que les temps d'installation des activités synaptiques sont très courts devant ceux du signal BOLD (dans un rapport de 1 à 15) : la dynamique du système neural est donc quasiment instantanée devant celle des mécanismes hémodynamiques engendrant le signal BOLD, justifiant rétrospectivement l'hypothèse du (7.1.1.1).

#### 7.1.1.2.2 Analyse linéaire au voisinage de l'état stationnaire

Pouvons-nous estimer a priori ces temps caractéristiques du système neural cérébral ? Et pouvons-nous repérer quels paramètres ou valeurs au point fixe déterminent ces temps, par voie analytique ?

L'approche classique de la **théorie des systèmes dynamiques** consiste à procéder à une analyse linéaire de stabilité au voisinage du point fixe.

S'il existe et est unique<sup>6</sup>, le point fixe atteint par le système est défini par la solution du

6. nous ne le montrerons pas ici, mais dans toutes les configurations et paramétrisations étudiées, ce fut le cas constaté

système d'équations non linéaires suivant :

$$\vec{f} = 0 \rightarrow \begin{cases} f n_{rc}(s n_{rc}, r n_{rc}) = 0 = -\beta^E s n_{rc}^* + \alpha^E T_{glu}(1 - s n_{rc}^*) r n_{rc}^* \\ f g_{rc}(s g_{rc}, r g_{rc}) = 0 = -\beta^I s g_{rc}^* + \alpha^I T_{gaba}(1 - s g_{rc}^*) r g_{rc}^* \\ f n_T(s n_c, r n_c) = 0 = -\beta^E s n_c^* + \alpha^E T_{glu}(1 - s n_c^*) r n_c^* \\ f g_T(s g_c, r g_c) = 0 = -\beta^I s g_c^* + \alpha^I T_{gaba}(1 - s g_c^*) r g_c^* \end{cases} \quad (7.3)$$

La linéarisation du système d'équations différentielles précédent [7.2] autour du point fixe  $\vec{s}^* = (s n^*, s g^*)$  et  $\vec{r}^* = (r n^*, r g^*)$  aboutit à un nouveau système d'équations différentielles, appelé **système linéarisé tangent**, et qui donne l'évolution d'une perturbation autour du point fixe  $(\delta s n, \delta s g)$  qui s'écrit comme :

$$\begin{pmatrix} \delta \dot{s} n \\ \delta \dot{s} g \end{pmatrix}_i = \underbrace{f_i(s n^*, r n^*)}_{=0 \text{ par définition du point fixe}} + \left. \frac{\partial f}{\partial s} \right|_i^* \begin{pmatrix} \delta s n \\ \delta s g \end{pmatrix}_{i(t)} + \left. \frac{\partial f}{\partial r} \right|_i^* \begin{pmatrix} \delta r n \\ \delta r g \end{pmatrix}_{i(t)} \quad (7.4)$$

Nous omettrons par la suite de préciser que toutes ces expressions sont prises au voisinage du point fixe noté \*, et nous avons :

$$\begin{pmatrix} \delta \dot{s} n_{rc} \\ \delta \dot{s} g_{rc} \\ \delta \dot{s} n_c \\ \delta \dot{s} g_T \end{pmatrix}_{(t)} = \underbrace{\begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{A}_{13} & \mathcal{A}_{14} \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{A}_{23} & \mathcal{A}_{24} \\ \mathcal{A}_{31} & \mathcal{A}_{32} & \mathcal{A}_{33} & \mathcal{A}_{34} \\ \mathcal{A}_{41} & \mathcal{A}_{42} & \mathcal{A}_{43} & \mathcal{A}_{44} \end{pmatrix}}_{\text{matrice } \mathcal{A}^*} \begin{pmatrix} \delta s n_{rc} \\ \delta s g_{rc} \\ \delta s n_c \\ \delta s g_c \end{pmatrix}_{(t)} \quad (7.5)$$

Ce système étant maintenant linéaire par construction, les évolutions au voisinage du point fixe peuvent s'exprimer sous forme d'une combinaison de solutions exponentielles :  $\sum_k a_k e^{\lambda_k t}$ , où les  $\lambda_k$  sont les valeurs propres de la matrice  $\mathcal{A}$ .

Comme nous l'avons déjà vu ci-dessus, si une de ces valeurs propres a une partie réelle positive, l'évolution tendra vers l'infini et le point fixe sera diagnostiqué comme instable.

Dans le cas où toutes les parties réelles sont négatives, le système dynamique est stable et nous obtenons le spectre des temps caractéristiques, où les  $\tau_k = \frac{1}{|\Re(\lambda_k)|}$ .

### Résultats numériques de la configuration de référence

Par intégration numérique, après application d'un stimulus de référence (0.1 nA), nous obtenons la matrice d'avance  $\mathcal{A}^*$  au point fixe et nous obtenons également des valeurs propres  $\lambda_k$ , toutes réelles (et à partie réelle négative, ce qui assure la stabilité du point fixe).

Nous en déduisons les temps caractéristiques  $\tau_k$  (en secondes) :

$$\begin{cases} \tau_1 \simeq 0.075 \\ \tau_2 \simeq 0.031 \\ \tau_3 \simeq 1.73 \cdot 10^{-3} \\ \tau_4 \simeq 1.54 \cdot 10^{-3} \end{cases} \quad (7.6)$$

La valeur propre possédant la partie réelle la plus petite en valeur absolue<sup>7</sup> donne également le temps caractéristique  $\tau_1$  du système dynamique<sup>8</sup>, qui est donc ici de l'ordre du dixième de seconde, et donc bien de l'ordre du dixième du temps d'installation observé sur la figure 7.1.

### Conclusion en termes de temps caractéristiques

La détermination analytique des temps caractéristiques du modèle neural à deux régions confirme les résultats obtenus par la simulation illustrative.

Dans le développement du questionnement de thèse, nous verrons plus loin que d'autres modes de connectivité que celui inclus dans le modèle de départ seront développés. Nous avons vérifié les différents temps caractéristiques du système à deux régions en fonction des différents modes de connectivité : les temps caractéristiques ne changent que de quelques pourcents d'un mode à l'autre, gardant cette conclusion inchangée.

De même, nous avons calculé les temps caractéristiques pour des systèmes composés de nombreuses régions cérébrales (11 régions de contrôle et 4 régions cibles) : là encore, les temps caractéristiques restent dans le même ordre de grandeur, c'est-à-dire très petits devant le temps caractéristiques du signal BOLD.

#### 7.1.1.3 Que devient le modèle ?

Le temps d'installation du modèle neural est court devant celui de la réponse BOLD. En conséquence, le régime transitoire du modèle neural entre deux points fixes au repos et stimulé n'a que très peu d'effet sur la variation BOLD associée.

Nous pouvons donc considérer le modèle neural au régime stationnaire en considérant qu'il "saute" instantanément d'un point fixe de repos à un point fixe stimulé. En revanche, nous ne pouvons que continuer à travailler en dynamique pour les modèles hémodynamique et BOLD. Le modèle complet que nous retenons pour la suite s'écrit dès lors comme suit :

Modèle neural :

$$\begin{cases} 0 = -\beta^E sn_i^* + \alpha^E T_{glu} r n_i^* (1 - sn_i^*) \\ 0 = -\beta^I sg_i^* + \alpha^I T_{gaba} r g_i^* (1 - sg_i^*) \end{cases} \quad (7.7)$$

$$\begin{cases} r n_i^* = \frac{a_E x n_i^* - b_E}{1 - e^{-d_E(a_E x n_i^* - b_E)}} \\ r g_i^* = \frac{a_I x g_i^* - b_I}{1 - e^{-d_I(a_I x g_i^* - b_I)}} \end{cases} \quad (7.8)$$

$$\begin{cases} x n_i^* = W_+ J_{nmda} s n_i^* - J_{gaba,i} s g_i^* + W_E I_0 + I_{stim_i} U(t) + G J_{nmda} \sum_{j \neq i} C_{ij} s n_j^* \\ x g_i^* = J_{nmda} s n_i^* - W_- s g_i^* + W_I I_0 \end{cases} \quad (7.9)$$

---

7. La valeur propre possédant la partie réelle la plus petite en valeur absolue est appelée **exposant de Liapunov**

8. i.e. le temps caractéristique le plus grand, qui est celui qui décrit globalement l'évolution du système, les autres composantes exponentielles étant atténuées bien avant ce temps.

Avec la fonction créneau  $U$  définie telle que :

$$U(t) = \begin{cases} 1 & \text{en période de stimulation si } i \text{ est une région de contrôle} \\ 0 & \text{hors période de stimulation} \end{cases} \quad (7.10)$$

Modèle hémodynamique

$$\begin{cases} \dot{z}_i(t) = sn_i^* - \kappa z_i(t) - \gamma(f_i(t) - 1) \\ \dot{f}_i(t) = z_i(t) \\ \dot{v}_i(t) = \frac{1}{\tau} \left( f_i(t) - v_i(t)^{\frac{1}{\alpha}} \right) \\ \dot{q}_i(t) = \frac{f_i(t)}{\tau \rho} \left( 1 - (1 - \rho)^{\frac{1}{f_i(t)}} \right) - \frac{1}{\tau} q_i(t) v_i(t)^{\frac{1}{\alpha} - 1} \end{cases} \quad (7.11)$$

Modèle BOLD

$$y_i(t) = V_0 \left( k_1(1 - q_i(t)) + k_2(1 - q_i(t)v_i(t)^{-1}) + k_3(1 - v_i(t)) \right) \quad (7.12)$$

### 7.1.2 Relation entre le signal BOLD $y_i^*$ émis par une région cérébrale et son activité synaptique à l'état stationnaire $sn_i^*$

Dans la section 6.3.1, le modèle hémodynamique d'une région cérébrale a été détaillé, ainsi que son lien avec le signal BOLD observé.

En absence de stimulus externe agissant sur l'activité neurale, il existe une activité synaptique dite "basale", que nous dénoterons  $sn_B$ , et pour une région isolée, il existe donc un état stationnaire non trivial du modèle hémodynamique (et par voie de conséquence du signal BOLD) :

On peut exprimer analytiquement l'état stationnaire du signal BOLD pour une valeur non nulle de l'activité du pool excitateur  $sn_i^*$ , en présence ou non de stimulus.

Des deux premières équations différentielles, on tire :

$$f_i^* = sn_i^* / \gamma + 1 \quad (7.13)$$

qui est une fonction affine de  $sn_i^*$ .

De la 3ème équation, on tire la fonction du volume :

$$v_i^* = f_i^{*\alpha} \quad (7.14)$$

Et de la 4ème équation, on tire, après quelques arrangements :

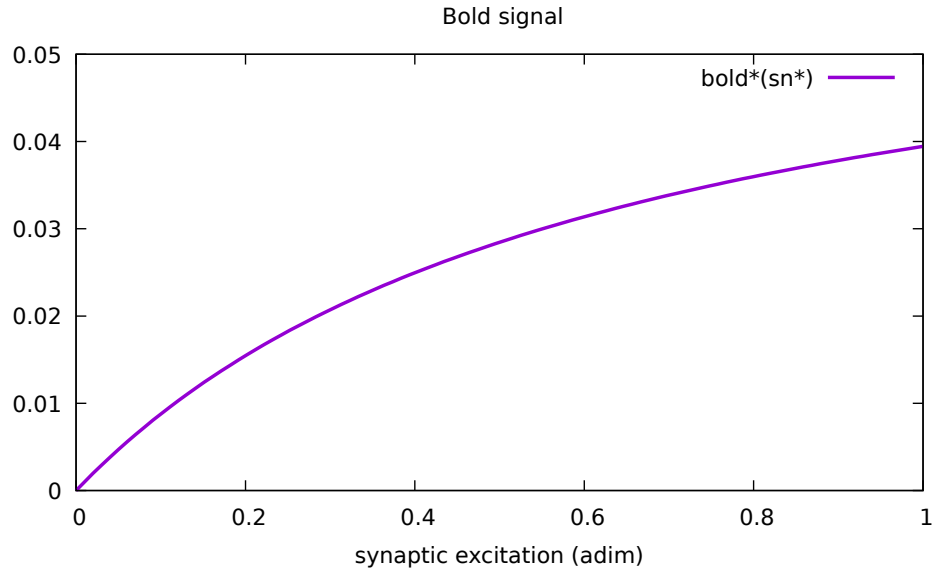
$$q_i^* = v_i^* \left\{ \left( 1 - (1 - \rho)^{\frac{1}{f_i^*}} \right) \right\} / \rho \quad (7.15)$$

En reportant successivement dans les équations du modèle, il ne reste plus, pour obtenir le signal BOLD  $y_i^*$  à l'état stationnaire qu'à remplacer ces fonctions de  $sn_i^*$  dans l'expression précédente.

On obtient ainsi la relation entre le signal BOLD stationnaire d'une région donnée et le "forçage" donné par le  $sn^*$  de la même région :

$$y_i^* = V_0 \{k_1(1 - q_i^*) + k_2(1 - q_i^* v_i^{*-1}) + k_3(1 - v_i^*)\} \quad (7.16)$$

Dans le domaine de variation de  $sn_i^*$ ,  $[0, 1]$ , on peut alors représenter la dépendance fonctionnelle du signal BOLD stationnaire en fonction de  $sn_i^*$  par la courbe ci-dessous :



Conclusion : il existe une relation bijective croissante entre l'activité synaptique  $sn_i^*$  et le signal BOLD  $y_i^*$ . Un point central pour la procédure d'inversion du BOLD vers le neural est de constater que le domaine de définition de la variable neurale entre 0 et 1 ne peut produire qu'un signal BOLD contraint entre 0 et 0.04 : cela justifiera une normalisation du signal BOLD à partir des données brutes obtenues par IRMf.

Sur le plan dynamique, pour un jeu de paramètre fixé, le signal BOLD  $y_i(t)$  tend au cours du temps vers  $y_i^*$ . Hors période de stimulation, le signal BOLD d'une région se rapproche de son état de base  $y_{B_i}^*$ , d'autant plus que la période sans stimulation est longue. De même, en période de stimulation, le signal BOLD de cette région s'approche de son état stimulé  $y_{stim_i}^*$ , d'autant plus que la période de stimulation est longue.

Le signe des variations du signal BOLD entre les périodes de stimulation et de non stimulation est identique à celui de la différence entre  $y_{stim_i}^*$  et  $y_{B_i}^*$ , même si ces périodes sont trop courtes pour que les points fixes soient atteints.

D'après la bijection, pour avoir une augmentation de l'activité BOLD de la région de contrôle en période de stimulation,  $\Delta y_{rc}^* = y_{stim_{rc}}^* - y_{B_{rc}}^* > 0$ , il faut donc que  $\Delta sn_{rc}^* > 0$ . Pour augmenter l'activité BOLD d'une région de contrôle en période de stimulation, il est nécessaire

d'augmenter son activité synaptique  $sn_{rc}^*$  au régime stationnaire. Inversement, pour diminuer l'activité BOLD d'une région cible, il est nécessaire de diminuer son activité synaptique  $sn_c^*$  au régime stationnaire :  $\Delta sn_c^* < 0$ .

Ce résultat est très important pour l'analyse qui va suivre, puisque qu'il met en évidence une relation bijective entre l'activité synaptique du système neural associée à une région et le signal BOLD qui sera observé pour cette même région au régime stationnaire. C'est le support principal justifiant toute l'analyse au voisinage de l'état stationnaire des propriétés du modèle neural (sensibilités aux forçages et aux paramètres, gains de rétroaction, etc.) des systèmes de régions.

### 7.1.3 Conclusion pour la précision de la problématique

#### 7.1.3.1 Pour le modèle direct

La comparaison des temps d'installation des modèles neural et hémodynamique, ainsi que la bijection entre  $sn_i^*$  et  $y_i^*$  permettent de conclure que :

- La variation d'activité BOLD entre les périodes de stimulation et de repos (positive dans la région de contrôle et négative dans la cible) peut se traduire en terme de variation de l'activité neurale  $sn^*$  au régime stationnaire ( $\Delta sn^*$ ), de même signe.

Dans la suite de la thèse, nous allons donc nous focaliser uniquement sur les propriétés du modèle neural au régime stationnaire, en particulier sur des  $\Delta sn^*$ , où nous chercherons les situations caractérisant le contrôle inhibiteur au niveau neural, quand :

- Pour une région de contrôle :  $\Delta sn_{rc}^* = sn_{stim_{rc}}^* - sn_{B_{rc}}^* > 0$
- Pour une région cible :  $\Delta sn_c^* = sn_{stim_c}^* - sn_{B_c}^* < 0$

#### 7.1.3.2 Pour l'inversion

La comparaison des temps d'installation des modèles neural et hémodynamique nous permet d'écrire le modèle neural au régime stationnaire. Nous allons donc pouvoir construire une procédure d'inversion en deux étapes découplées en traitant d'abord l'inversion des signaux BOLD temporels  $y_i(t)$  vers les  $sn_i^*$  (en se servant uniquement du modèle hémodynamique (7.11) et BOLD (7.12)), puis l'inversion des  $sn_i^*$  aux paramètres neuraux d'intérêt (avec le modèle neural au régime stationnaire).

Le calcul des temps d'installation du modèle hémodynamique, qui détermine celui du BOLD<sup>9</sup>, nous impose en effet la nécessité de travailler avec la dynamique dans la première étape de l'inversion.

Si les périodes avec et sans stimulation étaient suffisamment longues pour que le signal BOLD atteigne un point fixe entre le départ et l'arrêt des stimulations, nous aurions pu utiliser directement la bijection entre  $y_i^*$  et  $sn_i^*$  dans la première étape de l'inversion. Cependant, dans les protocoles classiquement utilisés dans les tâches d'IRMf de type TNT<sup>10</sup>, leur durées sont

9. car le signal BOLD est une fonction non linéaire des variables d'état hémodynamiques

10. bien entendu construits et justifiés par d'autres rationalités que notre tentative d'inversion, telles que la nécessité de prendre en compte les processus attentionnels, la patience des patients, le nombre de données à recueillir au regard du temps passé dans l'IRM...

courtes (de l'ordre de 3 secondes, cf [Mary et al., 2020]) par rapport à la durée d'installation calculée ci-dessus (de l'ordre de 30 secondes). Sous nos hypothèses de modélisation, le signal BOLD n'a donc pas le temps d'atteindre un point fixe et nous ne pouvons pas nous servir simplement de la bijection pour inverser les  $sn_i^*$ . Nous avons besoin de prendre en compte les évolutions temporelles du signal BOLD et pour cela nous devons travailler avec le modèle direct dynamique, ce que permettra de faire la toolbox VBA.

## 7.2 Etude d'une région isolée : stimulation d'un des deux pools

Afin de donner un premier éclairage sur le comportement qu'on peut attendre du modèle, nous nous intéressons à l'effet sur  $sn^*$  d'une augmentation du forçage soit sur le pool exciteur soit sur le pool inhibiteur d'une région isolée. Nous allons d'abord étudier l'effet d'une perturbation infinitésimale.

### 7.2.1 Formulation mathématique du modèle au régime stationnaire

Pour une région, le modèle dynamique s'écrit :

$$\begin{cases} \dot{sn}(t) = -\beta^E sn(t) + \alpha^E T_{glu}(1 - sn(t))rn(t) \equiv fn(sn(t), rn(t)) \\ \dot{sg}(t) = -\beta^I sg(t) + \alpha^I T_{gaba}(1 - sg(t))rg(t) \equiv fg(sg(t), rg(t)) \end{cases} \quad (7.17)$$

$$\begin{cases} rn(t) = \frac{a_E xn(t) - b_E}{1 - e^{-d_E(a_E xn(t) - b_E)}} \equiv hn(xn(t)) \\ rg(t) = \frac{a_I xg(t) - b_I}{1 - e^{-d_I(a_I xg(t) - b_I)}} \equiv hg(xg(t)) \end{cases} \quad (7.18)$$

$$\begin{cases} xn(t) = W_+ J_{nmda} sn(t) - J_{gaba} sg(t) + x_E \equiv wn(sn(t), sg(t), J_{gaba}, x_E) \\ xg(t) = J_{nmda} sn(t) - W_- sg(t) + x_I \equiv wg(sn(t), sg(t)) \end{cases} \quad (7.19)$$

avec  $x_E$  et  $x_I$  les forçages sur les pools excitateurs et inhibiteurs.

Au régime stationnaire, le modèle mathématique d'intérêt pour une région peut s'écrire :

$$\begin{cases} 0 = fn(sn^*, rn^*) \\ 0 = fg(sg^*, rg^*) \end{cases} \quad (7.20)$$

$$\begin{cases} rn^* = hn(xn^*) \\ rg^* = hg(xg^*) \end{cases} \quad (7.21)$$



$$\begin{cases} xn^* = wn(sn^*, sg^*, J_{gaba}, x_E) \\ xg^* = wg(sn^*, sg^*, x_I) \end{cases} \quad (7.22)$$

Nous définissons :

$$\vec{s} = \begin{pmatrix} sn^* \\ sg^* \end{pmatrix}, \quad \vec{r} = \begin{pmatrix} rn^* \\ rg^* \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} xn^* \\ xg^* \end{pmatrix} \quad (7.23)$$

Au point fixe, nous pouvons écrire :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{s}, x_E, x_I) \end{cases} \quad (7.24)$$

### 7.2.2 Etude de sensibilité par rapport au forçage sur le pool exciteur ou inhibiteur

Nous considérons une perturbation soit de  $x_E$ , soit de  $x_I$ , notée  $x_\square$  pour le moment. Le nouveau modèle perturbé s'écrit :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{s}, x_\square + \delta x_\square) \end{cases} \quad (7.25)$$

La linéarisation autour du point fixe fournit le modèle en perturbation :

$$\begin{cases} \frac{\partial \vec{f}}{\partial s} \delta \vec{s} + \frac{\partial \vec{f}}{\partial r} \delta \vec{r} = \vec{0} \\ \delta \vec{r} = \frac{\partial \vec{h}}{\partial x} \delta \vec{x} \\ \delta \vec{x} = \frac{\partial \vec{w}}{\partial s} \delta \vec{s} + \frac{\partial \vec{w}}{\partial x_\square} \delta x_\square \end{cases} \quad (7.26)$$

En injectant les deux dernières expressions dans la première, nous obtenons :

$$\frac{\partial \vec{f}}{\partial s} \delta \vec{s} + \frac{\partial \vec{f}}{\partial r} \frac{\partial \vec{h}}{\partial x} \left( \frac{\partial \vec{w}}{\partial s} \delta \vec{s} + \frac{\partial \vec{w}}{\partial x_\square} \delta x_\square \right) = \vec{0} \quad (7.27)$$

que nous réarrangeons en :

$$\left(\frac{\overline{\partial f}}{\partial s} + \frac{\overline{\partial f}}{\partial r} \frac{\partial h}{\partial x} \frac{\partial w}{\partial s}\right) \vec{\delta s} = - \frac{\overline{\partial f}}{\partial r} \frac{\partial h}{\partial x} \frac{\partial w}{\partial x_{\square}} \delta x_{\square} \quad (7.28)$$

Nous posons :

$$\overline{\overline{S}} = - \frac{\overline{\partial f}^{-1}}{\partial s} \frac{\overline{\partial f}}{\partial r} \frac{\partial h}{\partial x} \quad (7.29)$$

La variation infinitésimale  $\vec{\delta s}$  s'écrit alors :

$$\vec{\delta s} = \left(\overline{\mathbb{I}} - \overline{\overline{S}} \frac{\overline{\partial w}}{\partial s}\right)^{-1} \overline{\overline{S}} \frac{\partial w}{\partial x_{\square}} \delta x_{\square} \quad (7.30)$$

Or, dans l'article du chapitre 9, nous avons calculé

$$\begin{cases} \left(\overline{\mathbb{I}} - \overline{\overline{S}} \frac{\overline{\partial w}}{\partial s}\right)^{-1} \overline{\overline{S}} \frac{\partial w}{\partial x_E} = \frac{\vec{\delta s}}{\delta x_E} \equiv \begin{bmatrix} \mathcal{A}_{sn,x_E} \\ \mathcal{A}_{sg,x_E} \end{bmatrix} \\ \left(\overline{\mathbb{I}} - \overline{\overline{S}} \frac{\overline{\partial w}}{\partial s}\right)^{-1} \overline{\overline{S}} \frac{\partial w}{\partial x_I} = \frac{\vec{\delta s}}{\delta x_I} \equiv \begin{bmatrix} \mathcal{A}_{sn,x_I} \\ \mathcal{A}_{sg,x_I} \end{bmatrix} \end{cases} \quad (7.31)$$

Si nous perturbons le forçage sur le pool exciteur  $x_E$  et que nous regardons uniquement la variable d'état  $sn^*$ , nous avons :

$$\boxed{\delta sn_{/x_E} = \frac{\delta sn}{\delta x_E} \delta x_E \equiv \mathcal{A}_{sn,x_E} \delta x_E} \quad (7.32)$$

Nous montrons dans l'article du chapitre 9 que  $\mathcal{A}_{sn,x_E} > 0$  pour tout jeu de paramètres. Une augmentation infinitésimale de  $x_E$  produit donc une augmentation infinitésimale de  $sn^*$ . Les figures 7(A) et 7(B) de l'article donnent  $sn^*$  et la sensibilité de  $sn^*$  à  $x_E$ , en fonction de  $x_E$ .

De la même manière, si nous perturbons le forçage sur le pool inhibiteur  $x_I$ , nous obtenons :

$$\boxed{\delta sn_{/x_I} = \frac{\delta sn}{\delta x_I} \delta x_I \equiv \mathcal{A}_{sn,x_I} \delta x_I} \quad (7.33)$$

Nous montrons dans l'article du chapitre 9 que  $\mathcal{A}_{sn,x_I} < 0$  pour tout jeu de paramètres. Une augmentation infinitésimale de  $x_I$  produit donc une diminution infinitésimale de  $sn^*$ . Les figures 7(D) et 7(E) de l'article donnent  $sn^*$  et la sensibilité de  $sn^*$  à  $x_I$ , en fonction de  $x_I$ .

### 7.2.3 Conclusion pour la précision de la problématique

Sous l'hypothèse des modèles hémodynamique et BOLD, pour abaisser l'activité BOLD de la région cible, nous avons vu qu'il était nécessaire de faire diminuer son activité neurale au régime stationnaire  $sn^*$ . Pour cela, deux options sont possibles :

- augmenter le forçage  $x_I$  sur le pool inhibiteur
- abaisser le forçage  $x_E$  sur le pool excitateur

Or, nous avons vu que les régions communiquent entre elles en projetant des fibres de nature excitatrice depuis leur pool excitateur. Cela se traduit mathématiquement par des coefficients de couplage entre régions qui sont toujours positifs :  $GJ_{nmda}C_{ij} > 0$ .

Dans un système composé de deux régions couplées, la région de contrôle participe uniquement au courant  $xn_c^*$  du pool excitateur de la cible. Cette contribution est égale à  $GJ_{nmda}sn_{rc}^*$  (avec  $C_{c,rc} = 1$ ). En période de stimulation, toute rétroactions prises en compte, l'activité  $sn_{rc}^*$  de la région de contrôle augmente de  $\Delta sn_{rc}^*$ , ce qui fait augmenter  $xn_c^*$  de  $GJ_{nmda}\Delta sn_{rc}^*$ . Pour prédire la variation d'activité de la cible, nous pouvons considérer la région cible comme isolée et évaluer le forçage en prenant en compte le forçage intrinsèque et la partie venant de la région de contrôle (la formalisation complète se trouve dans l'annexe de l'article au chapitre 9). Augmenter l'activité de la région de contrôle pendant une tâche de contrôle inhibiteur ne peut conduire qu'à une augmentation de  $x_E$  de  $GJ_{nmda}\Delta sn_{rc}^* > 0$ , c'est-à-dire à augmenter l'excitation du pool excitateur de la région cible. L'activité de la cible ne peut donc pas diminuer.

Au contraire, si la région de contrôle participait uniquement au courant  $xg_c^*$  du pool inhibiteur de la cible, augmenter l'activité de la région de contrôle peut se traduire par une augmentation de l'excitation du pool inhibiteur de la région cible. Nous pouvons alors prédire une diminution de l'activité de la région cible.

Pour nous permettre d'abaisser l'activité neurale excitatrice au régime stationnaire dans la région cible (i.e.  $\Delta sn_c^* = sn_{stim_c}^* - sn_{B_c}^* < 0$ ), quand l'activité neurale excitatrice au régime stationnaire dans la région de contrôle augmente (i.e.  $\Delta sn_{rc}^* = sn_{stim_{rc}}^* - sn_{B_{rc}}^* > 0$ ), nous devons nous intéresser à la manière de connecter les régions entre elles. Dans le modèle actuel, la région de contrôle projette des fibres excitatrices sur le pool excitateur de la région cible, et pas sur son pool inhibiteur.

## 7.3 Identification de paramètres permettant de modifier l'état de la balance E/I locale

### 7.3.1 Notion de balance locale, paramètres locaux E et I

Les paramètres d'une région isolée peuvent être classés en deux catégories en fonction de l'effet excitateur ou inhibiteur de leur perturbation sur l'activité  $sn^*$  de la région d'intérêt. La perturbation d'un paramètre "excitateur" a pour effet d'augmenter l'activité  $sn^*$ , alors que celle d'un paramètre "inhibiteur" a pour effet de la faire diminuer.

La balance de Roberval est une métaphore qui permet de se représenter des effets compensatoires sur les niveaux des plateaux lorsqu'on en modifie les poids. Supposons que l'ensemble

des paramètres "excitateurs" soit placé sur un plateau et l'ensemble des paramètres "inhibiteurs" sur l'autre. Le niveau des plateaux donne un niveau d'activité  $sn^*$  de la région. C'est la situation de référence.

Si on augmente le poids d'un paramètre exciteur, la balance penche du côté du plateau "excitateur" (situation 1) et l'activité  $sn^*$  de la région augmente par rapport à la situation de référence. On rajoute maintenant un poids "inhibiteur" sur l'autre plateau. Cet ajout a pour effet de réduire, annuler voire inverser le déplacement des plateaux de la situation 1. L'augmentation supplémentaire du paramètre inhibiteur conduit alors à réduire l'augmentation d'activité  $sn^*$  de la situation 1 (variation qui reste toujours positive par rapport à la situation de référence), voire à diminuer  $sn^*$  par rapport à la situation de référence. La métaphore de la balance de Roberval permet de se représenter des effets compensatoires sur la perturbation d'activité  $\delta sn^*$  si deux paramètres exciteur et inhibiteur sont perturbés en même temps.

Dans le modèle à une région isolée, nous voyons dans le paragraphe qui suit qu'une perturbation du paramètre  $J_{gaba}$  de  $\delta J_{gaba}$  a pour effet de produire un  $\delta sn^* < 0$ . Inversement, une perturbation du forçage  $x_E$  de  $\delta x_E$  conduit à une perturbation  $\delta sn^* > 0$ . La perturbation simultanée des deux paramètres donne lieu à des effets de compensations.

## 7.3.2 Approche analytique

### 7.3.2.1 Etude de sensibilité par rapport au forçage exciteur sur le pool exciteur et par rapport à $J_{gaba}$

Au régime stationnaire, le modèle mathématique d'intérêt pour une région peut s'écrire :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{s}, x_E, J_{gaba}) \end{cases} \quad (7.34)$$

Si nous considérons des perturbations de  $x_E$  et  $J_{gaba}$ , le nouveau modèle perturbé s'écrit :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{s}, x_E + \delta x_E, J_{gaba} + \delta J_{gaba}) \end{cases} \quad (7.35)$$

La linéarisation autour du point fixe produit :

$$\begin{cases} \frac{\partial \vec{f}}{\partial s} \vec{\delta s} + \frac{\partial \vec{f}}{\partial r} \vec{\delta r} = \vec{0} \\ \vec{\delta r} = \frac{\partial \vec{h}}{\partial x} \vec{\delta x} \\ \vec{\delta x} = \frac{\partial \vec{w}}{\partial s} \vec{\delta s} + \frac{\partial \vec{w}}{\partial x_E} \delta x_E + \frac{\partial \vec{w}}{\partial J_{gaba}} \delta J_{gaba} \end{cases} \quad (7.36)$$

En injectant les deux dernières dans la première, on obtient :

$$\frac{\overline{\partial f}}{\partial s} \vec{\delta s} + \frac{\overline{\partial f} \overline{\partial h}}{\partial r \partial x} \left( \frac{\overline{\partial w}}{\partial s} \vec{\delta s} + \frac{\partial \vec{w}}{\partial x_E} \delta x_E + \frac{\partial \vec{w}}{\partial J_{gaba}} \delta J_{gaba} \right) = \vec{0} \quad (7.37)$$

que nous réarrangeons en :

$$\left( \frac{\overline{\partial f}}{\partial s} + \frac{\overline{\partial f} \overline{\partial h} \overline{\partial w}}{\partial r \partial x \partial s} \right) \vec{\delta s} = - \frac{\overline{\partial f} \overline{\partial h}}{\partial r \partial x} \frac{\partial \vec{w}}{\partial x_E} \delta x_E - \frac{\overline{\partial f} \overline{\partial h}}{\partial r \partial x} \frac{\partial \vec{w}}{\partial J_{gaba}} \delta J_{gaba} \quad (7.38)$$

Nous posons

$$\overline{S} = - \frac{\overline{\partial f}^{-1} \overline{\partial f} \overline{\partial h}}{\partial s \partial r \partial x} \quad (7.39)$$

La variation infinitésimale  $\vec{\delta s}$  s'écrit alors :

$$\vec{\delta s} = \left( \overline{\mathbb{I}} - \overline{S} \frac{\overline{\partial w}}{\partial s} \right)^{-1} \overline{S} \frac{\partial \vec{w}}{\partial x_E} \delta x_E + \left( \overline{\mathbb{I}} - \overline{S} \frac{\overline{\partial w}}{\partial s} \right)^{-1} \overline{S} \frac{\partial \vec{w}}{\partial J_{gaba}} \delta J_{gaba} \quad (7.40)$$

Or

$$\frac{\partial \vec{w}}{\partial J_{gaba}} = \begin{bmatrix} -sg \\ 0 \end{bmatrix} = -sg \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -sg \frac{\partial \vec{w}}{\partial x_E} \quad (7.41)$$

Nous en déduisons :

$$\vec{\delta s} = \left( \overline{\mathbb{I}} - \overline{S} \frac{\overline{\partial w}}{\partial s} \right)^{-1} \overline{S} \frac{\partial \vec{w}}{\partial x_E} (\delta x_E - sg \delta J_{gaba}) \quad (7.42)$$

Dans l'article du chapitre 9, nous avons calculé :

$$\left( \overline{\mathbb{I}} - \overline{S} \frac{\overline{\partial w}}{\partial s} \right)^{-1} \overline{S} \frac{\partial \vec{w}}{\partial x_E} = \frac{\vec{\delta s}}{\delta x_E} \equiv \begin{bmatrix} \mathcal{A}_{sn, x_E} \\ \mathcal{A}_{sg, x_E} \end{bmatrix} \quad (7.43)$$

Si nous regardons en particulier la variable d'état  $sn^*$ , nous obtenons :

$$\boxed{\delta sn = \frac{\delta sn}{\delta x_E} (\delta x_E - sg \delta J_{gaba}) \equiv \mathcal{A}_{sn, x_E} (\delta x_E - sg \delta J_{gaba})} \quad (7.44)$$

### Etude de signe :

- Perturbation de  $x_E$  :  $\delta x_E \neq 0$  et  $\delta J_{gaba} = 0$

$$\frac{\delta sn}{\delta x_E} = \mathcal{A}_{sn, x_E} \quad (7.45)$$

Nous montrons que  $\mathcal{A}_{sn, x_E} > 0$  (la justification formelle est donnée dans l'article du chapitre 9). Une augmentation de  $x_E$  produit bien une augmentation de  $sn^*$ .  $x_E$  est un paramètre exciteur (voir Illustration 7.2).

- Perturbation de  $J_{gaba}$  :  $\delta x_E = 0$  et  $\delta J_{gaba} \neq 0$

$$\frac{\delta sn}{\delta J_{gaba}} = -sg \frac{\delta sn}{\delta x_E} = -sg \mathcal{A}_{sn, x_E} \quad (7.46)$$

Puisque  $\frac{\delta sn}{\delta x_E} = \mathcal{A}_{sn, x_E} > 0$  et  $sg > 0$ , alors  $\frac{\delta sn}{\delta J_{gaba}} < 0$ .

Une augmentation de  $J_{gaba}$  produit bien une diminution de  $sn^*$ .  $J_{gaba}$  est un paramètre inhibiteur (voir Illustration 7.3).

- Perturbation de  $x_E$  et  $J_{gaba}$  :

Dans l'équation (7.44), les perturbations  $\delta x_E$  et  $\delta J_{gaba}$  donnent des contributions à  $\delta sn^*$  de signes opposés : les effets des perturbations des paramètres  $x_E$  et  $J_{gaba}$  se compensent. Nous retrouvons dans le calcul analytique l'image de la balance décrite à la section 7.3.1.

En conclusion, le forçage sur le pool exciteur  $x_E$  et le paramètre du couplage inhibiteur  $J_{gaba}$  représentent des propriétés excitatrice et inhibitrice neurales locales qui permettent de faire pencher la balance E/I d'un côté ou de l'autre.

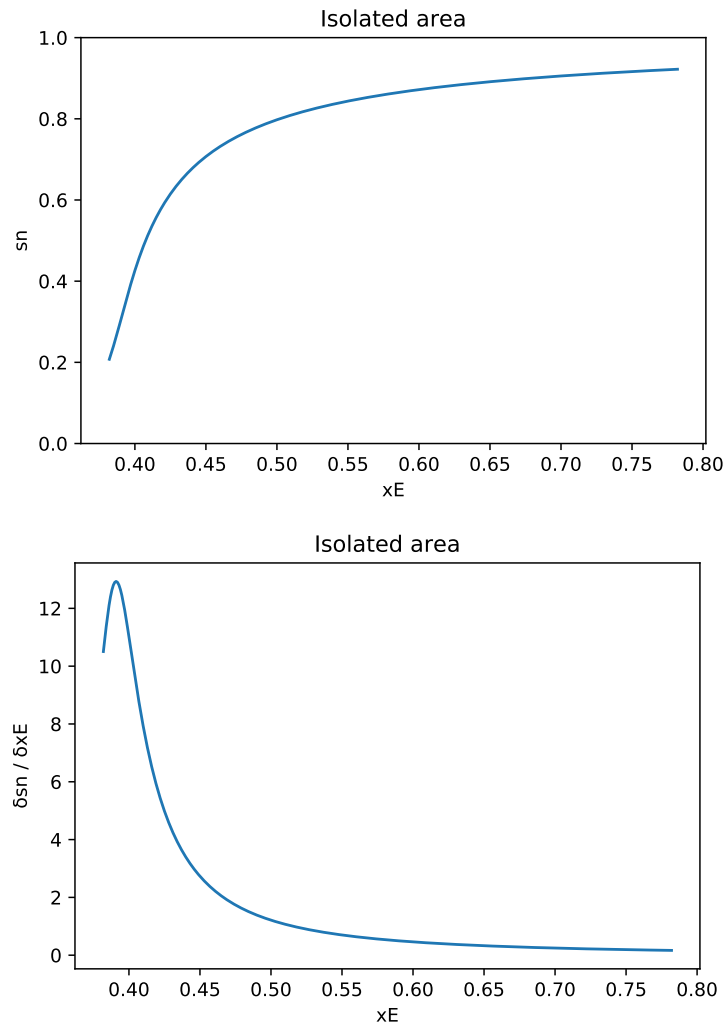


**Figure 7.2** – Haut : pour une valeur de  $J_{gaba}$  fixée (1.222 nA),  $sn$  augmente avec  $x_E$ .  
 Bas : la sensibilité est la dérivée de la courbe de gauche, et est donc bien positive.

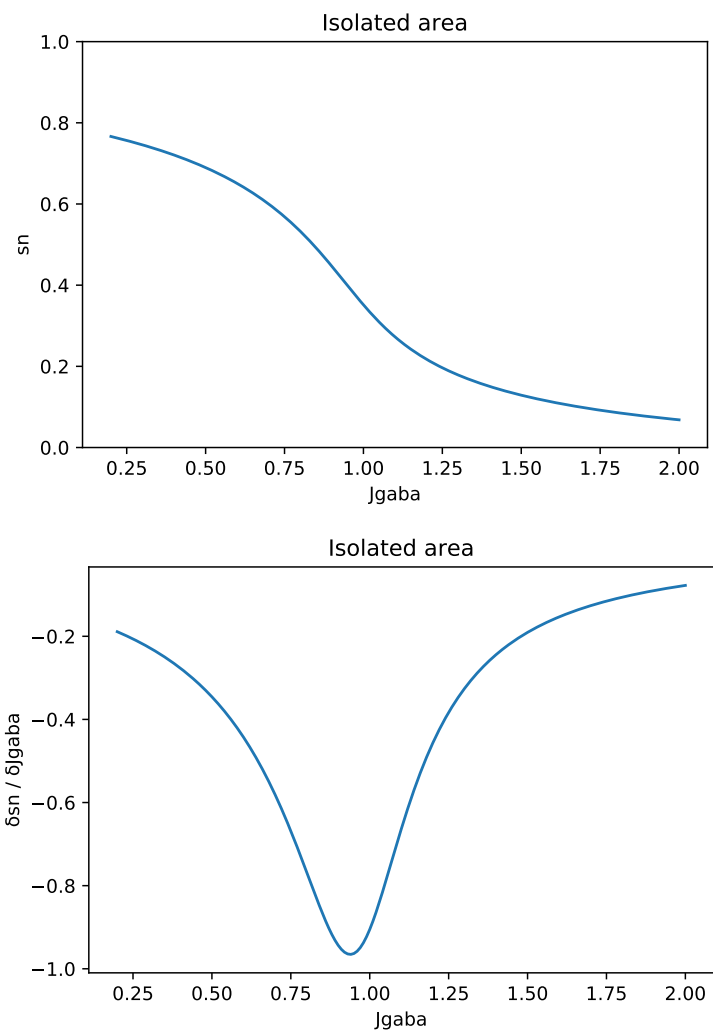


**Figure 7.3** – Haut : pour une valeur de  $x_E$  fixée (0.382),  $sn$  diminue avec  $J_{gaba}$ .  
 Bas : la sensibilité est la dérivée de la courbe de gauche, et est donc bien négative.



### 7.3.3 Choix des paramètres pour le système à deux régions

#### 7.3.3.1 Guidé par la question de la variabilité interindividuelle

Au chapitre 3, en nous basant sur des mesures expérimentales de contrôle inhibiteur des intrusions mnésiques ([Schmitz et al., 2017]), nous avons décrit deux sources possibles de variabilité interindividuelle du contrôle inhibiteur :

- l'émission : une activation neurale insuffisante des régions de contrôle
- la réception : modification du système gabaergique dans les régions cibles

Ces deux hypothèses nous guident sur le choix des paramètres intéressants à estimer pendant une tâche de contrôle inhibiteur, pour étudier la variabilité interindividuelle. Nous testerons l'effet de deux catégories de paramètres sur l'augmentation d'activité de régions de contrôle et la baisse d'activité de régions cibles : les forçages  $I_{stim_{rc}}$  dans les régions de contrôle (paramètres excitateurs)<sup>11</sup> et les coefficients de couplage  $J_{gaba_c}$  dans les régions cibles (paramètres inhibiteurs).  $J_{gaba_c}$  peut être interprété comme le poids synaptique du pool inhibiteur sur le pool excitateur (qui dépend, par exemple, du nombre de synapses, du nombre de canaux par synapse, ou d'autres éléments présentés au chapitre 3 dans l'encadré 3.3.3).

Ces hypothèses portent sur deux paramètres qui diffèreraient entre les sujets. Elles constituent un "métamodèle", c'est-à-dire un modèle sur des différences attendues entre les sujets<sup>12</sup>. Nous pouvons préciser l'effet attendu d'une variation de chacun de ces paramètres sur la baisse d'activité de la région cible, i.e. notre observable  $\Delta sn_c < 0$ .

#### 7.3.3.2 Hypothèse sur l'effet d'une augmentation de $I_{stim}$ dans la RC

Plus  $I_{stim}$  augmente, plus la baisse d'activité de la cible est grande (i.e.  $\Delta sn_c < 0$  augmente en valeur absolue).

#### 7.3.3.3 Hypothèse sur l'effet de $J_{gaba_c}$

Plus  $J_{gaba_c}$  augmente, plus la baisse d'activité de la cible est grande (i.e.  $\Delta sn_c < 0$  augmente en valeur absolue).

#### 7.3.3.4 Précision des hypothèses en lien avec la question de la variabilité interindividuelle

Nous pouvons alors "affiner le métamodèle", en précisant les hypothèses sur les relations d'ordre entre les paramètres d'intérêt ( $I_{stim}$  dans la région de contrôle et  $J_{gaba_c}$  dans la cible) pour différents sujets. Supposons deux groupes de sujets qui se distingueraient par

---

11. Pour rappel, les résultats de [Schmitz et al., 2017] n'indiquent pas que le "poids inhibiteur" dans les régions de contrôle des participants testés, puisse participer à la variabilité interindividuelle des mécanismes d'abaissement d'activité de l'hippocampe (sections 5.1.1.2, 5.4.4, 5.5.2). Nous n'interrogeons pas  $J_{gaba_{rc}}$ .

12. C'est un deuxième modèle par dessus un premier modèle qui donne les activités BOLD d'un participant au cours de la tâche de contrôle inhibiteur.

leur capacité à inhiber l'activité de la région cible pendant une tâche de contrôle inhibiteur des intrusions mnésiques : un groupe high-forgetters (HF) et un groupe low-forgetters (LF), comme qualifiés dans [Benoit and Anderson, 2012]. Suite à une activation de la région de contrôle, le groupe HF abaisserait davantage l'activité de la région cible que le groupe LF. Dans notre contexte de modélisation, tous les autres paramètres non mentionnés étant supposés égaux entre ces deux groupes, nous pouvons écrire une relation d'ordre en :

- Emission :  $I_{stim_{rc}, HF} > I_{stim_{rc}, LF}$
- Réception :  $J_{gaba_c, HF} > J_{gaba_c, LF}$

## 7.4 Précision des questions de recherche : explicitation des objectifs précis de la thèse

L'ensemble du travail présenté dans ce chapitre permet de préciser les questions de la section 6.4. Les objectifs de la thèse portent sur :

- A) Comment connecter les régions au niveau neural pour obtenir une baisse de l'activité neurale excitatrice au régime stationnaire dans la région cible (i.e.  $\Delta sn_c^* = sn_{stim_c}^* - sn_{B_c}^* < 0$ ), quand l'activité neurale excitatrice au régime stationnaire dans la région de contrôle augmente (i.e.  $\Delta sn_{rc}^* = sn_{stim_{rc}}^* - sn_{B_{rc}}^* > 0$ ) ?
- B) Pour une architecture de connectivité fixée, comment  $I_{stim}$  dans les régions de contrôle et  $J_{gaba_c}$  dans les régions cibles influencent-ils les variations des activités neurales excitatrices au régime stationnaire, i.e.  $\Delta sn_i^*$ , des régions impliquées dans une tâche de contrôle inhibiteur ; en particulier la baisse d'activité des régions cibles  $\Delta sn_c^*$  due à l'augmentation d'activité des régions de contrôle  $\Delta sn_{rc}^*$  ?
- C) Comment l'analyse du modèle direct éclaire-t-elle le problème de l'inversion et nous permet de construire une procédure d'inversion (en 2 étapes) pour inférer les paramètres neuraux  $I_{stim}$  et  $J_{gaba}$  de chaque région à partir de signaux BOLD recueillis pendant une tâche de contrôle inhibiteur (quel choix de modèle direct pertinent, sur quelle observable porte l'inversion, comment construire une procédure d'inversion etc. ) ?
- D) La mise en oeuvre numérique du point C) nous permet d'estimer les paramètres neuraux  $I_{stim}$  et  $J_{gaba}$  à partir de mesures BOLD pendant une tâche de contrôle inhibiteur. Nous prenons comme cas d'étude le contrôle inhibiteur des intrusions mnésiques (testé avec le paradigme TNT). Est-il possible de présenter une preuve de concept de l'inversion sur les données d'un sujet (un participant non exposé de l'étude REMEMBER) ?

## 7.5 Présentation des étapes du travail

La thèse s'organise en deux parties :

- Première partie (chapitre 9) : Analyse du modèle neural direct menée sur 2 régions. Ce travail permet d'éclairer les questions A) et B).

- Deuxième partie (chapitres 10 et 11) : Construction de la procédure d'inversion et inversion des données d'un sujet (une personne servant de témoin non exposé dans l'étude REMEMBER).

Nous nous servons du travail mené dans la première partie pour construire la procédure d'inversion en 2 étapes. Les réponses apportées à A) permettent de fixer une architecture de connectivité pour obtenir une baisse de l'activité neurale excitatrice au régime stationnaire dans la région cible ( $\Delta sn_c^* < 0$ ), quand l'activité neurale excitatrice au régime stationnaire dans la région de contrôle augmente ( $\Delta sn_{rc}^* > 0$ ). L'exploration de la question B) nous assure de l'impossibilité d'inverser sur les  $\Delta sn^*$  et donc sur les variations du signal BOLD, mais la possibilité d'inverser sur les décours temporels.

Après avoir formulé la méthode d'inversion, nous l'implémenterons numériquement pour la tester avec des simulations. Enfin, nous la mettrons en oeuvre sur un jeu de données BOLD obtenues pendant une tâche de contrôle inhibiteur des intrusions mnésiques (paradigme TNT).

# Troisième partie

## Analyse du modèle direct



# Chapitre 8

## Préambule de l'article

Le contexte général de cette thèse est de tenter de faire un lien entre des observables caractéristiques du contrôle inhibiteur, telles qu'obtenues en IRMf (baisse du signal BOLD dans certaines régions cérébrales lors de tâches de type TNT), et des descripteurs à l'échelle neurale.

Une fois posés les choix de modélisation, cette mise en correspondance des échelles vise ultimement à réaliser une identification des paramètres d'intérêt à partir des données (procédure d'inversion statistique). Si on peut toujours tenter de "faire tourner" une procédure d'inversion (e.g. de type "toolbox VBA" qui met en oeuvre une procédure d'inversion bayésienne variationnelle), il est d'une grande utilité d'analyser, par ailleurs, le comportement du modèle afin de s'assurer d'une part de la pertinence du modèle pour implémenter le comportement recherché (baisse du signal BOLD dans les régions sous l'effet de l'activation des régions de contrôle), et pour éclairer, d'autre part, la possibilité d'une inversion effective (e.g. importance du niveau de bruit, effets retards du BOLD sur le neural, plages de paramètres pertinentes, modélisation des distributions a priori...).

Dans le cas présent, ayant adopté le modèle de [Naskar et al., 2021] pour l'échelle neurale, la première question qui s'est posée a concerné le type de connectivité à introduire entre les régions, pour obtenir le comportement recherché. Nous aurions pu, comme il se fait le plus souvent dans la communauté des modélisations neuronales, procéder par des séries de simulations numériques pour balayer l'espace des (nombreux) paramètres. La collaboration avec un groupe de physiciens des modèles du climat nous a, en fait, ouvert la voie à des développements analytiques qui permettent de tirer des conclusions valant quelques soient les valeurs des paramètres. L'article présente une partie de ces développements, où le modèle neural est réduit à un système simplifié composé de deux régions : une région de contrôle et une région cible. Chaque région est composée de deux pools (un pool excitateur et un pool inhibiteur), et nous gardons la contrainte classique que seul le pool excitateur envoient des projections vers l'autre région, mais nous considérons que ces projections peuvent se faire sur l'un ou l'autre pool (ou les deux) de la région d'arrivée .

Le premier objectif étant de comprendre le rôle de la connectivité entre les régions, nous avons exploré les quatre schémas de connectivité les plus tranchés :

- Excitation mutuelle : les deux pools excitateurs projettent vers le pool excitateur de l'autre région,

- La cible inhibe le contrôle : le pool excitateur de la région de contrôle projette vers le pool excitateur de la région cible, et celui de la région cible projette vers le pool inhibiteur de la région de contrôle,
- Le contrôle inhibe la cible : les rôles sont inversés par rapport au schéma précédent,
- Inhibition mutuelle : les pools excitateurs projettent tous deux vers le pool inhibiteur de l'autre région.

En construisant analytiquement l'expression de la sensibilité de la réponse de la région cible à une perturbation positive de la région de contrôle, nous montrons que seuls les deux derniers schémas peuvent implémenter le contrôle inhibiteur, et ce, quelques soient les paramètres.

Nous explorons ensuite, pour les quatre schémas de connectivité, le rôle de deux paramètres d'intérêt crucial : le premier décrit la variation d'activité dans la région de contrôle qui active le contrôle inhibiteur, le second décrit l'intensité de l'auto-inhibition de la région par elle-même. Ces deux paramètres font l'objet de la procédure d'inversion présentée en quatrième Partie du manuscrit. Au regard de cette dernière, un résultat important de l'article est qu'une variation d'activité dans la région cible peut correspondre à deux valeurs très différentes du second paramètre, et que donc l'inversion n'est pas univoque sur un signal différentiel.

## Chapitre 9

# Analyse du comportement du modèle pour 2 régions

### 9.1 Article



# Sensitivity analysis enlightens inhibitory connectivity in a Neural Mass Model under Control-Target mode

Vallet Anaïs<sup>1</sup>, Blanco Stéphane<sup>2</sup>, Chevallier Coline<sup>2,3</sup>, Eustache Francis<sup>1</sup>, Gautrais Jacques<sup>2,3,\*</sup>, Granpeix Jean-Yves<sup>4</sup>, Joly Jean-Louis<sup>2</sup>, Segobin Shailendra<sup>1</sup>, Gagnepain Pierre<sup>1</sup>

**1** Normandie Univ, Neuropsychol & Imagerie Memoire Humaine, GIP Cyceron, UNICAEN, PSL Res Univ, EPHE, INSERM, U1077, CHU Caen, F-14000 Caen, France

**2** LAPLACE, UMR 5213 - Université Paul Sabatier, 118, Route de Narbonne, Toulouse Cedex 31062, France

**3** Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, France

**4** LMD/IPSL/SU, CNRS, Paris 75005, France

\* [jacques.gautrais@cnrs.fr](mailto:jacques.gautrais@cnrs.fr)

## Abstract

Inhibitory control allows a person to repel irrelevant or inappropriate responses. Using functional MRI (fMRI), inhibition is characterized by a decrease in BOLD activity in some Target brain regions when some Control brain regions are subject to an increase in BOLD activity. At neural level, these regions have a pool of excitation (E) and inhibitory (I) neurons that are connected to one another to achieve the desired output (inhibition or not). Building from the neural model by Naskar et al. (Network Neuroscience, 2021), our aim was to understand how these brain regions can be connected at neural level to implement inhibitory control.

Focusing on this neural model in its stationary state (Fixed Point, FP), we built a sensitivity analysis in the neighbourhood of the FP using two brain areas as the simplest configuration model, a control area and a target area. We explore four kinds of connectivities: mutual excitation (E-E), where regions are connected by their excitatory pools, Control-to-Target inhibition (E-I), where pool E of Control project on pool I of Target, Target-to-Control inhibition (I-E), and mutual inhibition (I-I).

Using this two-areas model, we show analytically that sensitivities of Target activity to Control perturbation are always positive in E-E and I-E cases, and always negative in E-I and I-I cases. Hence, only E-I and I-I connectivities can implement inhibitory control. We next explore and explain the structure of sensitivities at the levels of isolated pools, of isolated regions, and of the full system, in the four cases of connectivities, and with a focus upon a parameter controlling the strength of the self-inhibition within Target region.

Finally, analyzing the Target response to a finite-size perturbation of the external forcing in the Control area as a function of this parameter, we observed an unanticipated effect : the amplitude of Target response evolves non monotonically with this parameter, and is maximal for an optimal middle range value. As a side effect, Target can display the same response for two values of self-inhibition, that are respectively lower and higher than the optimal value. Any “causal” viewpoint between activation of Control and inhibition of Target should then be tempered, if feedback from Target to Control has to be considered.

Beyond the configurations explored in the present work, our methodology to build analytical sensitivities by nesting accross levels (pool, region, system) lays the groundwork for building nested sensitivities for more complex network configurations, which would help identify the network of regions for inhibitory control neural implementation.

# Contents

|          |                                                                                           |           |
|----------|-------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                       | <b>2</b>  |
| 1.1      | Motivation . . . . .                                                                      | 2         |
| 1.2      | The model . . . . .                                                                       | 3         |
| 1.2.1    | Modeling background . . . . .                                                             | 3         |
| 1.2.2    | Formal expression of the model . . . . .                                                  | 4         |
| 1.2.3    | Generalizing connectivity . . . . .                                                       | 5         |
| 1.3      | Analysis of connectivity effect based on sensitivities . . . . .                          | 6         |
| 1.3.1    | Control-Target mode . . . . .                                                             | 6         |
| 1.3.2    | Formalizing the sensitivity of interest as a hierarchy of nested sensitivities . . . . .  | 7         |
| <b>2</b> | <b>Building a hierarchy of nested sensitivities</b>                                       | <b>8</b>  |
| 2.1      | At the pool level . . . . .                                                               | 8         |
| 2.2      | At the area level . . . . .                                                               | 10        |
| 2.3      | At two-area system level . . . . .                                                        | 11        |
| 2.4      | Control-Target system . . . . .                                                           | 11        |
| 2.5      | Effect of connectivity upon Target sensitivity . . . . .                                  | 12        |
| <b>3</b> | <b>Interpretation and discussion</b>                                                      | <b>15</b> |
| 3.1      | Effect of self-coupling in pools . . . . .                                                | 15        |
| 3.2      | Effect of coupling two pools . . . . .                                                    | 15        |
| 3.3      | Effect of connectivity upon sensitivities in Control-Target system . . . .                | 17        |
| 3.4      | Consequences for system responses upon step activation of Control area                    | 19        |
| 3.4.1    | Illustrations with BOLD signals . . . . .                                                 | 19        |
| 3.4.2    | Effect of self-inhibitory strength within Target area depending on connectivity . . . . . | 20        |
| <b>4</b> | <b>Conclusion and perspectives</b>                                                        | <b>22</b> |
| <b>5</b> | <b>Supporting information</b>                                                             | <b>23</b> |
| 5.1      | S1 Appendix. . . . .                                                                      | 23        |
| 5.2      | S2 Appendix. . . . .                                                                      | 23        |
| <b>6</b> | <b>Acknowledgments</b>                                                                    | <b>23</b> |
| <b>7</b> | <b>Figures</b>                                                                            | <b>26</b> |

|            |                                                                                                                                                                                                                                                                                                                                          |                  |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| <b>1</b>   | <b>Introduction</b>                                                                                                                                                                                                                                                                                                                      | <b>1</b>         |
| <b>1.1</b> | <b>Motivation</b>                                                                                                                                                                                                                                                                                                                        | <b>2</b>         |
|            | In recent years, a large effort has been made to design and implement large-scale simulations of human brain, in relation to functional neuroimaging data (EEG, MEG, functional MRI, ...) in order to draw inferences regarding neurophysiological mechanisms [1–4]. In this domain, the class of so-called biophysical models represent | 3<br>4<br>5<br>6 |

the brain as a graph of inter-connected neural masses, and the neural dynamics are used to link anatomical connectivity (aka structural connectivity, or structural connectome) to functional imaging data (e.g. BOLD signals from fMRI) [5–10]. A widespread practice is to use recordings of fMRI to infer the so-called functional connectivity from correlations between time courses of areas’ activity level. This functional connectivity can then be instrumental to tune biophysical models’ parameters [10]. In such approaches, connectivity among cortical areas plays then a crucial role into shaping dynamics and a huge effort has been made to obtain and characterize a precise mapping of brain connectivity by *in vivo* tractography [11–15].

Connectivity among cortical areas is known to be supported by long-range connections that can be only excitatory, and in most biophysical models where neural masses activity level are represented by one state variable, they are considered as mutually excitatory. There are however some contexts in which the influence from one area upon another one is clearly considered as inhibitory by nature, the most prominent paradigm being every kind of inhibitory control [16–19], associated with fronto-basal-ganglia circuits [20, 21]. In such context, it must be considered that the activity coming from a control region may have an inhibitory effect upon target populations [2], namely inhibitory control should translate increasing activities in some control areas into decreased activities in their target areas.

In this paper, we offer an analytical account of how such inhibitory connectivity can affect activity levels in a widely used biophysical model [6]. For this purpose, we develop a methodological framework for building a hierarchy of nested sensitivities and we analyse the structure of the sensitivities of target area responses, depending on connectivity.

## 1.2 The model

### 1.2.1 Modeling background

We start from the dynamic mean field (DMF) model of excitatory/inhibitory neural populations first introduced by Deco et al. [6] (after [22]), in an effort to account for resting-state networks (RSNs). For each area, mean-field description summarises the corresponding underlying spiking neural network, in which excitatory pyramidal neurones and GABAergic inhibitory neurones are mutually interconnected. At the mean field scale of description, areas are interconnected by long range excitatory signals (NMDA-mediated connections), weighted by anatomical connectivity and a global coupling strength.

These mean field equations are then integrated numerically, spanning initial conditions, in order to study the stationary state landscape. They tuned the global inter-areal coupling strength in the model to fMRI-based functional connectivity in humans and showed that resting brain operates at the edge of multi-stability [6]. This DMF model was then used to explore more exhaustively the dynamics of the system [23], especially how noise propagation reflects the double effect of anatomical connectivity and slow dynamics around the spontaneous low-firing stationary state, resulting in the functional connectivity (i.e. pairwise correlations between cortical areas). Here, they show that global inter-areal coupling strength fitted to fMRI data rather indicates that brain operates at full multi-stability, yet with resting state being a stable stationary state. Furthermore, they show that pairwise correlations depend on the sensitivities of each area to the inputs coming from other areas. Since these sensitivities depend in turn upon the activity levels at the considered stationary state, affecting these activity levels (e.g. by over-activating a subset of areas) results in a different stationary state, and hence to a different functional connectivity, while anatomical connectivity remains the same [23, 24]. The model has then been used to

infer anatomical connectivity from functional connectivity, relating structure and function [25], to explore the possibility of dynamical transitions between multiple RSNs depending on noise level [8], to suggest that human brain anatomical connectivity may be evolutionary tuned to display the largest diversity in activities of area clusters in response to an increased inter-regional coupling [9]...

In a refinement of this model, Deco et al. [7] propose to better describe neural mass areas by explicitly considering that they are composed of two pools of neurones: an excitatory one and an inhibitory one, so that the local feedback inhibition can be locally constrained. Under these constraints, the empirical finding that intra-areal correlations should remain poor is recovered, whereas it is violated when using inter-areal coupling in single-pool modelling. With this choice, the intra-areal activity can be made decorrelated while the inter-areal correlations keep reflecting functional connectivity. Moreover, and as a consequence of this modelling choice, long range excitatory signals can now target either excitatory pools in other areas or their inhibitory pools, or both. We then turned to this version of the model since it allows to represent feedforward inhibition, when one area can stimulate the inhibitory pool of another area. Also, Deco et al. [7] consider the effect of stimulating some random subset (10%) of areas by exogenous stimuli, representing task-related signals. Interestingly, in this version, the dynamics yield only one stable stationary state at resting (for reasonable coupling strength avoiding epileptical divergence), and must be stimulated by these exogenous stimuli to switch state. They advocate that this refinement yield more robust prediction of functional connectivity, and more realistic responses to external stimuli [7]. This two-pool area model at resting state has been used by Glomb et al. to recover the spatio-temporal structure of RSNs from fluctuations around the single stationary state [26]. As a final extension, Naskar et al. justify the kinetic parameters driving the average synaptic gating variable based on GABA and glutamate concentrations [27]. We start from this version.

### 1.2.2 Formal expression of the model

In this model, the state variables at neural mass scale,  $sn(t)$  and  $sg(t)$ , are called *average synaptic gating variables* and represent the open fraction of NMDA (resp. GABA) synaptic channels for the excitatory (resp. inhibitory) pools. For each area  $i$ , their dynamics are given by the following ODEs :

$$\begin{cases} \dot{sn}_i(t) = -\beta^E sn_i(t) + \alpha^E T_{glu} rn_i(t)(1 - sn_i(t)) + \sigma \nu_i(t) \\ \dot{sg}_i(t) = -\beta^I sg_i(t) + \alpha^I T_{gaba} rg_i(t)(1 - sg_i(t)) + \sigma \nu_i(t) \end{cases} \quad (1)$$

where the intermediate variables  $rn(t)$  and  $rg(t)$  represent the mean-field firing rates of each pool, which are given by input-output functions (inspired from [28]):

$$\begin{cases} rn_i(t) = \frac{a_E x n_i(t) - b_E}{1 - e^{-d_E(a_E x n_i(t) - b_E)}} \\ rg_i(t) = \frac{a_I x g_i(t) - b_I}{1 - e^{-d_I(a_I x g_i(t) - b_I)}} \end{cases} \quad (2)$$

and depend, in turn, upon the input currents  $xn(t)$  and  $xg(t)$ , that sum up all incoming contributions:

$$\begin{cases} xn_i(t) = W_E I_0 + W_+ J_{nmda} sn_i(t) - J_{gaba,i} sg_i(t) + G J_{nmda} \sum_{j \neq i} C_{ij} sn_j(t) \\ xg_i(t) = W_I I_0 + J_{nmda} sn_i(t) - sg_i(t) \end{cases} \quad (3)$$

In these representations, coupling are made explicit in Eq 3, namely coupling are modelled as flows of information about input currents.

At the area scale, there are four couplings: self-excitatory coupling of the excitatory pool, excitatory coupling from the excitatory pool upon the inhibitory pool, self-inhibitory coupling of the inhibitory pool and inhibitory coupling from the inhibitory pool upon the excitatory pool (see Fig 1).

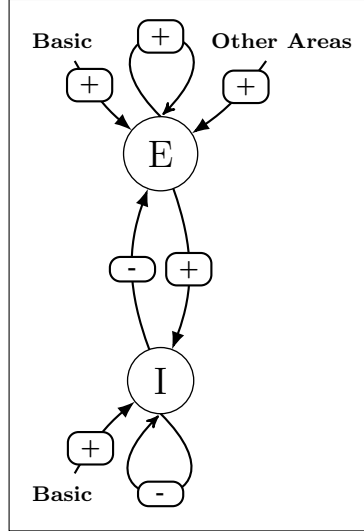


**Fig 1.** Graphical representation of Eqs. 1 to 3 for a given area.

Coupling among areas are represented through the sum term where  $C_{ij}$  represent normalised anatomical connectivity,  $J_{nmda}$  represents excitatory synaptic coupling strength, and  $G$  is a positive free parameter which scales the global coupling strength between areas. The parameter  $J_{gaba,i}$  represents inhibitory synaptic coupling strength, and can be locally adjusted to regulate the balance between long-range excitation and local feedback inhibition so as to ensure homeostasis [7].

Note that without the so-called *basic input current*,  $I_0$ , the system would present no activity other than noise  $\sigma$ , which is set to a negligible level throughout the papers considered above.

### 1.2.3 Generalizing connectivity

In the model above, coupling between areas is only between excitatory pools. To allow flexibility, we introduce a further parameter  $k_{E_{ij}} \in [0; 1]$  which modulates the fraction of long range excitatory input coming from area  $j$  that will project onto the excitatory pool of area  $i$ , the remaining part being considered to project onto its inhibitory pool. We also denote  $B_{E_i}$  and  $B_{I_i}$  the effective external input current so that they can vary among areas, and we introduce a parameter  $J_-$  to get an homogenous expression.

With these extensions, our system only modifies Eq 3, which now reads:

$$\begin{cases} xn_i(t) = B_{E_i} & + W_+ J_{nmda} sn_i(t) & - J_{gaba,i} sg_i(t) & + \sum_{j \neq i} k_{E_{ij}} \kappa_{ij} sn_j(t) \\ xg_i(t) = B_{I_i} & + J_{nmda} sn_i(t) & - J_- sg_i(t) & + \sum_{j \neq i} (1 - k_{E_{ij}}) \kappa_{ij} sn_j(t) \end{cases} \quad (4)$$

where  $\kappa_{ij} \equiv G J_{nmda} C_{ij}$ ,  $W_+ = 1.4$ ,  $J_- = 1$  nA in [27].

116

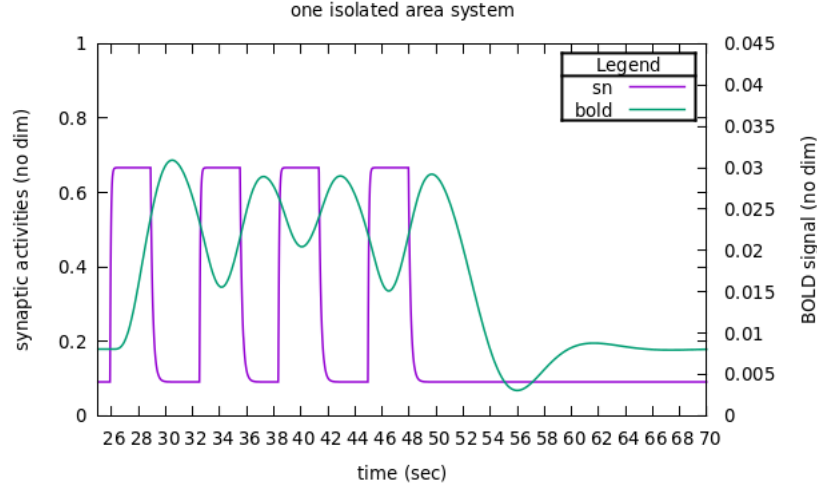


**Fig 2.** Basic response of an isolated area to inputs.

### 1.3 Analysis of connectivity effect based on sensitivities

117

#### 1.3.1 Control-Target mode

118

Since we are interested in inhibitory control process, we can distinguish two kinds of areas, splitting the system into a subset of control areas, and a subset of target areas.

119

In the aim to analyse how connectivity among areas will shape the behavioral response of the system, the main picture is then to understand how target areas respond to upregulation applied to control areas, once we take into account that target areas can themselves project back to control areas and have a feedback effect upon their upregulation.

121

122

123

124

125

For a single area, a typical time course of the response of the excitatory pool to inputs is shown in Fig 2 (purple line) together with the corresponding BOLD signal (green line), as modelled in [7]. In response to alternate switching of  $B_E$  value (between 0.382 and 0.482 nA), one can see that the activity level  $sn(t)$  adjusts very quickly, whereas BOLD signal follows with an observable delay. At the neural mass level, we can then study the behavior of the system considering only the value  $sn(t)$  at steady states, which are fixed points of the dynamics. Hence, we can address the question of the effect of connectivity looking only how the fixed points are affected by parameters, namely to predict how the target activity level will be affected.

126

127

128

129

130

131

132

133

134

For this, we will develop a *sensitivity* analysis, in which we build an analytical expression for how the system responds to an infinitesimal perturbation upon some control areas, and how it affects activity levels in all areas (including those upon which the perturbation is applied). The point of interest is then how the activity levels of target areas are modified, whether they are up-regulated or down-regulated, which is given by the sign of sensitivities (i.e. the sign of the derivative of activity level with regards to the perturbation). Once such sensitivities expression are established, they are used to analyze the behavior of the system, depending on the kind of connectivity.

135

136

137

138

139

140

141

142

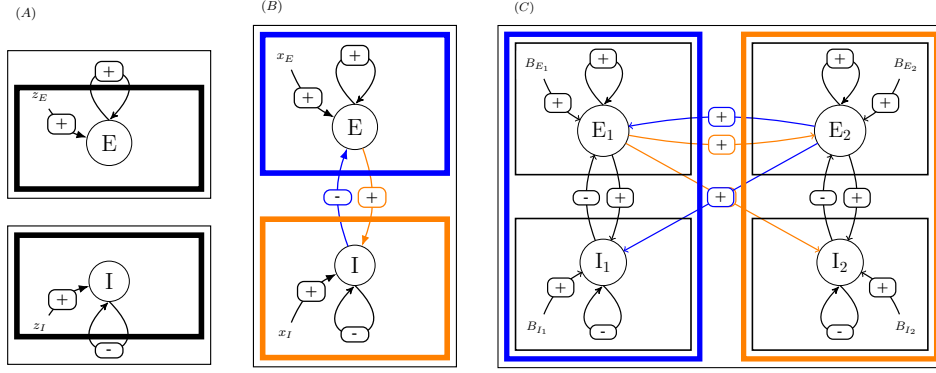


**Fig 3.** *Principle of the method.* (A) One excitatory pool, either excitatory or inhibitory. (B) One area with two coupled pools. (C) Two coupled areas. In case (A), the sensitivity of the activity level of the pool to perturbation upon its external forcing ( $z_E$ ) is expressed as function of its sensitivity when uncoupled with itself. In case (B), the sensitivity of the activity levels of a pool to perturbation upon its external forcing ( $x_E$  or  $x_I$ ) is expressed as a function of its sensitivity when uncoupled from the alternate pool, namely, the expression found in case (A). In case (C), the sensitivity of activity level of the excitatory pool of one area is expressed as a function of its sensitivity when uncoupled from the alternate area, namely, the expression found in case (B). In the three cases, sensitivities at system level are expressed as functions of sensitivities at component level (colored boxes), by opening the feedback loop (cutting colored connections).

### 1.3.2 Formalizing the sensitivity of interest as a hierarchy of nested sensitivities

The sensitivities are to be estimated at the fixed point corresponding to the resting state. Following Naskar et al. [27], the parameters  $J_{gaba_i}$  will be adjusted on a per-area basis such that resting state corresponds to a firing rate of the excitatory pools prescribed at 3 Hz.

Once this fixed point is established, we apply an infinitesimal perturbation  $\delta B_{E_i}$  upon  $B_{E_i}$  in control areas, and we will observe its effects upon state variables  $sn_j$  in target areas.

Hence, basically sensitivities are expressed as  $\frac{\delta sn_j}{\delta B_{E_i}}$ .

To derive the analytical expression, we will proceed with a hierarchy of sensitivities, illustrated in Fig 3. Starting from the sensitivities of excitatory and inhibitory pools when considered isolated, we can express the sensitivity for one area where the two pools are coupled. At the next level, the sensitivity of coupled areas can be expressed as a function of sensitivities of each area when considered isolated. At the next level, the sensitivity for areas belonging to two coupled subsystems can be expressed as functions of sensitivities of areas when considered in their subsystem alone. In the present paper, we will focus on a system consisting of two areas, one control (denoted as  $C$ ) and one target (denoted as  $T$ ), and leave generalization to larger systems to a future companion paper.

Hence, we look for a formal expression for  $\frac{\delta sn_T}{\delta B_{E_C}}$ .

## 2 Building a hierarchy of nested sensitivities

The complete system can be regarded as made of three nested levels : the system itself, which contains areas, which contain pools. Here, we will expose how we build the analytical expressions of the sensitivities at one level as functions of sensitivities at lower level. Since, the principle to build sensitivities at one level from sensitivities at the lower level is generic across levels, we will fully expose this principle in the simplest case (from pool level to area level). For further levels integration, the formal details are given in [S1 Appendix](#) and we only report the corresponding results.

### 2.1 At the pool level

Focusing on one isolated pool (here, one excitatory pool, see Fig 3(A)) at stationary state, its state variables must obey:

$$\begin{cases} 0 = -\beta^E sn^* + \alpha^E T_{glu}(1 - sn^*)rn^* \equiv fn(sn^*, rn^*) \\ rn^* = \frac{a_E xn^* - b_E}{1 - e^{-d_E(a_E xn^* - b_E)}} \equiv hn(xn^*) \\ xn^* = W_+ J_{nmda} sn^* + z_E \equiv wn(sn^*, z_E) \end{cases} \quad (5)$$

where the *star* notation is to designate fixed point values and  $z_E$  represents the external forcing.

If we now consider a perturbation upon  $z_E$ , we get a new model, which reads:

$$\begin{cases} fn(sn, rn) = 0 \\ rn = hn(xn) \\ xn = wn(sn, z_E + \delta z_E) \end{cases} \quad (6)$$

and we want the formal expression for the sensitivity  $\frac{\delta sn}{\delta z_E}$ .

The linearization for the perturbation yields:

$$\begin{cases} \frac{\partial fn}{\partial sn} \delta sn + \frac{\partial fn}{\partial rn} \delta rn = 0 \\ \delta rn = \frac{dhn}{dxn} \delta xn \equiv hn' \delta xn \\ \delta xn = \frac{\partial wn}{\partial sn} \delta sn + \frac{\partial wn}{\partial z_E} \delta z_E \end{cases} \quad (7)$$

where the derivatives are to be evaluated at fixed point values ( the expression for  $hn'$  is given in [S1 Appendix](#)).

Plugging the last two equations into the first, we get:

$$\frac{\partial fn}{\partial sn} \delta sn + \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \left[ \frac{\partial wn}{\partial sn} \delta sn + \frac{\partial wn}{\partial z_E} \delta z_E \right] = 0 \quad (8)$$

Rearranging to read how  $\delta sn$  depends upon  $\delta z_E$ :

$$\left( 1 + \frac{\partial fn}{\partial sn}^{-1} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial sn} \right) \delta sn = - \frac{\partial fn}{\partial sn}^{-1} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial z_E} \delta z_E \quad (9)$$

This expression is of a primary interest in the methodology to build hierarchical sensitivities, as it represents the *canonical* form for identifying open loop and close loop responses to perturbation upon  $z_E$ , that are related through the feedback gain  $g_{sn, z_E}$ , that would be identified as:



$$(1 - g_{sn,z_E})\delta sn = \delta sn^O \quad (10)$$

Indeed, in the present case of the isolated excitatory pool, what appears as the support of non linearity is the self-amplification of the excitatory pool by the recurrent connectivity, which is expressed in  $wn$  through the parameter  $W_+$ .

In the absence of this recurrent connectivity (setting  $W_+ = 0$ , hence, *opening the loop* between  $sn^*$  and itself), we would have  $\frac{\partial wn}{\partial sn} = 0$ , so that the only effect of perturbing  $z_E$  would translate directly into a perturbation of  $sn^*$  in the absence of the feedback loop. This would amount to consider *opening* the feedback loop of  $sn$  upon itself and we can denote the open loop response to perturbation as  $\delta sn^O$ .

Then, the open loop sensitivity is defined as:

$$\mathcal{P}_{sn,z_E}^O \equiv \frac{\delta sn^O}{\delta z_E} = -\frac{\partial fn}{\partial sn}^{-1} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial z_E} \quad (11)$$

Then, we can write the *close loop* sensitivity  $\mathcal{P}_{sn,z_E}$  as a function of the *open loop* sensitivity, following:

$$\mathcal{P}_{sn,z_E} \equiv \frac{\delta sn}{\delta z_E} = \frac{1}{1 - g_{sn,z_E}} \mathcal{P}_{sn,z_E}^O \quad (12)$$

where

$$g_{sn,z_E} \equiv -\frac{\partial fn}{\partial sn}^{-1} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial sn} \quad (13)$$

defines the feedback gain for  $sn$  upon perturbation of  $z_E$  through the feedback loop between  $sn^*$  and itself. From here, we can analyze the role of closing the loop, depending on the value of  $g$ , hence predicting the behaviour of the system in response to a perturbation on external input, namely:

- If  $g < 0$ , then  $\frac{1}{1-g} \in ]0; 1[$ , so that  $\mathcal{P}_{sn,z_E} < \mathcal{P}_{sn,z_E}^O$ . Coupling with a negative feedback gain dampens the sensitivity to inputs.
- If  $0 < g < 1$ , then  $\frac{1}{1-g} \in ]0; +\infty[$ , so that  $\mathcal{P}_{sn,z_E} > \mathcal{P}_{sn,z_E}^O$ . Coupling with a positive feedback gain lower than 1 enhances the sensitivity to inputs.
- $g \geq 1$  leads the system in uncharted territories that have to be studied on a per-case basis.

In the case of the isolated excitatory pool, the feedback gain is positive by construction (see Fig. 5).

Finally, we can conclude that (see S1 Appendix for details):

$$\mathcal{P}_{sn,z_E} = \left( \frac{\beta_E + \alpha_E T_{glu} hn(wn(sn^*, z_E))}{\alpha_E T_{glu} (1 - sn^*) hn'(wn(sn^*, z_E))} - W_+ J_{nmda} \right)^{-1} \equiv \varphi_{n_E}(sn^*, z_E) \quad (14)$$

where we define the function  $\varphi_{n_E}$  so as to explicitly express the dependencies of the sensitivity for the isolated excitatory pool.

The main point here is to consider that this analytical expression for the sensitivity depends both upon the external forcing, and upon the fixed point under consideration (there can be more than one). The expression will remain the same when we will connect the excitatory pool to the inhibitory one, even if its value will change, since the coupling will modify the values of  $sn^*$  and  $z_E$  at which to compute it. In the case of the

isolated pool,  $z_E$  is prescribed as the effective external input, which in turn fixes the  
fixed point value  $sn^*$ , whereas in coupled situations,  $z_E$  will also depend upon the  
inputs coming from other parts of the system, which will depend, in turn, upon the  
stationary values they take in the coupled situation.

Regarding the inhibitory pool, we proceed the same way (see [S1 Appendix](#)), and we  
get:

$$\mathcal{P}_{S_g, z_I} = \left( \frac{\beta_I + \alpha_I T_{gaba} h g(w g(sg^*, z_I))}{\alpha_I T_{gaba} (1 - sg^*) h g'(w g(sg^*, z_I))} + J_- \right)^{-1} \equiv \varphi_{g_I}(sg^*, z_I) \quad (15)$$

where function  $\varphi_{g_I}$  explicitly express the dependencies of the sensitivity for the  
isolated inhibitory pool.

In the case of the isolated inhibitory pool, the feedback gain is negative by  
construction (see [Fig. 6](#)).

## 2.2 At the area level

Each area contains coupled excitatory and inhibitory pools (see [Fig 3\(B\)](#)). Here, we are  
interested in the sensitivity of excitatory and inhibitory pool activity to excitatory  
forcings.

In order to build an analytical expression of these sensitivities, we will consider here  
again open loop condition versus close loop, now considering that the loop to be opened  
is the coupling between the two pools.

Open loop sensitivities become the sensitivities of the pools to a perturbation of  
either  $x_E$  or  $x_I$  when the connectivity between the two pools is nullified (i.e. in [Fig 3\(B\)](#),  
when the colored connectivity is cut), i.e. the sensitivities to a perturbation upon  
external forcings (now denoted  $x_E$  and  $x_I$ ) when the feedback between the two pools is  
absent, namely the pool-level sensitivities that have just been built in the previous  
section.

We denote  $\mathcal{A}_{sn, x_E}$  (respectively  $\mathcal{A}_{sn, x_I}$ ) the sensitivity of the excitatory pool activity  
 $sn$  to the forcing  $x_E$  applied upon the excitatory pool (respectively  $x_I$  applied upon the  
inhibitory pool), and  $\mathcal{A}_{sg, x_E}$  (respectively  $\mathcal{A}_{sg, x_I}$ ) the sensitivity of the inhibitory pool  
activity  $sg$  to the same forcing. We denote  $\mathcal{A}_{sn, x_E}^O$  (respectively  $\mathcal{A}_{sn, x_I}^O, \mathcal{A}_{sg, x_E}^O, \mathcal{A}_{sg, x_I}^O$ )  
the corresponding open loop sensitivities.

Following the same lines as for the isolated pools (see [S1 Appendix](#), Sec 3), we  
express close loop sensitivities (sensitivities for the pools when they are coupled) as  
functions of open loop sensitivities (sensitivities for the pools with only their recurrent  
coupling).

Finally, we obtain:

$$\begin{pmatrix} \mathcal{A}_{sn, x_E} & \mathcal{A}_{sn, x_I} \\ \mathcal{A}_{sg, x_E} & \mathcal{A}_{sg, x_I} \end{pmatrix} = \frac{1}{1 + J_{nmda} J_{gaba} \mathcal{A}_{sn, x_E}^O \mathcal{A}_{sg, x_I}^O} \times \begin{bmatrix} \mathcal{A}_{sn, x_E}^O & -J_{gaba} \mathcal{A}_{sn, x_E}^O \mathcal{A}_{sg, x_I}^O \\ J_{nmda} \mathcal{A}_{sn, x_E}^O \mathcal{A}_{sg, x_I}^O & \mathcal{A}_{sg, x_I}^O \end{bmatrix} \quad (16)$$

As mentioned above, the main point here is to consider that open loop sensitivities  
have the same functional form as the isolated pool sensitivities, only they have to be  
evaluated at the fixed points  $sn^*$  and  $sg^*$  yielded by the coupled system, and considering

the total amount of external forcings (i.e. basic forcing  $x_E$  or  $x_I$  together with the forcing coming from the alternate pool:  $z_E = -J_{gaba}sg^* + x_E$  et  $z_I = J_{nmda}sn^* + x_I$ ).

Hence, we can then write the functional forms  $\Phi_{n_E}$ ,  $\Phi_{g_E}$ ,  $\Phi_{n_I}$  and  $\Phi_{g_I}$  of the sensitivities for an isolated region to explicitly express their dependencies :

$$\left\{ \begin{array}{lcl} \mathcal{A}_{sn,x_E}^O & = & \varphi_{n_E}(sn^*, -J_{gaba}sg^* + x_E) \\ \mathcal{A}_{sg,x_I}^O & = & \varphi_{g_I}(sg^*, J_{nmda}sn^* + x_I) \\ \Phi_{n_E}(\vec{s}^*, x_E, x_I) & = & \frac{\mathcal{A}_{sn,x_E}^O}{1 + J_{nmda}J_{gaba}\mathcal{A}_{sn,x_E}^O\mathcal{A}_{sg,x_I}^O} \\ \Phi_{g_E}(\vec{s}^*, x_E, x_I) & = & \frac{J_{nmda}\mathcal{A}_{sn,x_E}^O\mathcal{A}_{sg,x_I}^O}{1 + J_{nmda}J_{gaba}\mathcal{A}_{sn,x_E}^O\mathcal{A}_{sg,x_I}^O} \\ \Phi_{n_I}(\vec{s}^*, x_E, x_I) & = & \frac{-J_{gaba}\mathcal{A}_{sn,x_E}^O\mathcal{A}_{sg,x_I}^O}{1 + J_{nmda}J_{gaba}\mathcal{A}_{sn,x_E}^O\mathcal{A}_{sg,x_I}^O} \\ \Phi_{g_I}(\vec{s}^*, x_E, x_I) & = & \frac{\mathcal{A}_{sg,x_I}^O}{1 + J_{nmda}J_{gaba}\mathcal{A}_{sn,x_E}^O\mathcal{A}_{sg,x_I}^O} \end{array} \right. \quad (17)$$

where  $\varphi_{n_E}$  and  $\varphi_{g_I}$  are given by Eq 14 and 15.

### 2.3 At two-area system level

We now turn to the coupling between two areas (see Fig 3(C)).

In the same way that we have built isolated area sensitivities as functions of isolated pool sensitivities, we will identify open loop sensitivities of areas in the two-area system with their isolated area sensitivities counterparts, so the loop to be opened is now the connections between the two areas.

In a first step, we have built, in full generality for any two-area system, the expression of the matrix of sensitivities to a perturbation upon the input current in one of the four pools, and they are expressed as functions of the sensitivities in the single-area system (S1 Appendix Sec 3.3).

### 2.4 Control-Target system

In a second step, we use this general expression to address the question of how sensitivities would drive the response of the excitatory pool of one area to the activation of the excitatory pool of the other one, depending on the connectivity between the two areas. Hence, we attribute a role to each area: the area receiving a perturbed input current will be called "Control" area (denoted by C), and the other area will be called "Target area" (denoted by T). Since we are interested in understanding how BOLD signals can be affected by connectivity, we will focus on the expression of the perturbations of their excitatory pools  $\delta sn_C$  and  $\delta sn_T$  (which drive changes in BOLD signals) in response to a perturbation  $\delta B_{EC}$  applied to the input current of Control Area.

We obtain (S1 Appendix Sec 3.3):

$$\begin{pmatrix} \delta sn_C \\ \delta sn_T \end{pmatrix} = \frac{1}{1 - G_{CT}G_{TC}} \begin{pmatrix} \mathcal{A}_{sn_C,x_{EC}} \\ G_{TC}\mathcal{A}_{sn_C,x_{EC}} \end{pmatrix} \delta B_{EC} \quad (18)$$

with

$$\begin{cases} G_{CT} = \mathcal{A}_{sn_C, x_{E_C}} k_{E_{CT}} \kappa_{CT} + \mathcal{A}_{sn_C, x_{I_C}} (1 - k_{E_{CT}}) \kappa_{CT} \\ G_{TC} = \mathcal{A}_{sn_T, x_{E_T}} k_{E_{TC}} \kappa_{TC} + \mathcal{A}_{sn_T, x_{I_T}} (1 - k_{E_{TC}}) \kappa_{TC} \end{cases} \quad (19)$$

Written in the canonical form, we then have for the Control area:

$$(1 - G_{CT}G_{TC})\delta sn_C = \mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C} \quad (20)$$

so we can define  $g_{sn_C, B_{E_C}} = G_{CT}G_{TC}$  the feedback gain for  $\delta B_{E_C}$  acting on  $\delta sn_C$  through the feedback loop  $\{(C, T), (T, C)\}$ .

For the Target area, we have:

$$(1 - G_{CT}G_{TC})\delta sn_T = G_{TC} \mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C} \quad (21)$$

so that the feedback gain  $g_{sn_T, B_{E_C}}$  acting on  $\delta sn_T$  is the same, namely:

$$\begin{aligned} g_{sn_C|T, B_{E_C}} &= (\mathcal{A}_{sn_C, x_{E_C}} k_{E_{CT}} \kappa_{CT} + \mathcal{A}_{sn_C, x_{I_C}} (1 - k_{E_{CT}}) \kappa_{CT}) \\ &\times (\mathcal{A}_{sn_T, x_{E_T}} k_{E_{TC}} \kappa_{TC} + \mathcal{A}_{sn_T, x_{I_T}} (1 - k_{E_{TC}}) \kappa_{TC}) \end{aligned} \quad (22)$$

## 2.5 Effect of connectivity upon Target sensitivity

So far, we have derived the formal expressions for sensitivities and gains for any connectivity, i.e. for any values of  $k_{E_{ij}}$  that governs how excitatory and inhibitory pools are connected between areas. We will now consider archetypal possibilities by assigning binary values to  $k_{E_{CT}}$  and  $k_{E_{TC}}$ .

Considering that (long range) connexions always originate from excitatory pools, we then have four possibilities, depending for each area on whether it receives excitatory signal upon its excitatory pool or its inhibitory one (see Fig 4). We can then produce predictions on system behavior for the schemes of connectivity representing inhibitory control models: Target-to-Control inhibition (I-E), mutual inhibition (I-I) or Control-to-Target inhibition (E-I), w.r.t. the "model of reference" which is mutual excitation (E-E).

We convene to denote the connectivity by  $\square - \square$  where  $\square$  denotes the reception pools ( $E$  or  $I$ ) for the Control and Target areas respectively.

Accordingly, we will denote the corresponding feedback gain  $g_{sn_T, B_{E_C}}$  as:  $g_{\square - \square}$ .

We can then write a generic expression of feedback gains and sensitivities for any connectivity:

$$\begin{cases} g_{\square - \square} = \mathcal{A}_{sn_C, x_{\square_C}} \kappa_{CT} \mathcal{A}_{sn_T, x_{\square_T}} \kappa_{TC} \\ \delta sn_C = \frac{\mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C}}{1 - g_{\square - \square}} \\ \delta sn_T = \frac{\mathcal{A}_{sn_T, x_{\square_T}} \kappa_{TC} \mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C}}{1 - g_{\square - \square}} = \mathcal{A}_{sn_T, x_{\square_T}} \kappa_{TC} \delta sn_C \end{cases} \quad (23)$$

yielding the sensitivities given in table 1 for each connectivity.

From this, we can now predict the signs of sensitivities:

- For  $g < 1$ , since  $\mathcal{A}_{sn_C, x_{E_C}} > 0$ , we always have  $\delta sn_C > 0$ .
- Sensitivities for Target area are given by:

|                     |                                     |               |
|---------------------|-------------------------------------|---------------|
| $\square - \square$ | $\mathcal{A}_{sn_T, x_{\square_T}}$ | $\delta sn_T$ |
| $E - E$             | $\mathcal{A}_{sn_T, x_{E_T}} > 0$   | $> 0$         |
| $I - E$             | $\mathcal{A}_{sn_T, x_{E_T}} > 0$   | $> 0$         |
| $I - I$             | $\mathcal{A}_{sn_T, x_{I_T}} < 0$   | $< 0$         |
| $E - I$             | $\mathcal{A}_{sn_T, x_{I_T}} < 0$   | $< 0$         |

310

We conclude that only I-I or E-I connectivity could translate upregulation in Control area into downregulation in Target area.

311

312

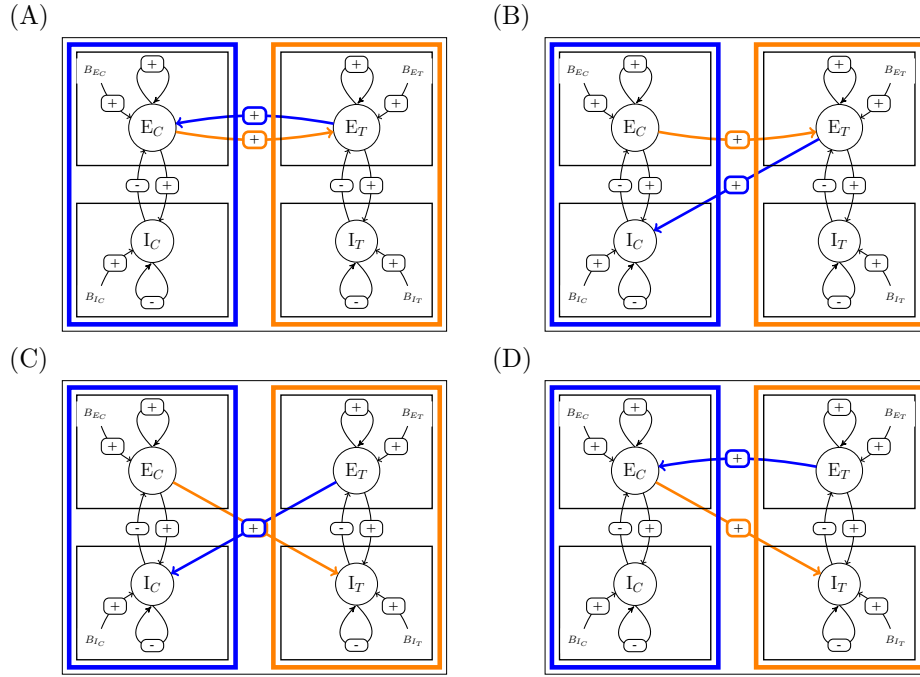


**Fig 4.** Connectivity that are considered. (A) Mutual excitation (E-E), (B) Target-to-Control inhibition (I-E), (C) Mutual inhibition (I-I), (D) Control-to-Target inhibition (E-I).

| $\square - \square$ | $g_{\square-\square}$                                                             | $\delta sn_C$                                                    | $\delta sn_T$                                         |
|---------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------|-------------------------------------------------------|
| $E - E$             | $\mathcal{A}_{sn_C, x_{E_C}} \kappa_{CT} \mathcal{A}_{sn_T, x_{E_T}} \kappa_{TC}$ | $\frac{\mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C}}{1 - g_{E-E}}$ | $\mathcal{A}_{sn_T, x_{E_T}} \kappa_{TC} \delta sn_C$ |
| $I - E$             | $\mathcal{A}_{sn_C, x_{I_C}} \kappa_{CT} \mathcal{A}_{sn_T, x_{E_T}} \kappa_{TC}$ | $\frac{\mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C}}{1 - g_{I-E}}$ | $\mathcal{A}_{sn_T, x_{E_T}} \kappa_{TC} \delta sn_C$ |
| $I - I$             | $\mathcal{A}_{sn_C, x_{I_C}} \kappa_{CT} \mathcal{A}_{sn_T, x_{I_T}} \kappa_{TC}$ | $\frac{\mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C}}{1 - g_{I-I}}$ | $\mathcal{A}_{sn_T, x_{I_T}} \kappa_{TC} \delta sn_C$ |
| $E - I$             | $\mathcal{A}_{sn_C, x_{E_C}} \kappa_{CT} \mathcal{A}_{sn_T, x_{I_T}} \kappa_{TC}$ | $\frac{\mathcal{A}_{sn_C, x_{E_C}} \delta B_{E_C}}{1 - g_{E-I}}$ | $\mathcal{A}_{sn_T, x_{I_T}} \kappa_{TC} \delta sn_C$ |

**Table 1.** Sensitivities for  $sn$  in Control and Target area to perturbation upon  $B_{E_C}$ , the input current into Control Area.

### 3 Interpretation and discussion

#### 3.1 Effect of self-coupling in pools

To illustrate the behavior of the basic unit of the system, the effect of incoming input  $z_E$  upon the activity level  $sn^*$  in an isolated excitatory pool is illustrated in Fig 5 for increasing values of its self-excitatory coupling parameter  $W_+$ .

As  $z_E$  increases, fixed points  $sn^*$  increase monotonically up to the saturating value 1 (Fig 5 (A)). With the set of parameters given in [27] (S2 Appendix), we observe a threshold effect, with an absence of reaction to input values lower than about  $z_E \simeq 0.3$  nA. This threshold effect is due to the filtering effect in the translation from input to  $sn$  through fire rate  $rn$  (Eq 2, Fig 5 (B)), so that if  $rn^* = 0$  in Eq 5, so must be  $sn^*$ .

Beyond this threshold, fixed points values increase strongly with small increments of  $z_E$ , such that the range of relevant input values is pretty narrow. If we consider a firing rate at 100 Hz as the reasonable maximal value for a highly activated pool, this range would span [0.3; 0.75] nA in absence of self-excitatory loop ( $W_+ = 0$ ) down to [0.3; 0.6] nA for  $W_+ = 1$ .

This high reactivity to input is explained by the feedback gain (Fig 5 (C)), which is positive by construction and can get strong values as the self-excitatory coupling parameter  $W_+$  is increased. When  $W_+ = 0$ , the gain is analytically null (blue curves), and the corresponding sensitivity is the open loop one (the lowest one in Fig 5 (D)). As  $W_+$  increases, the gain remains null for  $sn^* = 0$  corresponding to  $z_E < 0.3$  (since there is no input for self-amplification), and also saturates towards 0 for  $sn^*$  saturating to 1 (since there is no more possible gain). In between, the gain reaches a maximal value in the range of  $z_E$  which has a maximal impact upon  $sn^*$ . The close loop sensitivity is then scaled accordingly.

The effect of incoming input  $z_I$  upon the activity level  $sg^*$  in an isolated inhibitory pool is illustrated in Fig 6 for increasing values of its self-inhibitory coupling parameter  $J_-$ . As expected, increasing  $z_I$  increases fixed points  $sg^*$  values monotonically, but here, at a far lower rate than in excitatory pool (Fig 6 (A)). Since the self-coupling is inhibitory, the feedback gain is negative (Fig 6 (C)), hence the sensitivity is the highest in absence of it (blue curves, Fig 6 (D)). As a consequence, the self-inhibitory coupling will extend the range of relevant input values  $z_I$  for firing rates below 100 Hz (Fig 6 (B)), from about [0.2; 0.45] nA to about [0.2; 0.9] nA.

#### 3.2 Effect of coupling two pools

On Figs 5(A) and 6(A), dashed lines report the values representing *basic external inputs*  $x_E$  and  $x_I$  when the pools are connected to form an area, and that are given in [27]. From their location in the range of relevant values, we can anticipate the role of the inhibitory pool upon the basic level of activity in one isolated area: since  $x_I$  is close to the threshold, it will be up-regulated by the excitatory input from the excitatory pool, so that its induced inhibitory input backward to the excitatory pool will actually lower  $z_E$  towards values near the threshold. Namely, with the value of  $x_E$  prescribed by [27], the excitatory pool *needs* to be down-regulated to have reasonable excitation level at the resting state.

This effect of coupling the two pools is illustrated in Fig 7, where we submit the area to varied values of  $x_E$  or  $x_I$  around the values of references (vertical dashed lines) and for a range of  $J_{gaba}$  which controls the strength of feedback inhibition from inhibitory pool to excitatory pool upon activation of the latter (this feedback inhibition then acts as a self-inhibition at the area level).

We report on the left column how varying  $x_E$ , while keeping  $x_I$  at the reference value, affects the activity level of the excitatory pool. In this case, increasing  $x_E$  will, on the one hand, directly enhance activity level in the excitatory pool, on the other hand, this enhanced activity level will in turn enhance activity level in the inhibitory pool (by feedforward activation), which will in turn exert a feedback inhibition effect upon the excitatory pool. All feedbacks taken into account (self-excitation and feedback inhibition), the sensitivity  $\mathcal{A}_{sn, x_E}$  remains positive (Fig 7 (A)), albeit with a responsiveness that can be far smoother than for the isolated pool.

We first note that sensitivities have here again bell-shape curves. Indeed, for lower values of external input ( $x_E < 0.25$ ),  $sn^*$  values would slowly tend to 0, whereas for too high values of external input ( $x_E > 0.6$ ) it would slowly saturate to 1 (Fig 7B). In between, sensitivities must have a peak value.

The responsiveness to external inputs depends upon the strength of the self-inhibitory control at the area level, which is governed by  $J_{gaba}$ . For lower values of  $J_{gaba}$ , the responsiveness is steep, and the range of operational values of total input remains narrow. As a matter of fact, in the case  $J_{gaba} = 0$ , we recover the sensitivity of the isolated excitatory pool (Eq 14) for the high value  $W_+ = 1.4$  that is prescribed when coupling pools in [27], and we observe a divergence of the sensitivity around  $x_E \simeq 0.33$ , which translates into a jump for  $sn$  in Fig 7 (B).

This means that, in absence of a sufficient down-regulation of the excitatory pool by the feedback inhibition, the slightest variation in total input could translate into an all-or-nothing response, which is not a property we would consider as reasonable for a neural pool. This divergence appears no more for  $J_{gaba}$  greater than 0.5 nA.

As  $J_{gaba}$  value becomes larger and larger, the responsiveness is smoother and smoother as the peak of sensitivity to inputs tends to decrease, and the operational range of input values increases. Since sensitivities can be regarded as the derivatives of  $sn$  with respect to the input, their bell-shaped curves translate into *sigmoidal* shapes when considering  $sn$  (Fig 7(B)) and the enlargement of the operational range translates into a response of the pool that becomes gradual.

Depending on  $J_{gaba}$ , the  $sn$  level at the value of reference for  $x_E$  spreads over quite large a range (from about 0.6 for  $J_{gaba} = 0.75$  down to about 0.05 for  $J_{gaba} = 2.25$ ). This spread in the excitatory pool  $sn$  is however largely reduced when considering the total effective input  $xn$ , as it is tempered by the negative contribution from the inhibitory input. For instance, in the case  $J_{gaba} = 0.75$  where  $sn \simeq 0.6$ , the total input current is about  $xn \simeq 0.45$  (Fig 7(C)) and would translate into a limited firing rate ( $rn \simeq 16$  Hz whereas it would be  $rn = 45$  Hz for  $J_{gaba} = 0$ ).

Actually, for large enough a value for  $J_{gaba}$  (around 1.75nA), inhibitory feedback control at the area level would compensate exactly for self-excitatory feedback in the excitatory pool (Fig 7(C), where dotted line corresponds to  $xn = x_E$ ). With such a high value of  $J_{gaba}$ , an external input  $x_E = 0.75$  nA would be fully counter-balanced and yield a firing rate at 100 Hz, as if in an isolated excitatory pool with no self-excitation ( $W_+ = 0$ , blue curve in Fig 5(B)).

On the right column of Fig 7, we report the symmetrical effect of varying  $x_I$  while keeping  $x_E$  at the value of reference. The sensitivity  $\mathcal{A}_{sn, x_I}$  of the activity ( $sn$ ) in the excitatory pool with regards to positive perturbation upon input current into the inhibitory pool is, as expected, negative (Fig 5(D)) since increasing input current into the inhibitory pool will lower the input current into the excitatory pool (Fig 5(F)). We note that in absence of external input current into the inhibitory pool ( $x_I = 0$ ), we have a almost null activity in the inhibitory pool (i.e.  $sg = 0$ , not shown), meaning that the forward excitation from the excitatory pool is not enough *per se* to activate it (with the current set of parameters) so that we recover the high stationary state of the isolated excitatory pool ( $sn^* = 0.8$ ).



On the other hand, with too strong an activation of the inhibitory pool ( $x_I > 0.4$ ),  $sn^*$  would be flattened to 0 and lose any responsiveness.

In between, the shape of sensitivities curves are poorly affected by  $J_{gaba}$  (for  $J_{gaba} \geq 0.75$ ) so that the range of operational values for  $x_I$  remains about the same, yet with a median value which is lower and lower as  $J_{gaba}$  is increased. As shown in Fig 7(F), an important point is that, with  $J_{gaba} \geq 1.25$  and no additional input into the excitatory pool ( $x_E$  set at its value of reference  $x_{E,ref} = 0.382$  nA), any additional input current to the inhibitory pool with regards to its own value of reference for  $x_I$  (vertical dashed line) will translate into a decrease of the input current into the excitatory pool (i.e.  $xn < x_{E,ref}$ , horizontal dashed line).

To summarize, we will point out at four main points:

1. The sensitivities to positive perturbation of input current to either excitatory ( $\mathcal{A}_{sn,x_E}$ ) or inhibitory pool ( $\mathcal{A}_{sn,x_I}$ ) are respectively positive and negative, as predicted by the study of isolated pools. This point will be of importance for the interpretation of the Control-Target system behaviour.
2. At state of reference (i.e. setting  $x_E$  and  $x_I$  at their values of reference representing *basic inputs* with no additional input from other areas), the stationary values of the excitatory pool ( $sn^*$  and  $rn^*$ ) can be directly adjusted by tuning  $J_{gaba}$ .
3. With the set of parameters prescribed by [27], the responsiveness of the area can be steep when the inhibitory feedback, driven by  $J_{gaba}$  from the inhibitory pool to the excitatory pool upon positive perturbation of the latter, is too low. Moreover, even with a high strength of the self-inhibitory feedback loop within the area, the operational range for further inputs is not so large: with the given values of references, higher values of  $J_{gaba}$  would at most allow a doubling of input current with regards to basic input current.
4. Given the structure of the model, activation of inhibitory pool by external input can either translate into a lower excitation of the excitatory pool (a brake on self-amplification), or into a inhibition of the excitatory pool (it more than compensates its self-excitation), depending upon how the excitation of the inhibitory pool can have an effect upon total current input into the excitatory pool through  $J_{gaba}$ . A crossing value between the two behaviours appears to be around  $J_{gaba} = 1.25$ .

### 3.3 Effect of connectivity upon sensitivities in Control-Target system

We now turn to illustrations of the Control-Target system's behaviour, depending upon the connectivity, as described in Table 1. In all cases,  $J_{gaba,C}$  in the Control area is set such that a) the firing rate  $rn_C = 3$  Hz for the input current of reference (namely,  $J_{gaba,C,ref} = 1.758$  nA in E-E and E-I connectivities, and  $J_{gaba,C,ref} = 0.929$  nA in I-E and I-I connectivities), and b) input current into inhibitory pool is set to the value of reference in both areas. Then input current into the Control area is varied, for a range of  $J_{gaba,T}$  in the Target area.

In E-E connectivity (mutual excitation, Fig 8), the excitatory pools are in a mutual excitation regime (Fig 8 (A)) so that both sensitivities to perturbation upon  $B_{E_C}$  are positive (Fig 8 (C,D)), and Target activity can only increase upon Control excitation. Moreover, the feedback gain associated with the inter-area connection is positive (Fig 8 (B)). Upon stimulation of Control, this positive feedback will then amplify the effect of

the stimulation, and stabilize the activity levels of both areas at higher levels than in absence of mutual excitation. The capacity of Control area to increase Target activity is modulated by the value set for  $J_{gaba,T}$  (Fig 8 (E,F)). For a low value of  $J_{gaba,T}$  (blue curve in Fig 8 (F)), the self-inhibition at Target area level is less operative so that excitation of Control can translate in higher activity in Target (here illustrated by firing rates). We note however that the capacity to increase Target activity can be limited to narrow ranges of  $B_{EC}$ . For example, for  $J_{gaba,T} = 1.75$  nA, the effective range of control is about  $[0.35, 0.425]$ , i.e. a 20% increase, to take firing rate in Target from about 2 Hz to about 15 Hz. Beyond this interval, further increase of Control will have poor effect (e.g. doubling the firing rate in Control, from 25 Hz to 50 Hz, will translate into an additional 5 Hz in Target). Strikingly, sensitivities peak values, as core values of range for effective control, are quite disparate, so that control efficacy would highly depend upon the value set for basic input current into Control area, which should become a function of  $J_{gaba,T}$ .

In I-E connectivity (Target-to-Control inhibition, Fig 9), the Control area has an excitatory effect upon Target area, which in turn has an inhibitory effect upon Control area. Here again, both sensitivities to perturbation upon  $B_{EC}$  are positive (Fig 11 (C,D)), and Target activity can only increase upon Control excitation. However, the feedback gain associated with the inter-area connection becomes negative (Fig 11 (B)) so that the effect of the Control stimulation will be dampened in both areas, resulting in activity levels lower than in absence of the feedback loop. The capacity of Control area to increase Target activity is still modulated by the value set for  $J_{gaba,T}$  (Fig 11 (E,F)), but, by contrast with E-E connectivity, the range of effective control is enlarged and become pretty similar across  $J_{gaba,T}$  values (Fig 11 (C,D)), so that changing  $J_{gaba,T}$  mainly regulates the range for firing rates in presence versus in absence of Control activation.

Direct projection of Control area to inhibitory pool of Target area (I-I and E-I connectivity) completely changes the picture: activating Control can now *decrease* activity in Target area, because sensitivities of  $sn_T$  to perturbation upon  $B_{EC}$  become negative in both cases.

In I-I connectivity (mutual inhibition, Fig 10), the feedback gain is positive and sensitivities can reach high values and span narrow ranges of  $B_{E,C}$ , as in E-E connectivity (note that the case  $J_{gaba,T} = 0.75$  nA yields a jump for  $B_{E,C}$  close to 0.4 nA and can be disregarded), here again yielding narrow ranges for control. For instance, for the condition ensuring resting state at 3 Hz in both areas for basic input current set to their value of reference ( $J_{gaba,T} = J_{gaba,C} = 0.929 \simeq 1$  nA, hence near the green curve in Fig 10F), a slight activation of the Control area by  $\Delta B_{E,C} = 0.425 - x_{E,ref} = 0.043$  nA, i.e. a small 10% increase, would be sufficient to shut the activity of Target down to 0.5 Hz.

In E-I connectivity (Control-to-Target inhibition, Fig 11), by contrast, sensitivities of Target to activation of Control area span the full range considered for  $B_{E,C}$  and their peak values are quite lower (in absolute value), so that effect of Control is smoother and can be nicely gradual (as in I-E connectivity). For instance, for the 3 Hz condition ( $J_{gaba,C} = 1.758$  and  $J_{gaba,T} = 0.929$ , green curves), there would need now about  $\Delta B_{E,C} = 0.5 - x_{E,ref} = 0.118$  nA in Control to get 0.5 Hz in the Target, i.e. a 30% increase.

Note that in both cases however, the 10% increase in I-I and the 30% increase in E-I, the activity level in Control area would shift from 3 Hz to about the same 30 Hz.

To summarize, there are here two main points:

1. In mutual excitation E-E, the upregulation of Control is amplified by the feedback loop in both areas; in Target-to-Control inhibition I-E, the upregulation of

Control is dampened in both areas; in mutual inhibition I-I, the feedback loop amplifies the downregulation of Target triggered by upregulation of Control, and in Target-to-Control inhibition E-I, it is dampened,

2. In I-E and E-I connectivity (Fig 17A, B), control effectivity is more gradual with regards to  $J_{gaba,T}$  than in E-E and I-I (Fig 17C, D).

### 3.4 Consequences for system responses upon step activation of Control area

So far, we have analyzed the response of the system in terms of sensitivities, namely how  $Sn_C$  and  $Sn_T$  would be affected by an infinitesimal (positive) perturbation applied upon the excitatory forcing of the excitatory pool of Control area. Formal expression of sensitivities are the only path to derive analytical results that allow to understand the dynamical behavior of the system in full generality.

We now turn to illustrations that extend this analysis by direct comparison between states of the system at rest versus when submitted to a finite step of activation, as it is a typical case in macroscopic imagery such as fMRI, where the BOLD signal is recorded during tasks like Go-NoGo or Think-NoThink in inhibitory control studies [29].

To make the link with previous sections, we outline, as a first point, that the response of Target area to a finite step of activation of Control area can be read as an integration over the interval for the forcing of Control area of the sensitivities we proved analytically for the activity of the excitatory pool of Target. We can then predict that, for any set of parameters values, integrating always positive or always negative sensitivities depending on the kind of connectivity, will translate into the same behavioral response of Target area: E-E and E-I connectivities will trigger Target up-regulation, and only I-E and I-I connectivities can account for inhibitory control, where Target activity is down-regulated by activation of Control activity.

#### 3.4.1 Illustrations with BOLD signals

In Fig 12, we first illustrate the system dynamical response for E-E (left) and E-I (right) connectivities, when submitted to a step stimulation of 0.1 nA upon the excitatory pool of Control area. We report the behavior of synaptic activities  $sn$  and of the BOLD signal in both areas, when submitted to a fictitious experiment with a 20s long stimulation (upper row) or a more realistic experiment with four 3s long stimulations interspaced by random delay (lower row).

As expected from Fig. 2, the dynamical response of the synaptic activity to stimulus onset is step-like in all case, and, as expected from our sensitivities analysis, the Target responds by a decrease of activity (both in synaptic activity, and then in BOLD signal) in E-I connectivity.

The long stimulation gives an idea of the characteristic times of the dynamics. Contrasting to synaptic activities, the BOLD signal needs about 15s to reach its stimulus-induced steady state value and about the same delay to return to its rest value after the stimulus stop. If, in both connectivities, synaptic activities adjust to stimulus onset within less than a second, their relaxation back to resting state after stimulus stops is longer in E-E connectivity (about 6s) than in E-I connectivity (less than a second).

The longer delay in E-E connectivity is due to the positive feedback gain between the two areas: if we focus for instance on Control area, even after the stimulus has stopped, there is still some overactivation (w.r.t. basic forcing) due to stimulation by Target area, which in turn is sustained by the remaining activity level of Control area (and vice versa). Here, the positive feedback gain acts as a brake on relaxation.

By contrast, in E-I connectivity, the high activity level of Control area tends to decrease Target activity (w.r.t. resting state) so that there is an under-activation of Control area (even if less and less as relaxation proceeds) so that nothing slows down the return to resting state (a negative feedback gain would not act as a brake).

As a consequence of its long characteristic time, the BOLD signal does not have time enough to reach its steady state value in the case of a series of short stimulations: it oscillates around values slightly below the steady state value observed for the long stimulation. In such stimulation series pattern, the time-shifted bijectivity between synaptic activity and BOLD signal is lost.

Still, considering the short time adjustment of synaptic activities levels to stimulus onset, it remains relevant to consider finite differences between synaptic activities at stimulated state vs. non stimulated state as a valuable observable of system response, e.g. in Fig. 12, about  $\Delta Sn_C = (0.8 - 0.2) = 0.6$  and  $\Delta Sn_T = (0.6 - 0.2) = 0.4$  are a good identification of the response by the system to a stimulus of 0.1 nA in Control area in an E-E connectivity, once all feedback loops are taken into account...

### 3.4.2 Effect of self-inhibitory strength within Target area depending on connectivity

To understand how parameters can drive system dynamics through such feedback loops, we need to figure out the relationship between connectivity and the parameter  $J_{gaba_T}$  which controls the self-inhibition within target area. For this, we chose to look at the cross-effect between the self-inhibitory activity of the target region and the ability of the control region to have an effect upon the target region. In doing this exploration, we came to a far from intuitive conclusion: the strength of self-inhibition in the target region has a non-monotonic effect with respect to the response to activation, both in target area and in the control one (through feedback loop).

We find this effect interesting enough to expose it for each connectivity, in Figs 13 to 16. In each figure, on the left column, we report how the fixed point values of  $Sn_C$  and  $Sn_T$  are affected when a step activation is applied upon Control area (for different values of the step  $I_{stim}$ , coded by colors), taking into account all feedback loops, and for a range of effective values for  $J_{gaba_T}$ . On the right columns are reported the discrete differences between these stimulated fixed point values and the non-stimulated fixed point value, for each value of  $J_{gaba_T}$ , namely  $\Delta Sn_C$  and  $\Delta Sn_T$  as exemplified in Sec 3.4.1. For instance, in E-E connectivity (Fig 13), for the value of reference  $J_{gaba_T} = 1.758$  nA, the corresponding  $\Delta Sn_C, \Delta Sn_T$  are those illustrated in Fig 12A (about 0.6 and 0.4 respectively for  $I_{stim} = 0.1$  nA).

The first point to observe is that applying the same step activation (e.g.  $I_{stim} = 0.1$  nA, brown lines in all graphics) not only will affect differently  $\Delta Sn_T$  depending on  $J_{gaba_T}$ , but will also affect the corresponding change  $\Delta Sn_C$  in Control Area, as a result of feedback loops.

Indeed, both  $\Delta$  are to be understood as the contrast between resting values and activated values, and a major point is that changing  $J_{gaba_T}$  not only affects stimulated states but also affects resting values. Hence, changes are driven by two factors: activity levels before stimulation is applied, and the margin for increase or decrease.

In E-E connectivity (Fig. 13), in low range of  $J_{gaba_T}$ , self-inhibition in Target area is poor, such that its unstimulated activity is quite high (above 0.8, Fig. 13C), so that a) its margin for increase is about null which translates into  $\Delta Sn_T$  tending to zero (Fig. 13D), and b) its excitatory feedback to Control area is high (with an unstimulated activity level above 0.6, Fig. 13A) which in turn lets to Control area a limited margin to increase upon stimulation (Fig 13B). In the high range of  $J_{gaba_T}$ , the strong

self-inhibition of Target leads to a very low unstimulated activity, which opens a larger margin for increase upon excitation by Control area, but in the same time, the higher  $J_{gaba_T}$ , the less sensitive it must become to activation by stimulated Control (as shown by sensitivities Fig 8D, with  $BE_C$  near 0.5) so that  $\Delta Sn_T$  should also tends to zero. In between these two limits,  $\Delta Sn_T$  can not but cross a peak value, corresponding to the highest contrast between unstimulated and stimulated fixed points. As suggested by sensitivities analysis (Fig 8B,C,D) that present sharper functions around unstimulated inputs, this peak value is mainly driven by the alteration of fixed points in absence of stimulation (blue curves), whereas upon stimulation, the stimulation becomes dominant, and the effect of feedback loops operates on a smaller range (as exemplified by the tiny change of Control activity, Fig 13A, brown curve). The same lines of interpretation apply to I-E connectivity (Fig 9 and 14), although again this is a more flexible control scheme than E-E connectivity.

As a summary, we point out that the very same stimulation upon Control (e.g.  $I_{stim} = 0.1$ ) can translate into the same change of activity level in Target (and hence corresponding changes in BOLD signal), despite very different values of self-inhibitory strength of the latter (e.g.  $\Delta Sn_T \simeq 0.2$  for  $J_{gaba_T} \simeq 1.5$  or  $J_{gaba_T} \simeq 3$  in Fig 13D). In other words, a further increase in Control activity does not imply a further increase in Target activity: below the peak value, it does, beyond the peak value, it does not.

To make the link with previous sections, we outline, as a second point, that the finite-size response of Target area has to be understood as the integration of its sensitivity to perturbation of forcing in Control area within a interval starting from the input value of reference (dotted line in sensitivities figures). Hence, one can read from these figures that those integrals indeed vary non monotonically depending upon  $J_{gaba_T}$  values.

In the connectivities allowing inhibitory control (E-I and I-I), we observe the same kind of non monotonic effect of  $J_{gaba_T}$  upon Control and Target change of activity levels when Control Area is stimulated, yet with a now negative change of Target activity (Figs 15 and 16). In both connectivities, the peak values of  $\Delta Sn_T$ , and the range for effective control, correspond however to lower values of  $J_{gaba_T}$  than for E-E and E-I.

In E-I connectivity, activity in Control area translates into a direct inhibition of Target, which adds up to the internal self-inhibition, so that higher values of  $J_{gaba_T}$  lead to complete extinction of Target activity, even in absence of stimulation of Control (Fig 15C). Adding stimulation of Control area only amplifies the trend. In terms of effect of the stimulation of Control upon change in Target activity, we have then that the highest change in Control  $\Delta Sn_C$  (Fig 15B, brown curve) corresponds to a null change in Target (Fig 15D, brown curve).

In I-I connectivity, as in E-E connectivity, higher values of  $J_{gaba_T}$ , lowering Target activity by self-inhibition, let the inhibitory effect of Target upon Control relax, so that Control activity can be higher, even in absence of stimulation (Fig. 16A, blue curve). As a consequence, margin for an increased Control activity upon stimulation is limited and change in Control activity is lowered (Fig. 16A, brown curve). As a consequence, in contrast with E-I connectivity, here, the highest change in Control  $\Delta Sn_C$  (Fig 16B, brown curve) corresponds to the highest change in Target  $\Delta Sn_T$  (Fig 16D).

As a summary for E-I and I-I connectivities (allowing inhibitory control), we point out that a) they correspond to about the same effective ranges of  $J_{gaba_T}$ , that are narrower and quite lower than for the other two connectivities and b) they fully contrast with each other in terms of the relationship between peak changes in areas activity upon stimulation.

Overall, all these results show the major role of the within-area self-inhibition in Target area, depending on the connectivity. If tuning  $J_{gaba_T}$  for an effective control is

to be considered (considering as a proxy the “finite-size” sensitivity of Target in response to stimulation of Control area, as we did here), this would lead to quite different ranges of optimal values to be considered (Fig 17).

## 4 Conclusion and perspectives

In this study, our aim was to understand how brain regions can be connected at neural level so that, in a Target area, a decrease in BOLD activity would be observed when a Control area is subjected to an increase in BOLD activity. For the neural model, we took inspiration from the model by Naskar et al. [27], which describes the mean synaptic activities, both excitatory and inhibitory within each area. More specifically, we used the same differential equations and set of neurobiological parameters, including their proposed range of values. Their model includes an implicit type of connectivity which we termed E-E. For the coupling from neural state to BOLD signal, we used the Extended Balloon-Windkessel model as proposed by [30]. When this composite model (neural and BOLD) was studied as a dynamical model, it was observed to reach, quite rapidly, a stationary neural state (also called a fixed point, FP), as opposed to the BOLD signal which follows the fluctuations in neural signal with a certain delay and would hardly reach a FP. We have therefore focused on this neural model in its stationary state, as a first step. Using this as a foundation, we built a sensitivity analysis in the neighbourhood of the FP using two brain areas as a configuration model, more specifically, a control area and a target area. Using this two-areas model, we built a hierarchical system of nested sensitivities and we have shown that:

- To have a decrease in synaptic activity in a Target area, we needed to consider other types of connectivity than E-E, as it is unable to cause an inhibitory response. Four different types of connectivities have been explored, and we show that only E-I and I-I connectivities can reduce the synaptic activity in the Target region, with possibly E-I allowing a smoother control.
- The E-I and I-I connectivities, which both enable inhibitory control, contrast with each other in terms of the correlation between changes in activity in control and target areas.
- The intra-area self-inhibitory coefficient  $J_{gaba_T}$  in the Target area modulates this decrease in synaptic activity when the Control region is stimulated. By analyzing the sensitivity of a finite perturbation of the external forcing in the Control area as a function of  $J_{gaba_T}$  in the Target area, we observed an unanticipated effect of  $J_{gaba_T}$  beyond a certain threshold: a further increase of stimulation of Control leads to a lower and lower of the inhibitory effect on Target area. Any “causal” viewpoint between activation of Control and inhibition of target should then be tempered, if feedback from Target to Control has to be considered.

As for our perspectives, we currently work on the generalization of the method for building nested sensitivities for more complex network configurations, e.g. a system with several control and target areas grouped into two sub-systems: a sub-system of Control areas and a sub-system of Target areas. Such an approach would yield an *analytical* access to sensitivities for some configurations proposed in the context of inhibitory control, such as the one by Mary et al. [29] which points out a network of 14 regions, comprising of 9 Control and 5 Target areas as a drive for the inhibitory control of intrusive memory in Post-Traumatic Stress Disorder. Furthermore, this generalization to a system-level with more areas would allow for the integration of relays in the network, as observed in recent studies [31]. Analytical sensitivities could also help



identifying areas that mostly influence information flow in any brain network, with no  
need to fall back to numerical experiments [32,33].

## 5 Supporting information

### 5.1 S1 Appendix.

Formal developments for sensitivities.

### 5.2 S2 Appendix.

Parameters values of reference as found in Naskar et al. [27].

## 6 Acknowledgments

## References

1. Sanzleon P, Knock SA, Woodman MM, Domide L, Mersmann J, McIntosh AR, et al. The virtual brain: A simulator of primate brain network dynamics. *Frontiers in Neuroinformatics*. 2013;7(MAY). doi:10.3389/fninf.2013.00010.
2. Sanz-Leon P, Knock SA, Spiegler A, Jirsa VK. Mathematical framework for large-scale brain network modeling in The Virtual Brain. *NeuroImage*. 2015;111:385–430. doi:10.1016/j.neuroimage.2015.01.002.
3. Amunts K, Defelipe J, Pennartz C, Destexhe A, Migliore M, Rylvlin P, et al. Linking Brain Structure, Activity, and Cognitive Function through Computation. *eNeuro*. 2022;9(2). doi:10.1523/ENEURO.0316-21.2022.
4. Schirner M, Domide L, Perdakis D, Triebkorn P, Stefanovski L, Pai R, et al. Brain simulation as a cloud service: The Virtual Brain on EBRAINS. *NeuroImage*. 2022;251(November 2021):118973. doi:10.1016/j.neuroimage.2022.118973.
5. Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, Meuli R, et al. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(6):2035–2040. doi:10.1073/pnas.0811168106.
6. Deco G, Jirsa VK. Ongoing cortical activity at rest: Criticality, multistability, and ghost attractors. *Journal of Neuroscience*. 2012;32(10):3366–3375. doi:10.1523/JNEUROSCI.2523-11.2012.
7. Deco G, Ponce-Alvarez A, Hagmann P, Romani GL, Mantini D, Corbetta M. How local excitation-inhibition ratio impacts the whole brain dynamics. *Journal of Neuroscience*. 2014;34(23):7886–7898. doi:10.1523/JNEUROSCI.5068-13.2014.
8. Hansen ECA, Battaglia D, Spiegler A, Deco G, Jirsa VK. Functional connectivity dynamics: Modeling the switching behavior of the resting state. *NeuroImage*. 2015;105:525–535. doi:10.1016/j.neuroimage.2014.11.001.
9. Castro S, El-Dereby W, Battaglia D, Orio P. Cortical ignition dynamics is tightly linked to the core organisation of the human connectome. *PLoS Computational Biology*. 2020;16(7):1–23. doi:10.1371/journal.pcbi.1007686.

10. Kobeleva X, Varoquaux G, Dagher A, Adhikari M, Grefkes C, Gilson M. Advancing brain network models to reconcile functional neuroimaging and clinical research. *NeuroImage: Clinical*. 2022;36(August 2021). doi:10.1016/j.nicl.2022.103262.
11. Sporns O, Tononi G, Kötter R. The human connectome: A structural description of the human brain. *PLoS Computational Biology*. 2005;1(4):0245–0251. doi:10.1371/journal.pcbi.0010042.
12. Bullmore E, Sporns O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*. 2009;10(3):186–198. doi:10.1038/nrn2575.
13. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*. 2010;52(3):1059–1069. doi:10.1016/j.neuroimage.2009.10.003.
14. Betzel RF, Avena-Koenigsberger A, Goñi J, He Y, de Reus MA, Griffa A, et al. Generative models of the human connectome. *NeuroImage*. 2016;124:1054–1064. doi:10.1016/j.neuroimage.2015.09.041.
15. Elam JS, Glasser MF, Harms MP, Sotiropoulos SN, Andersson JLR, Burgess GC, et al. The Human Connectome Project: A retrospective. *NeuroImage*. 2021;244(August):118543. doi:10.1016/j.neuroimage.2021.118543.
16. Munakata Y, Herd SA, Chatham CH, Depue BE, Banich MT, O'Reilly RC. A unified framework for inhibitory control. *Trends in Cognitive Sciences*. 2011;15(10):453–459. doi:10.1016/j.tics.2011.07.011.
17. Anderson MC, Hulbert JC. Active Forgetting: Adaptation of Memory by Prefrontal Control. *Annual Review of Psychology*. 2021;72:1–36. doi:10.1146/annurev-psych-072720-094140.
18. Apšvalka D, Ferreira CS, Schmitz TW, Rowe JB, Anderson MC. Dynamic targeting enables domain-general inhibitory control over action and thought by the prefrontal cortex. *Nature Communications*. 2022;13(1):1–21. doi:10.1038/s41467-021-27926-w.
19. Wessel JR, Anderson MC. Neural mechanisms of domain-general inhibitory control. *Trends in Cognitive Sciences*. 2024;28(2):124–143. doi:10.1016/j.tics.2023.09.008.
20. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*. 2014;18(4):177–185. doi:10.1016/J.TICS.2013.12.003.
21. Cunillera T, Brignani D, Cucurell D, Fuentemilla L, Miniussi C. The right inferior frontal cortex in response inhibition: A tDCS–ERP co-registration study. *NeuroImage*. 2016;140:66–75. doi:10.1016/j.neuroimage.2015.11.044.
22. Wong KF, Wang XJ. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*. 2006;26(4):1314–1328. doi:10.1523/JNEUROSCI.3733-05.2006.
23. Deco G, Ponce-Alvarez A, Mantini D, Romani GL, Hagmann P, Corbetta M. Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience*. 2013;33(27):11239–11252. doi:10.1523/JNEUROSCI.1091-13.2013.



24. Deco G, Jirsa VK, McIntosh AR. Resting brains never rest: Computational insights into potential cognitive architectures. *Trends in Neurosciences*. 2013;36(5):268–274. doi:10.1016/j.tins.2013.03.001.
25. Deco G, McIntosh AR, Shen K, Matthew Hutchison R, Menon RS, Everling S, et al. Identification of optimal structural connectivity using functional connectivity and neural modeling. *Journal of Neuroscience*. 2014;34(23):7910–7916. doi:10.1523/JNEUROSCI.4423-13.2014.
26. Glomb K, Ponce-Alvarez A, Gilson M, Ritter P, Deco G. Resting state networks in empirical and simulated dynamic functional connectivity. *NeuroImage*. 2017;159(July):388–402. doi:10.1016/j.neuroimage.2017.07.065.
27. Naskar A, Vattikonda A, Deco G, Roy D, Banerjee A. Multiscale dynamic mean field (MDMF) model relates resting-state brain dynamics with local cortical excitatory–inhibitory neurotransmitter homeostasis. *Network Neuroscience*. 2021;5(3):757–782. doi:10.1162/netn\_a.00197.
28. Abbott LF, Chance FS. Drivers and modulators from push-pull and balanced synaptic input. *Progress in Brain Research*. 2005;149:147–155. doi:10.1016/S0079-6123(05)49011-1.
29. Mary A, Dayan J, Leone G, Postel C, Fraisse F, Malle C, et al. Resilience after trauma: The role of memory suppression. *Science*. 2020;367(6479). doi:10.1126/science.aay8477.
30. Wang P, Kong R, Kong X, Liégeois R, Orban C, Deco G, et al.. Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain; 2019. Available from: <http://advances.sciencemag.org/>.
31. Anderson MC, Bunce JG, Barbas H. Prefrontal–hippocampal pathways underlying inhibitory control over memory; 2016.
32. Harush U, Barzel B. Dynamic patterns of information flow in complex networks. *Nature Communications*. 2017;8(1):1–11. doi:10.1038/s41467-017-01916-3.
33. Mohan VM, Banerjee A. A perturbative approach to study information communication in brain networks. *Network Neuroscience*. 2022;6(4):1275–1295. doi:10.1162/netn\_a.00260.

## 7 Figures

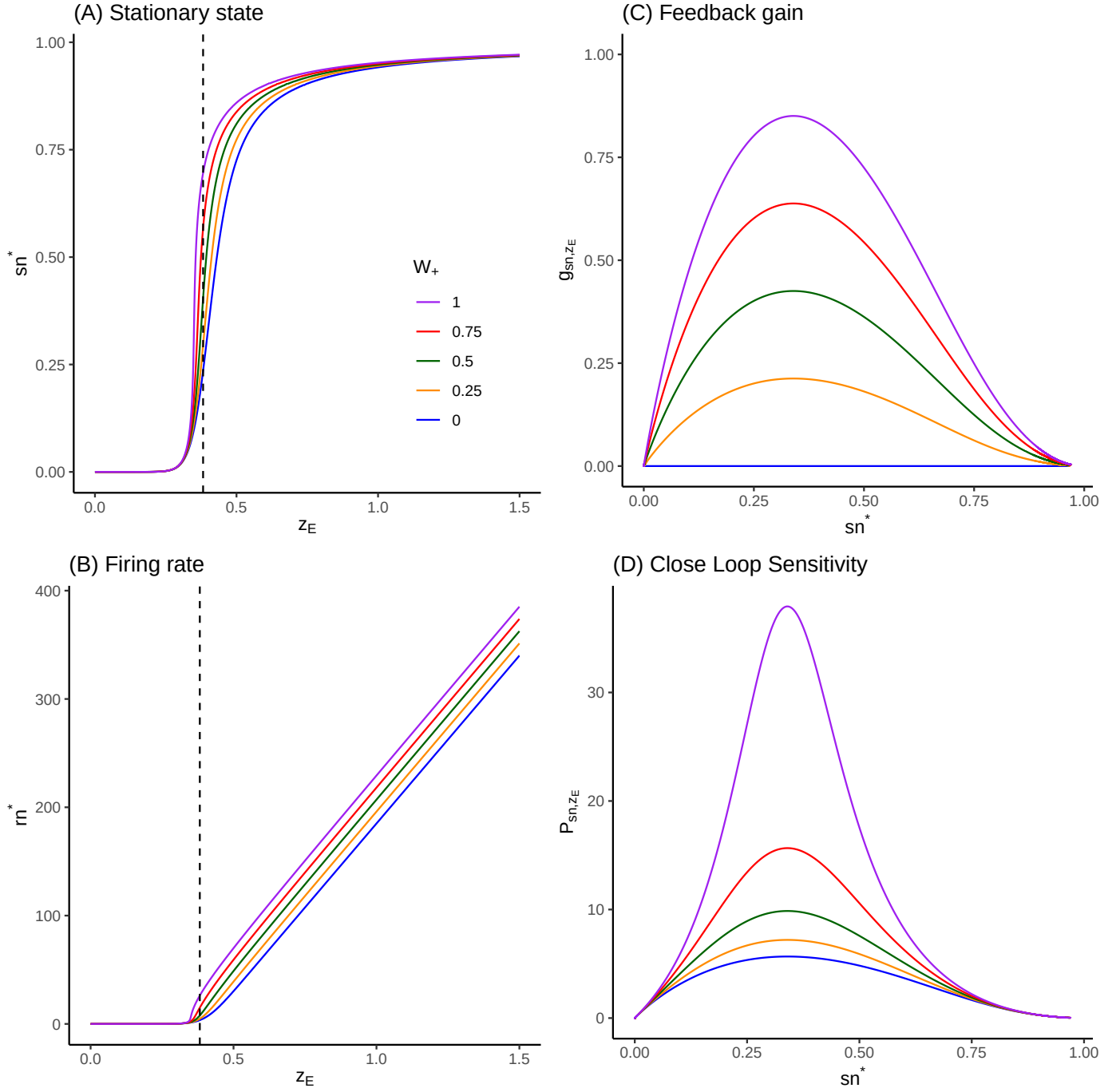


**Fig 5.** *Response of excitatory pool to inputs, feedback gain and sensitivities.* (A) Fixed point values of  $sn^*$  are numerically extracted by root-finding of Eq 5, for values of  $z_E \in [0; 1.5]$  and  $W_+ \in (0, 0.25, 0.5, 0.75, 1)$ . Dashed line correspond to  $z_E = x_{E,ref} = 0.382$  nA. (B) Corresponding fire rates, following Eq 2. (C) Feedback gain values are reported as function of  $sn^*$ , following Eq 13. (D) Corresponding close loop sensitivities are reported, following Eq 15.

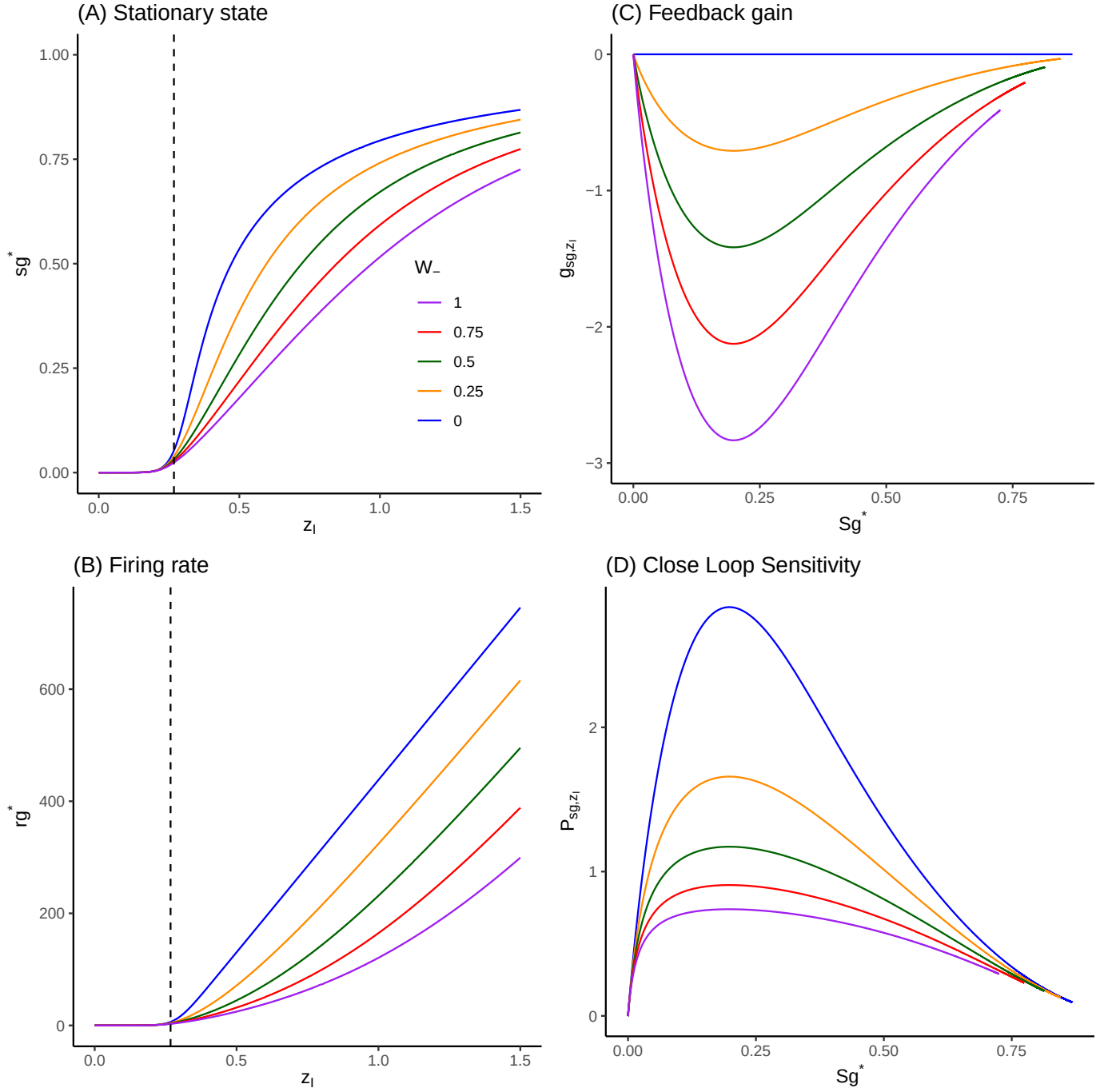


**Fig 6.** *Response of inhibitory pool to inputs, feedback gain and sensitivities.* (A) Fixed point values of  $Sg^*$  are numerically extracted by root-finding, for values of  $z_I \in [0; 1.5]$  and  $J_- \in (0, 0.25, 0.5, 0.75, 1)$ . Dashed line correspond to  $z_I = x_{I,ref} = 0.382 * 0.7$  nA. (B) Corresponding fire rates, following Eq 2. (C) Feedback gain values are reported as function of  $Sg^*$ . (D) Corresponding (close loop) sensitivities are reported following 15.

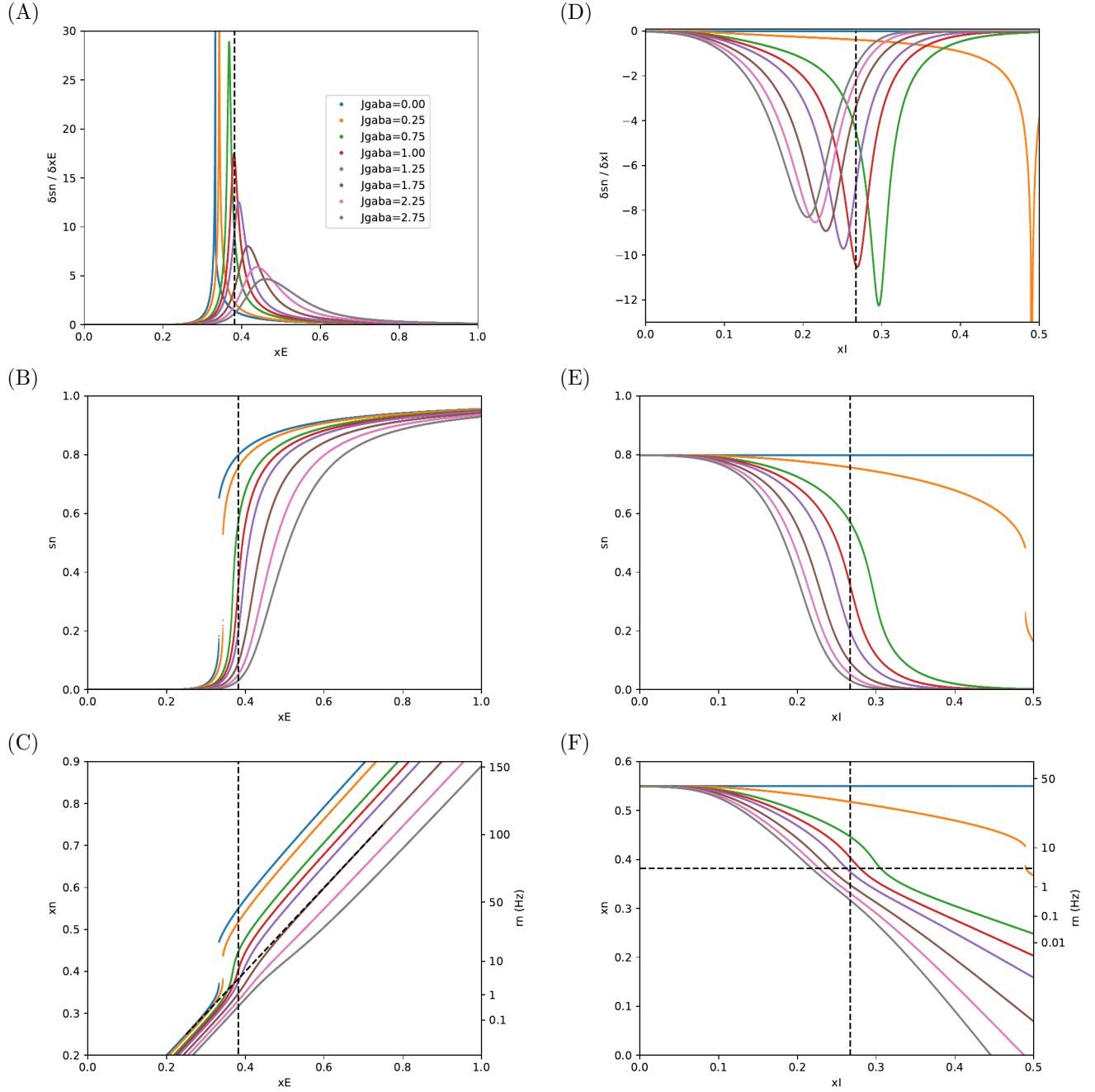


**Fig 7.** *Sensitivities and responses of excitatory pool in an isolated area.* (A—C) Response to external input upon the excitatory pool, (D—F) Response to external input upon the inhibitory pool. (A, D) Sensitivities, (B, E)  $sn^*$ , (C, F)  $rn^*$ . See comments in section 3.2.

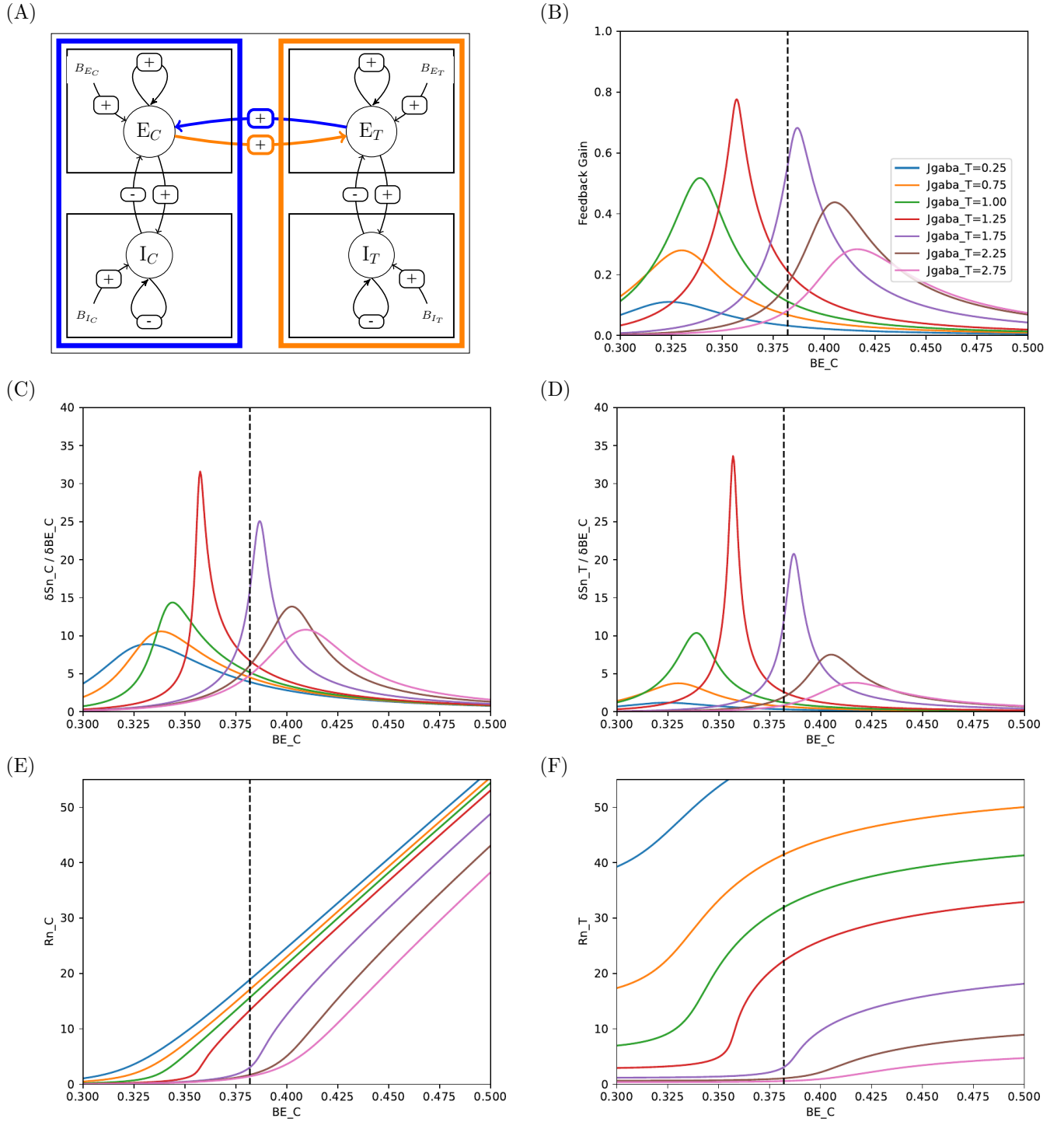


**Fig 8.** Responses and sensitivities of excitatory pools in two-area system with connectivity E-E. (A) Illustration of the connectivity. (B) Feedback gain  $g_{sn_C|T, BE_C}$  (Eq 22). (C, D) Sensitivities of  $sn$  to perturbations upon excitatory pool of Control Area, in Control and Target areas respectively. (E, F) Corresponding firing rates.

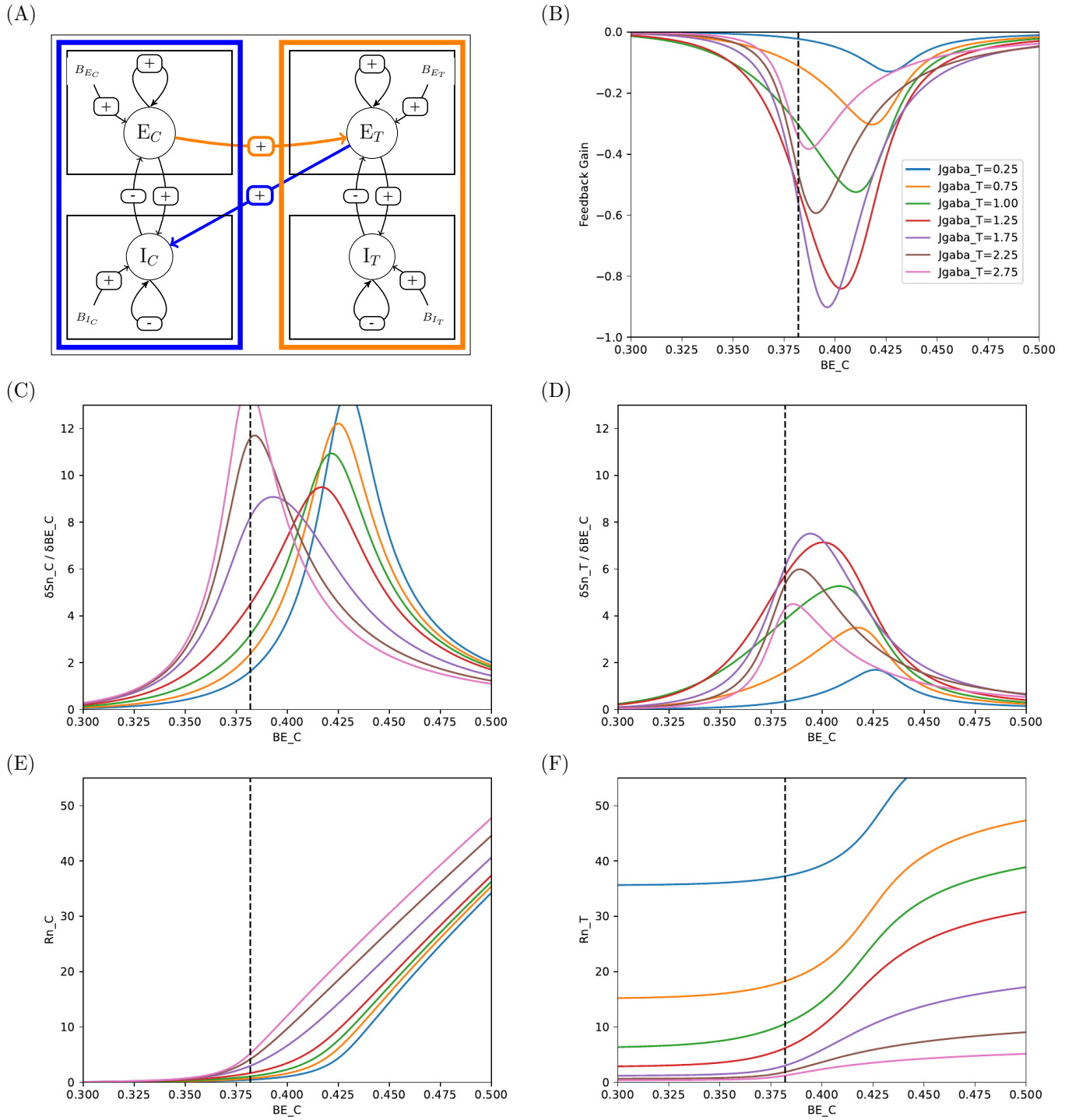


**Fig 9.** Responses and sensitivities of excitatory pools in two-area system with connectivity I-E. Same legend as in Fig 8.

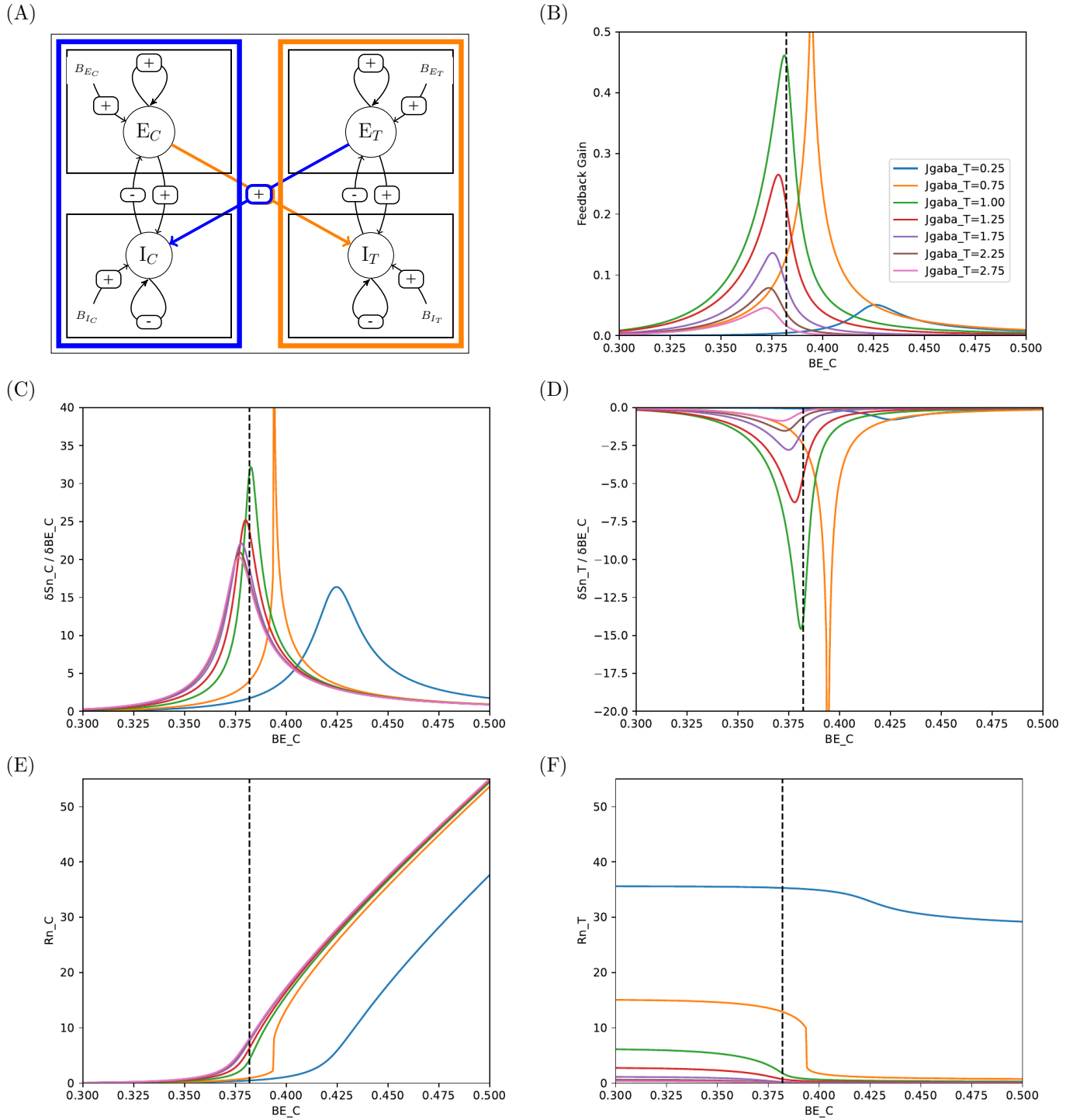


**Fig 10.** Responses and sensitivities of excitatory pools in two-area system with connectivity I-I. Same legend as in Fig 8



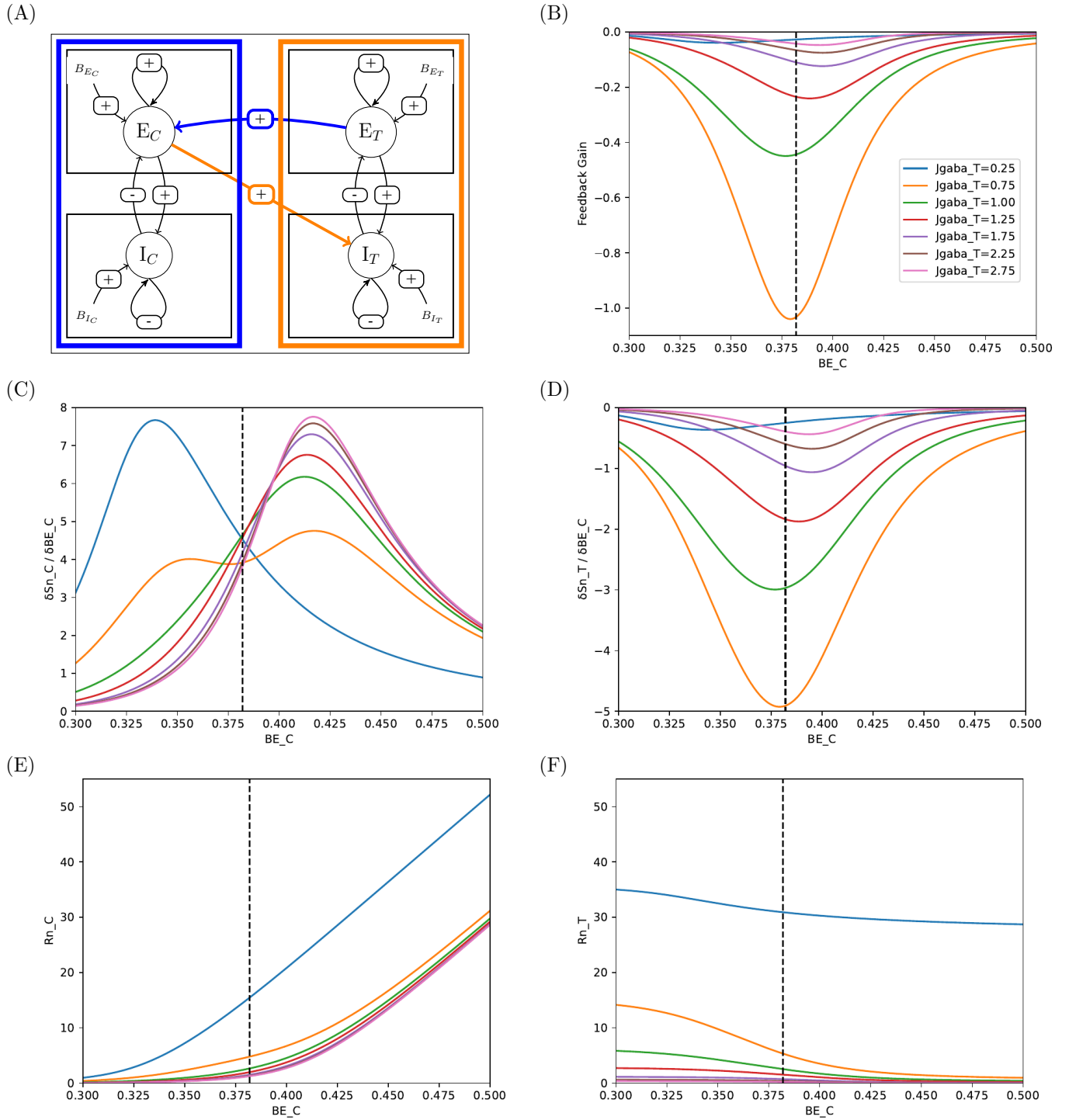
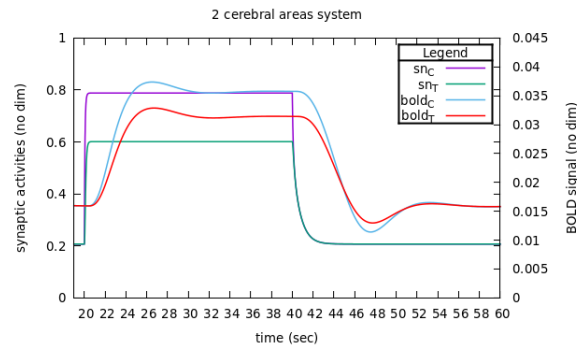
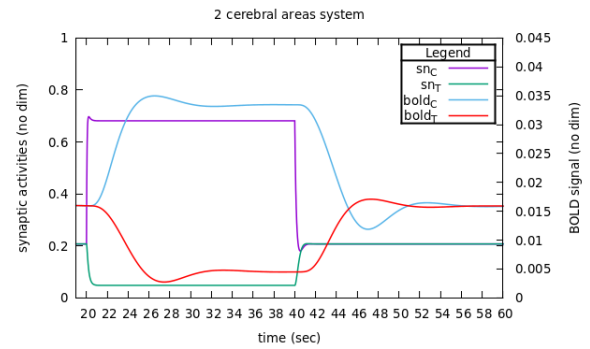


**Fig 11.** Responses and sensitivities of excitatory pools in two-area system with connectivity  $E-I$ . Same legend as in Fig 8

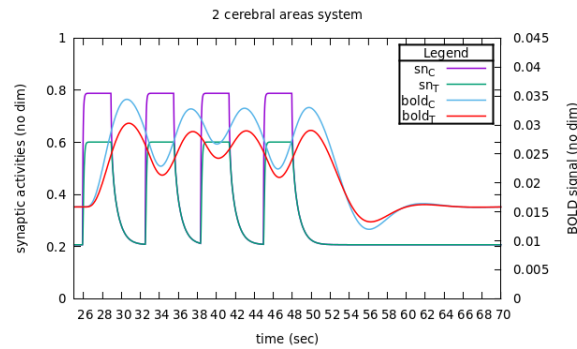
(A)



(B)



(C)



(D)

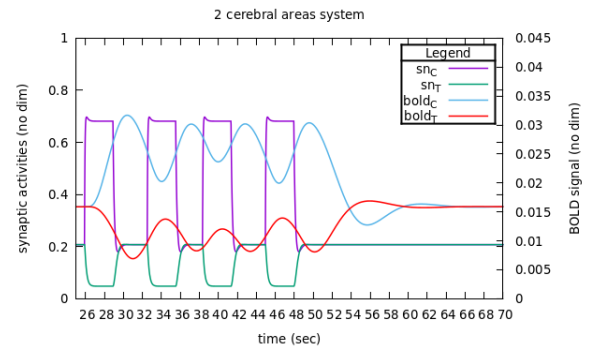


**Fig 12.** Response as BOLD signals in E-E and E-I connectivities. Left : E-E connectivity, right E-I connectivity. (A, B) A stimulation of 0.1 nA is applied upon the excitatory pool of the Control area during 20 s. (C, D) Time course of stimulations are made typical of a TNT task with stimulation steps during 3 s, separated by random intervals between 2.4 and 3.6 s. See comments in Sec 3.4.1.

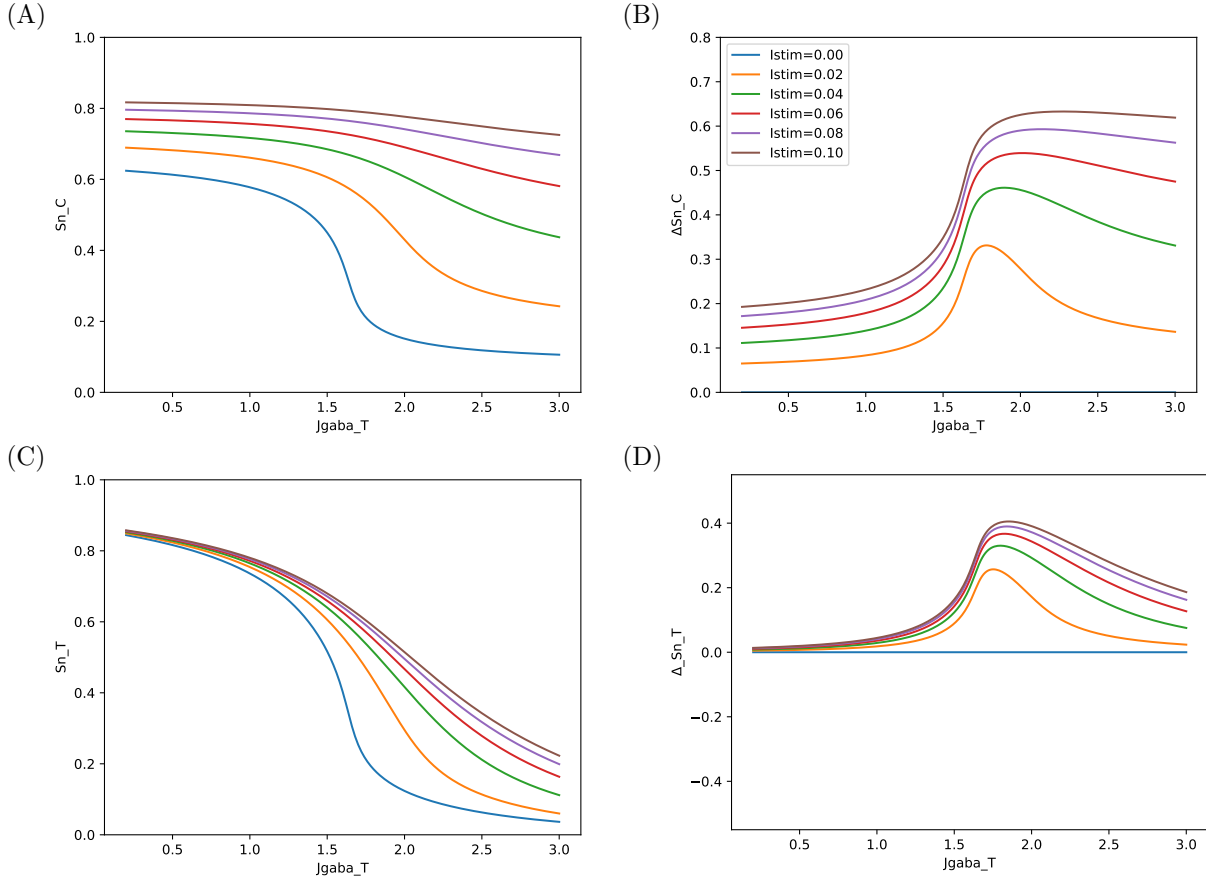


**Fig 13.** Effect of varying  $J_{gaba\_T}$ , the self-inhibitory strength within Target area, upon response to step activation of Control area, in E-E connectivity. A: Fixed point values of activity in Control area once stimulus is on (color legend in B), as functions of  $J_{gaba\_T}$ . The basic activity before stimulation is given by the blue line ( $I_{stim} = 0$ ). B: Corresponding finite size response  $\Delta Sn_C$  in Control area to different strength of step activation, as a function of  $J_{gaba\_T}$ , e.g. orange line corresponds to the difference between orange and blue lines in A. C: Corresponding fixed point values of activity in Target area. D: Corresponding finite size response  $\Delta Sn_T$  in Target Area. See comments in Sec 3.4.2.

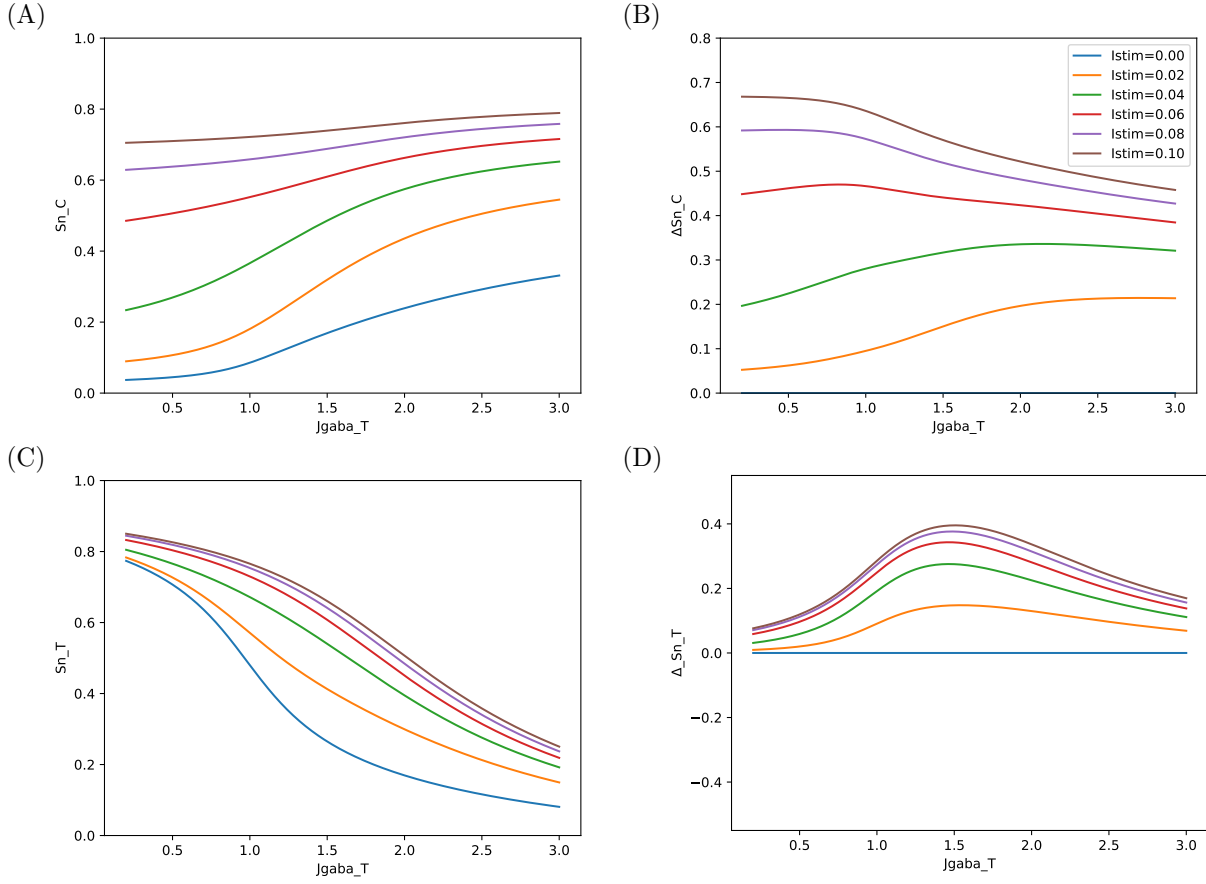


**Fig 14.** Effect of varying  $J_{gaba\_T}$ , the self-inhibitory strength within Target area, upon response to step activation of Control area, in I-E connectivity. Same legend as in Fig. 13.

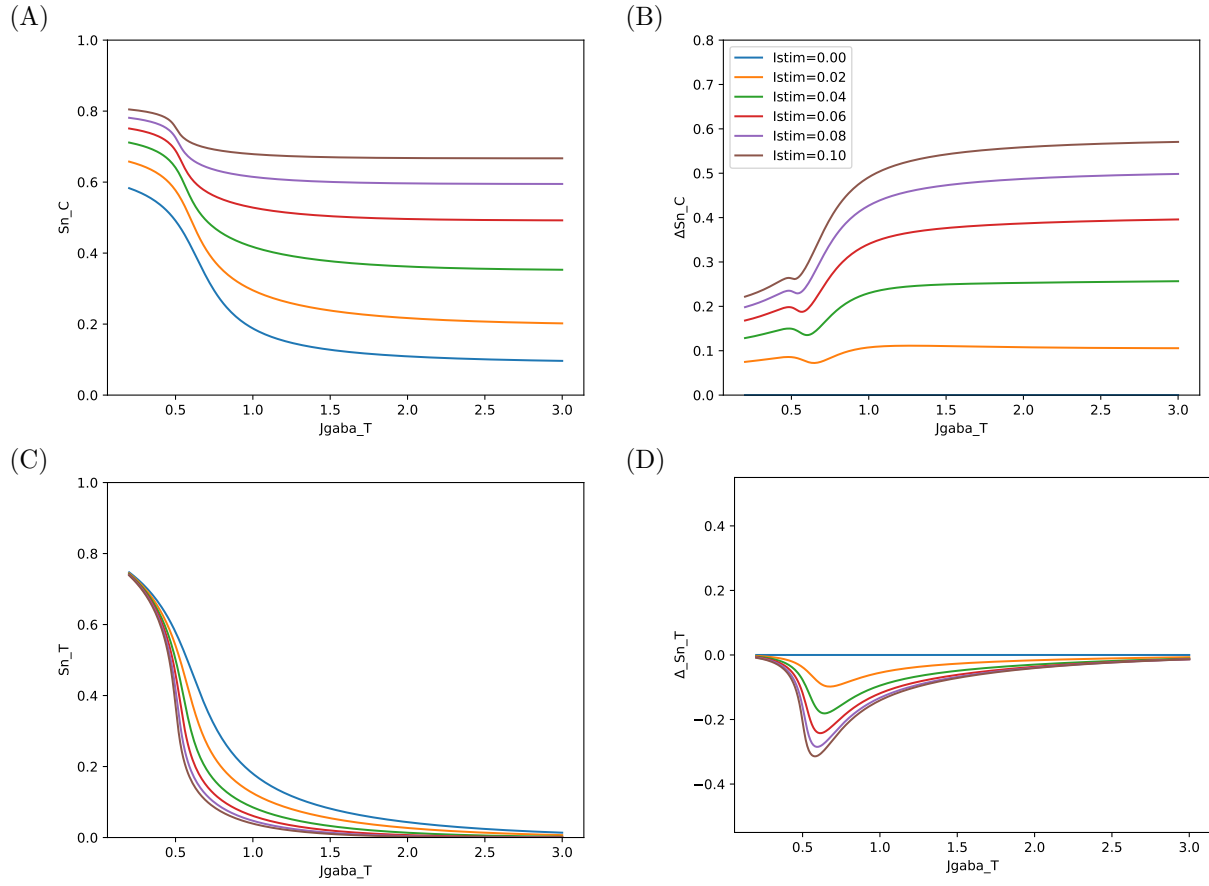


**Fig 15.** Effect of varying  $J_{gaba\_T}$ , the self-inhibitory strength within Target area, upon response to step activation of Control area, in E-I connectivity. Same legend as in Fig. 13.

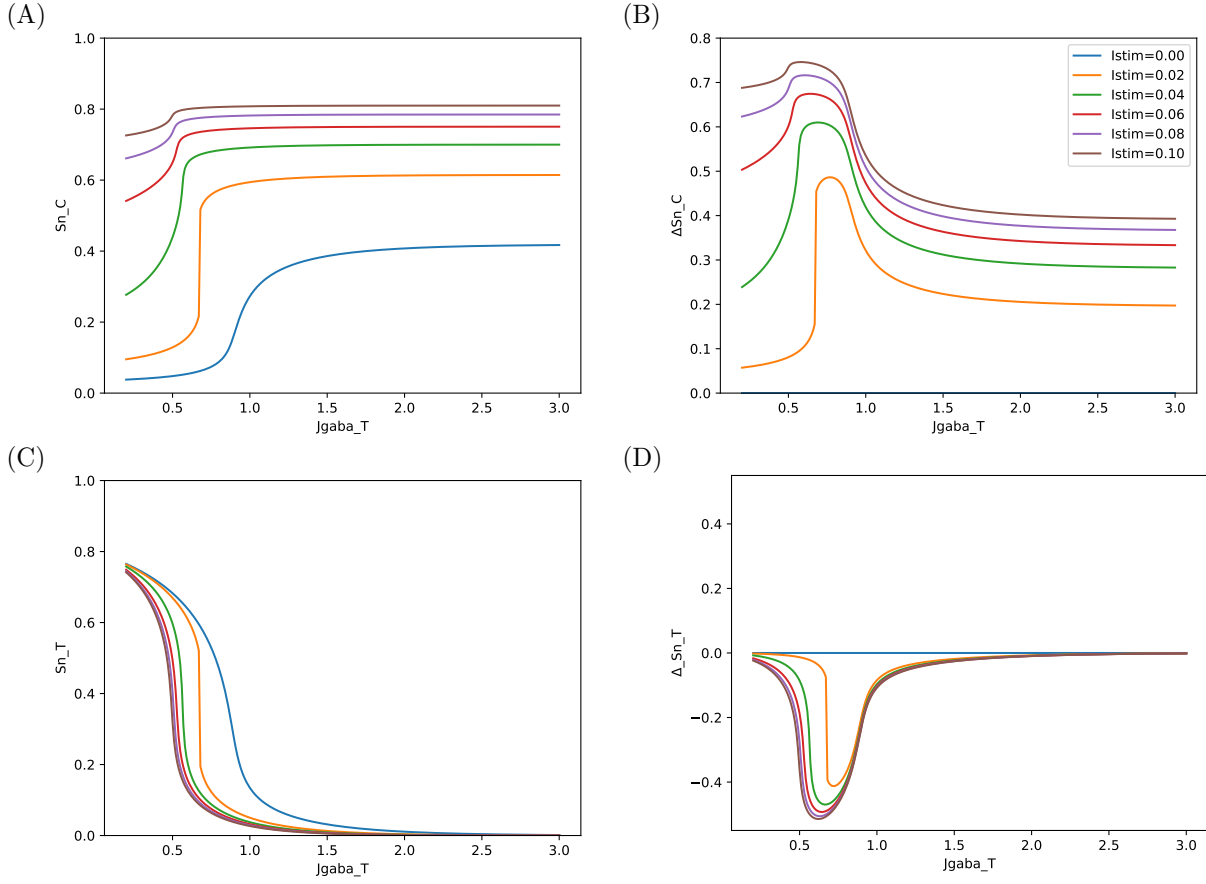


**Fig 16.** Effect of varying  $J_{gaba\_T}$ , the self-inhibitory strength within Target area, upon response to step activation of Control area, in I-I connectivity. Same legend as in Fig. 13.

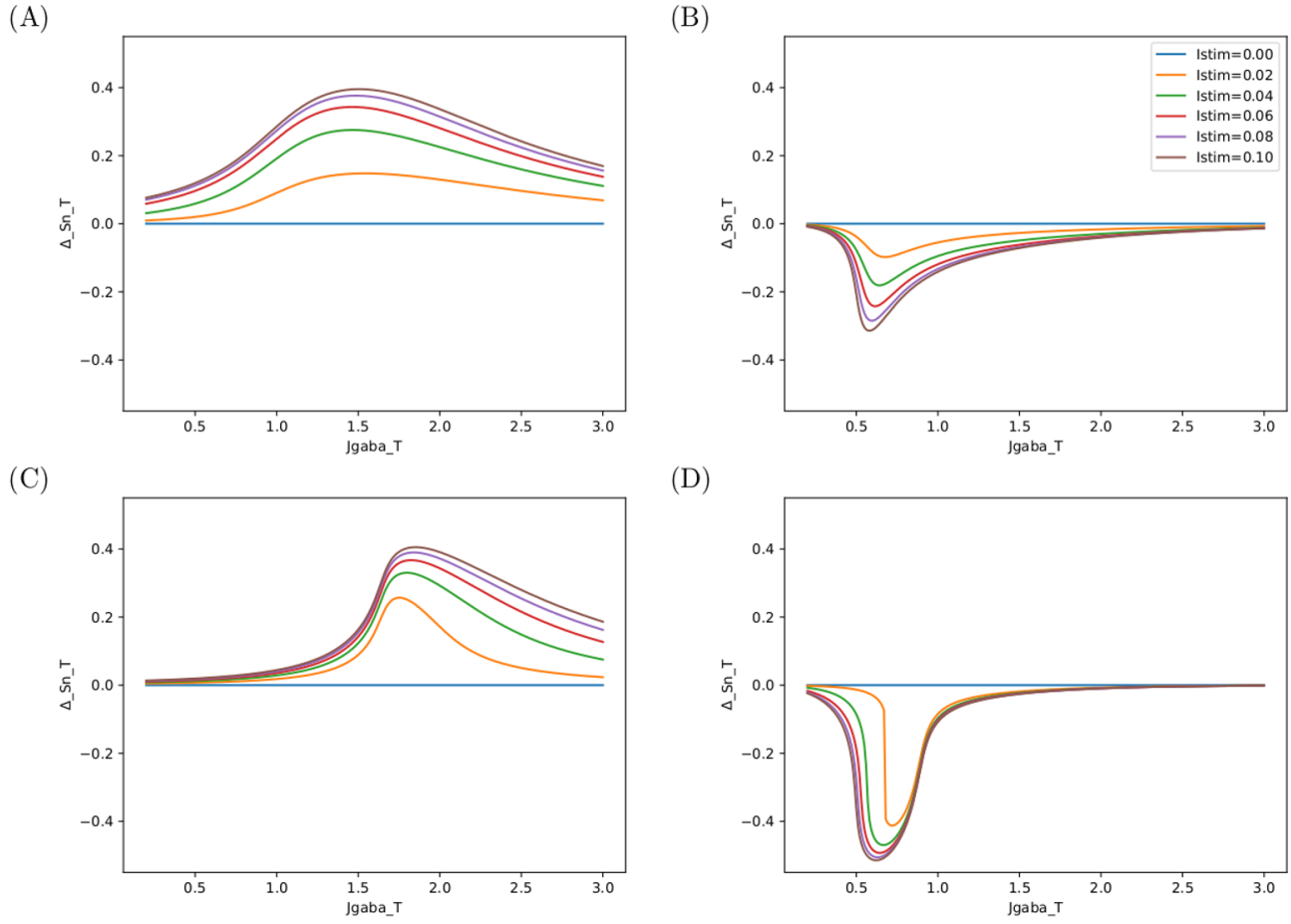


**Fig 17.** Summary figure for comparison of Figs 14,15,13,16(D) : Effect of varying  $J_{gaba\_T}$ , the self-inhibitory strength within Target area, upon response of Target Area to step activation of Control area, in (A) I-E, (B) E-I, (C) E-E and (D) I-I connectivities.

## 9.2 Annexes de l'article



# Supplemental Information for the paper: Sensitivity analysis enlightens inhibitory connectivity in a Neural Mass Model under Control-Target mode

Vallet Anaïs<sup>1</sup>, Blanco Stéphane<sup>2</sup>, Chevallier Coline<sup>2,3</sup>, Eustache Francis<sup>1</sup>, Gautrais Jacques<sup>2,3</sup>, Granpeix Jean-Yves<sup>4</sup>, Joly Jean-Louis<sup>2</sup>, Segobin Shail<sup>1</sup>, Gagnepain Pierre<sup>1</sup>,  
**1** Normandie Univ, Neuropsychol & Imagerie Memoire Humaine, GIP Cyceon, UNICAEN, PSL Res Univ, EPHE, INSERM, U1077, CHU Caen, F-14000 Caen, France  
**2** LAPLACE, UMR 5213 - Université Paul Sabatier, 118, Route de Narbonne, Toulouse Cedex 31062, France  
**3** Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, France  
**4** LMD/IPSL/SU, CNRS, Paris 75005, France

\* jacques.gautrais@cnrs.fr

## Abstract

This document exposes the formal developments to build sensitivities in the coupled system as a hierarchy of nested sensitivities.

## Contents

|          |                                                                                                        |           |
|----------|--------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Sensitivity analysis for the isolated pools</b>                                                     | <b>2</b>  |
| 1.1      | Isolated excitatory pool . . . . .                                                                     | 2         |
| 1.2      | Isolated inhibitory pool . . . . .                                                                     | 3         |
| 1.3      | Functional expressions of the sensitivities for the isolated excitatory and inhibitory pools . . . . . | 4         |
| <b>2</b> | <b>Sensitivity analysis for one isolated area</b>                                                      | <b>5</b>  |
| 2.1      | Splitting $x_{inter}$ from $x_{intra}$ . . . . .                                                       | 5         |
| 2.2      | Intermediate variable elimination . . . . .                                                            | 5         |
| 2.3      | Expressing sensitivities w.r.t. forcings . . . . .                                                     | 6         |
| 2.3.1    | Expressing Open Loop Sensitivities . . . . .                                                           | 7         |
| 2.3.2    | Using Isolated Pool Sensitivities . . . . .                                                            | 8         |
| 2.3.3    | Closed Loop Sensitivities as functions of Open Loop Sensitivities . . . . .                            | 9         |
| 2.4      | Functional expressions of closed loop sensitivities . . . . .                                          | 9         |
| <b>3</b> | <b>Two coupled area system</b>                                                                         | <b>11</b> |
| 3.1      | Splitting $x_{inter}$ from $x_{intra}$ . . . . .                                                       | 11        |
| 3.2      | Intermediate variables elimination . . . . .                                                           | 11        |
| 3.3      | Expressing sensitivities w.r.t. forcings . . . . .                                                     | 12        |
| 3.3.1    | Perturbation of one forcing $B$ . . . . .                                                              | 12        |
| 3.3.2    | Expressing Open Loop Sensitivities . . . . .                                                           | 13        |
| 3.3.3    | Using Isolated Area Sensitivities . . . . .                                                            | 14        |
| 3.3.4    | Closed loop sensitivities as functions of Open Loop Sensitivities . . . . .                            | 15        |
| 3.4      | Sensitivities in a Target-Control system . . . . .                                                     | 16        |
| 3.4.1    | Extracting the case of interest . . . . .                                                              | 16        |

# 1 Sensitivity analysis for the isolated pools

We consider in this section one isolated excitatory pool and one isolated inhibitory pool. Each pool contains an excitatory or inhibitory recurrent coupling and an excitatory forcing. Here, we are interested in the sensitivity of each excitatory and inhibitory pool activity to excitatory forcing.

## 1.1 Isolated excitatory pool

At the fixed point, the system reads:

$$\begin{cases} 0 = -\beta^E sn^* + \alpha^E T_{glu}(1 - sn^*)rn^* \equiv fn(sn^*, rn^*) \\ rn^* = \frac{a_E xn^* - b_E}{1 - e^{-d_E(a_E xn^* - b_E)}} \equiv hn(xn^*) \\ xn^* = W_+ J_{nmda} sn^* + z_E \equiv wn(sn^*, z_E) \end{cases} \quad (1)$$

With the perturbation of  $z_E$ , we get the new model:

$$\begin{cases} fn(sn^*, rn^*) = 0 \\ rn^* = hn(xn^*) \\ xn^* = wn(sn^*, z_E + \delta z_E) \end{cases} \quad (2)$$

We search the formal expression for the sensitivity  $\frac{\delta sn}{\delta z_E}$ .

The linearization for the perturbation around fixed point  $sn^*$  yields:

$$\begin{cases} \frac{\partial fn}{\partial sn} \delta sn + \frac{\partial fn}{\partial rn} \delta rn = 0 \\ \delta rn = \frac{dhn}{dxn} \delta xn \equiv hn' \delta xn \\ \delta xn = \frac{\partial wn}{\partial sn} \delta sn + \frac{\partial wn}{\partial z_E} \delta z_E \end{cases} \quad (3)$$

Plugging the last two equations into the first, we get:

$$\left( \frac{\partial fn}{\partial sn} + \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial sn} \right) \delta sn = - \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial z_E} \delta z_E \quad (4)$$

Rearranging to read how  $\delta sn$  depends upon  $\delta z_E$ :

$$\left( 1 + \frac{\partial fn}{\partial sn} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial sn} \right) \delta sn = - \frac{\partial fn}{\partial rn} \frac{\partial fn}{\partial sn} \frac{dhn}{dxn} \frac{\partial wn}{\partial z_E} \delta z_E \quad (5)$$

Then we get:

$$\frac{\delta sn}{\delta z_E} = \frac{- \frac{\partial fn}{\partial sn} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial z_E}}{1 + \frac{\partial fn}{\partial sn} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial sn}} \quad (6)$$

$$= - \frac{1}{\frac{\partial fn}{\partial sn} \frac{\partial fn}{\partial rn} \frac{dhn}{dxn} \frac{\partial wn}{\partial z_E} + \frac{\partial wn}{\partial sn} \frac{\partial wn}{\partial z_E} - 1} \quad (7)$$

Considering that

$$\frac{\partial fn}{\partial sn} = -\beta_E - \alpha_E T_{glu} r n^* \quad (8)$$

$$\frac{\partial fn}{\partial rn} = \alpha_E T_{glu} (1 - sn^*) \quad (9)$$

$$\frac{\partial wn}{\partial sn} = W_+ J_{nmda} \quad (10)$$

$$\frac{\partial wn}{\partial z_E} = 1 \quad (11)$$

We can calculate:

$$\frac{\delta sn}{\delta z_E} = \frac{1}{\frac{\beta_E + \alpha_E T_{glu} r n^*}{\alpha_E T_{glu} (1 - sn^*) \frac{dhn}{dxn}} - W_+ J_{nmda}} \quad (12)$$

## 1.2 Isolated inhibitory pool

At the fixed point, the system reads:

$$\begin{cases} 0 = -\beta^I sg^* + \alpha^I T_{gaba} (1 - sg^*) rg^* \equiv fg(sg^*, rg^*) \\ rg^* = \frac{a_I xg^* - b_I}{1 - e^{-d_I(a_I xg^* - b_I)}} \equiv hg(xg^*) \\ xg^* = -J_- sg^* + z_I \equiv wg(sg^*, z_I) \end{cases} \quad (13)$$

With the perturbation of  $z_I$ , we get the new model:

$$\begin{cases} fg(sg^*, rg^*) = 0 \\ rg^* = hg(xg^*) \\ xg^* = wg(sg^*, z_I + \delta z_I) \end{cases} \quad (14)$$

We search the formal expression for the sensitivity  $\frac{\delta sg}{\delta z_I}$ .

The linearization for the perturbation around fixed point  $sg^*$  yields:

$$\begin{cases} \frac{\partial fg}{\partial sg} \delta sg + \frac{\partial fg}{\partial rg} \delta rg = 0 \\ \delta rg = \frac{dhg}{dxg} \delta xg \equiv hg' \delta xg \\ \delta xg = \frac{\partial wg}{\partial sg} \delta sg + \frac{\partial wg}{\partial z_I} \delta z_I \end{cases} \quad (15)$$

Plugging the last two equations into the first, we get:

$$\left( \frac{\partial fg}{\partial sg} + \frac{\partial fg}{\partial rg} \frac{dhg}{dxg} \frac{\partial wg}{\partial sg} \right) \delta sg = - \frac{\partial fg}{\partial rg} \frac{dhg}{dxg} \frac{\partial wg}{\partial z_I} \delta z_I \quad (16)$$

Rearranging to read how  $\delta sg$  depends upon  $\delta z_I$ :

$$\left( 1 + \frac{\partial fg}{\partial sg}^{-1} \frac{\partial fg}{\partial rg} \frac{dhg}{dxg} \frac{\partial wg}{\partial sg} \right) \delta sg = - \frac{\partial fg}{\partial sg}^{-1} \frac{\partial fg}{\partial rg} \frac{dhg}{dxg} \frac{\partial wg}{\partial z_I} \delta z_I \quad (17)$$

Then we get:

$$\frac{\delta sg}{\delta z_I} = \frac{- \frac{\partial fg}{\partial sg}^{-1} \frac{\partial fg}{\partial rg} \frac{dhg}{dxg} \frac{\partial wg}{\partial z_I}}{1 + \frac{\partial fg}{\partial sg}^{-1} \frac{\partial fg}{\partial rg} \frac{dhg}{dxg} \frac{\partial wg}{\partial sg}} \quad (18)$$

$$= - \frac{1}{\frac{\partial fg}{\partial sg} \frac{\partial fg}{\partial rg}^{-1} \frac{dhg}{dxg}^{-1} \frac{\partial wg}{\partial z_I}^{-1} + \frac{\partial wg}{\partial sg} \frac{\partial wg}{\partial z_I}^{-1}} \quad (19)$$

Considering that

22

$$\frac{\partial fg}{\partial sg} = -\beta_I - \alpha_I T_{gaba} r g^* \quad (20)$$

$$\frac{\partial fg}{\partial rg} = \alpha_I T_{gaba} (1 - sg^*) \quad (21)$$

$$\frac{\partial wg}{\partial sg} = -J_- \quad (22)$$

$$\frac{\partial wg}{\partial z_I} = 1 \quad (23)$$

We can calculate:

23

$$\frac{\delta sg}{\delta z_I} = \frac{1}{\frac{\beta_I + \alpha_I T_{gaba} r g^*}{\alpha_I T_{gaba} (1 - sg^*) \frac{d hg}{d x g}} + J_-} \quad (24)$$

### 1.3 Functional expressions of the sensitivities for the isolated excitatory and inhibitory pools

24

25

We define the function  $\varphi n_E$  and  $\varphi g_I$  to explicitly express the dependencies of the sensitivities for an isolated excitatory pool and an isolated inhibitory pool:

26

27

$$\begin{cases} \varphi n_E(sn^*, z_E) \equiv \left( \frac{\beta_E + \alpha_E T_{glu} hn(w n(sn^*, z_E))}{\alpha_E T_{glu} (1 - sn^*) hn'(w n(sn^*, z_E))} - W_+ J_{nmda} \right)^{-1} \\ \varphi g_I(sg^*, z_I) \equiv \left( \frac{\beta_I + \alpha_I T_{gaba} hg(wg(sg^*, z_I))}{\alpha_I T_{gaba} (1 - sg^*) hg'(wg(sg^*, z_I))} + J_- \right)^{-1} \end{cases} \quad (25)$$

where

28

$$h_{\square}'(x) = \frac{(ae^{d(ax-b)}(e^{d(ax-b)} - adx + bd - 1))}{(e^{d(ad-b)} - 1)^2} \quad (26)$$

with  $a$ ,  $b$  and  $d$  to be taken accordingly.

29

## 2 Sensitivity analysis for one isolated area

At the fixed point, the system with the two coupled pools reads:

$$\begin{cases} 0 = -\beta^E sn^* + \alpha^E T_{glu}(1 - sn^*)rn^* \equiv fn(sn^*, rn^*) \\ 0 = -\beta^I sg^* + \alpha^I T_{gaba}(1 - sg^*)rg^* \equiv fg(sg^*, rg^*) \end{cases} \quad (27)$$

with

$$\begin{cases} rn^* = \frac{a_E xn^* - b_E}{1 - e^{-d_E(a_E xn^* - b_E)}} \equiv hn(xn^*) \\ rg^* = \frac{a_I xg^* - b_I}{1 - e^{-d_I(a_I xg^* - b_I)}} \equiv hg(xg^*) \end{cases} \quad (28)$$

in which  $xn^*$  and  $xg^*$  represent the respective total input currents:

$$\begin{cases} xn^* = W_+ J_{nmda} sn^* - J_{gaba} sg^* + x_E \\ xg^* = J_{nmda} sn^* - J_- sg^* + x_I \end{cases} \quad (29)$$

where  $x_E$  and  $x_I$  represent basal forcings (effective external inputs), that we will respectively perturbate.

### 2.1 Splitting $x_{inter}$ from $x_{intra}$

In order to express closed loop sensitivities of the pools as a function of their open loop sensitivities, we split explicitly the total input currents between an internal current, representing the effect of intra-pool recurrences:

$$\begin{cases} xn_{intra}^* = W_+ J_{nmda} sn^* \equiv wn_{intra}(sn^*) \\ xg_{intra}^* = -J_- sg^* \equiv wg_{intra}(sg^*) \end{cases} \quad (30)$$

and an external current, representing the total amount of forcing, due to the basic external forcing and the feedback current from the alternate pool:

$$\begin{cases} xn_{inter}^* = -J_{gaba} sg^* + x_E \equiv wn_{inter}(sg^*, x_E) \\ xg_{inter}^* = J_{nmda} sn^* + x_I \equiv wg_{inter}(sn^*, x_I) \end{cases} \quad (31)$$

so we have:

$$\begin{cases} xn^* = xn_{intra}^* + xn_{inter}^* \equiv wn(xn_{intra}^*, xn_{inter}^*) \\ xg^* = xg_{intra}^* + xg_{inter}^* \equiv wg(xg_{intra}^*, xg_{inter}^*) \end{cases} \quad (32)$$

### 2.2 Intermediate variable elimination

We define

$$\vec{s} = \begin{pmatrix} sn^* \\ sg^* \end{pmatrix}, \quad \vec{r} = \begin{pmatrix} rn^* \\ rg^* \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} xn^* \\ xg^* \end{pmatrix}, \quad \overrightarrow{x_{intra}} = \begin{pmatrix} xn_{intra}^* \\ xg_{intra}^* \end{pmatrix}, \quad \overrightarrow{x_{inter}} = \begin{pmatrix} xn_{inter}^* \\ xg_{inter}^* \end{pmatrix} \quad (33)$$

so that we can summarize the fixed point as:

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\overrightarrow{x_{intra}}, \overrightarrow{x_{inter}}) \\ \overrightarrow{x_{intra}} = \overrightarrow{w_{intra}}(\vec{s}) \\ \overrightarrow{x_{inter}} = \overrightarrow{w_{inter}}(\vec{s}, x_E, x_I) \end{cases} \quad (34)$$

Plugging intermediate variables  $\vec{r}$ ,  $\vec{x}$ ,  $\overrightarrow{x_{intra}}$  into the first equation, Eq 34 can be rewritten as:

$$\begin{cases} \vec{F}(\vec{s}, \overrightarrow{x_{inter}}) = \vec{0} \\ \overrightarrow{x_{inter}} = \overrightarrow{w_{inter}}(\vec{s}, x_E, x_I) \end{cases} \quad (35)$$

where

$$\begin{cases} Fn(sn^*, xn_{inter}^*) \equiv -\beta^E sn^* + \alpha^E T_{glu}(1 - sn^*) \frac{a_E(W_+ J_{nmda} sn^* + xn_{inter}^*) - b_E}{1 - e^{-d_E[a_E(W_+ J_{nmda} sn^* + xn_{inter}^*) - b_E]}} \\ Fg(sg^*, xg_{inter}^*) \equiv -\beta^I sg^* + \alpha^I T_{gaba}(1 - sg^*) \frac{a_I(-J_- sg^* + xg_{inter}^*) - b_I}{1 - e^{-d_I[a_I(-J_- sg^* + xg_{inter}^*) - b_I]}} \end{cases} \quad (36)$$

and where  $\overrightarrow{x_{inter}}$ , given by Eq 31, will be the support of information transfer between the two pools (hence, denoted *transfer variables*).

### 2.3 Expressing sensitivities w.r.t. forcings

We start from Eq 35 where we consider the dependency to  $x_E$ .

With the perturbation, we get the new model:

$$\begin{cases} \vec{F}(\vec{s}, \overrightarrow{x_{inter}}) = \vec{0} \\ \overrightarrow{x_{inter}} = \overrightarrow{w_{inter}}(\vec{s}, x_E + \delta x_E, x_I) \end{cases} \quad (37)$$

We search the formal expression for:

$$\begin{pmatrix} \mathcal{A}_{sn, x_E} \\ \mathcal{A}_{sg, x_E} \end{pmatrix} \equiv \begin{pmatrix} \frac{\delta sn}{\delta x_E} \\ \frac{\delta sg}{\delta x_E} \end{pmatrix} = \frac{\overrightarrow{\delta s}}{\delta x_E} \quad (38)$$

The linearization for the perturbation around fixed points  $\vec{s}^* = (sn^*, sg^*)$  and  $\overrightarrow{x_{inter}} = (xn_{inter}^*, xg_{inter}^*)$  yields:

$$\begin{cases} \overrightarrow{\frac{\partial F}{\partial s} \delta s} + \overrightarrow{\frac{\partial F}{\partial x_{inter}} \delta x_{inter}} = \vec{0} \\ \overrightarrow{\delta x_{inter}} = \overrightarrow{\frac{\partial w_{inter}}{\partial s} \delta s} + \overrightarrow{\frac{\partial w_{inter}}{\partial x_E} \delta x_E} \end{cases} \quad (39)$$

Plugging the last expression into the first, we get:

$$\overrightarrow{\frac{\partial F}{\partial s} \delta s} + \overrightarrow{\frac{\partial F}{\partial x_{inter}} \frac{\partial w_{inter}}{\partial s} \delta s} + \overrightarrow{\frac{\partial F}{\partial x_{inter}} \frac{\partial w_{inter}}{\partial x_E} \delta x_E} = \vec{0} \quad (40)$$

Rearranging to read how  $\overrightarrow{\delta s}$  depends upon  $\delta x_E$ :

$$\left(\bar{\mathbb{I}} + \frac{\overline{\frac{\partial F}{\partial s}}^{-1} \overline{\frac{\partial F}{\partial x_{inter}} \frac{\partial w_{inter}}{\partial s}}}\right) \overrightarrow{\delta s} = - \frac{\overline{\frac{\partial F}{\partial s}}^{-1} \overline{\frac{\partial F}{\partial x_{inter}} \frac{\partial w_{inter}}{\partial x_E}}}{\overline{\frac{\partial F}{\partial s}}} \delta x_E \quad (41)$$

Defining

$$\bar{S} = - \frac{\overline{\frac{\partial F}{\partial s}}^{-1} \overline{\frac{\partial F}{\partial x_{inter}}}}{\overline{\frac{\partial F}{\partial s}}} \quad (42)$$

we get:

$$\left(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}}\right) \overrightarrow{\delta s} = \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial x_E}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}} \delta x_E \quad (43)$$

so that:

$$\begin{pmatrix} \mathcal{A}_{sn, x_E} \\ \mathcal{A}_{sg, x_E} \end{pmatrix} = \frac{\overrightarrow{\delta s}}{\delta x_E} = \left(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}}\right)^{-1} \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial x_E}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}} \quad (44)$$

Now turning to  $x_I$ , we can proceed the same way, expressing at fixed point the model with perturbation::

$$\begin{cases} \vec{F}(\vec{s}, \vec{x}_{inter}) = \vec{0} \\ \vec{x}_{inter} = \vec{w}_{inter}(\vec{s}, x_E, x_I + \delta x_I) \end{cases} \quad (45)$$

and we obtain:

$$\begin{pmatrix} \mathcal{A}_{sn, x_I} \\ \mathcal{A}_{sg, x_I} \end{pmatrix} = \frac{\overrightarrow{\delta s}}{\delta x_I} = \left(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial x_I}}}\right)^{-1} \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial x_I}}}{\overline{\frac{\partial w_{inter}}{\partial x_I}}} \quad (46)$$

Using that:

$$\frac{\overline{\frac{\partial w_{inter}}{\partial x_E}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \frac{\overline{\frac{\partial w_{inter}}{\partial x_I}}}{\overline{\frac{\partial w_{inter}}{\partial x_I}}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (47)$$

we finally get a compact expression for the matrix of sensitivities to excitatory perturbations upon external forcings in either excitatory or inhibitory pool:

$$\begin{pmatrix} \mathcal{A}_{sn, x_E} & \mathcal{A}_{sn, x_I} \\ \mathcal{A}_{sg, x_E} & \mathcal{A}_{sg, x_I} \end{pmatrix} = \left(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial s}}}\right)^{-1} \bar{S} \quad (48)$$

### 2.3.1 Expressing Open Loop Sensitivities

In the closed loop situation, from Eq 31, we have that:

$$\frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial s}}} = \begin{pmatrix} 0 & -J_{gaba} \\ J_{nmda} & 0 \end{pmatrix} \quad (49)$$

and in the open loop situation, we set:  $\frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial s}}} = \vec{0}$ .

Under perturbation upon  $x_E$ , Eq 43 is

$$\underbrace{\left(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial s}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}}\right)}_{\bar{G}} \overrightarrow{\delta s} = \bar{S} \frac{\overline{\frac{\partial w_{inter}}{\partial x_E}}}{\overline{\frac{\partial w_{inter}}{\partial x_E}}} \delta x_E \quad (50)$$

where  $\overline{\overline{G}}$  is the matrix of feedback gain due to the closed loop between the two pools. In the open loop situation,  $\overline{\overline{G}}$  is then nullified while the r.h.s. remains untouched. In this case,

$$\overrightarrow{\delta s^O} = \overline{\overline{S}} \overrightarrow{\frac{\partial w_{inter}}{\partial x_E}} \delta x_E \quad (51)$$

represents the effect of perturbation upon fixed point values when the feedback loop has been opened.

The canonical form can then be written as:

$$(\overline{\mathbb{I}} - \overline{\overline{G}}) \overrightarrow{\delta s} = \overrightarrow{\delta s^O} \quad (52)$$

We have:

$$\frac{\overrightarrow{\delta s^O}}{\delta x_E} = \overline{\overline{S}} \underbrace{\overrightarrow{\frac{\partial w_{inter}}{\partial x_E}}}_{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \begin{pmatrix} \mathcal{A}_{sn,x_E}^O \\ 0 \end{pmatrix} \quad (53)$$

Obviously, in the open loop condition, the inhibitory pool is not affected by a perturbation upon  $x_E$ .

The same way, considering perturbation upon  $x_I$ , we get:

$$\frac{\overrightarrow{\delta s^O}}{\delta x_I} = \overline{\overline{S}} \underbrace{\overrightarrow{\frac{\partial w_{inter}}{\partial x_I}}}_{\begin{bmatrix} 0 \\ 1 \end{bmatrix}} = \begin{pmatrix} 0 \\ \mathcal{A}_{sg,x_I}^O \end{pmatrix} \quad (54)$$

In the open loop condition gain, the excitatory pool is not affected by a perturbation upon  $x_I$ .

Hence  $\overline{\overline{S}}$  reads:

$$\overline{\overline{S}} = \begin{pmatrix} \mathcal{A}_{sn,x_E}^O & 0 \\ 0 & \mathcal{A}_{sg,x_I}^O \end{pmatrix} \quad (55)$$

and represent the matrix of open loop sensitivities.

### 2.3.2 Using Isolated Pool Sensitivities

Now, lets consider  $z_E$  as:

$$z_E = -J_{gaba} sg^* + x_E = xn_{inter} = wn_{inter}(sg^*) \quad (56)$$

It represents the total amount of forcing at excitatory pool, due to the basic external forcing and the feedback current from the inhibitory pool.

Considering that perturbing here  $x_E$  is the same thing as perturbing  $z_E$  in the excitatory pool considered as an isolated pool, and that the same is true for the inhibitory pool, we then have

$$\begin{cases} \mathcal{A}_{sn,x_E}^O = \varphi n_E(sn^*, -J_{gaba} sg^* + x_E) \\ \mathcal{A}_{sg,x_I}^O = \varphi g_I(sg^*, J_{nmda} sn^* + x_I) \end{cases} \quad (57)$$

where  $\varphi n_E$  and  $\varphi g_I$  are given by Eq 25 and are to be evaluated at the fixed points  $sn^*$  and  $sg^*$  yielded by the closed loop system, and taking into account the total amount of external forcing.



### 2.3.3 Closed Loop Sensitivities as functions of Open Loop Sensitivities

We have

$$\frac{\overline{\overline{\partial w_{inter}}}}{\partial s} = \begin{bmatrix} \frac{\partial w_{inter}}{\partial sn} & \frac{\partial w_{inter}}{\partial sg} \\ \frac{\partial w_{inter}}{\partial sn} & \frac{\partial w_{inter}}{\partial sg} \end{bmatrix} = \begin{bmatrix} 0 & -J_{gaba} \\ J_{nmda} & 0 \end{bmatrix} \quad (58)$$

so, using Eq 55, we get:

$$(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\overline{\partial w_{inter}}}}{\partial s}) = \bar{\mathbb{I}} - \begin{bmatrix} \mathcal{A}_{sn,xE}^O & 0 \\ 0 & \mathcal{A}_{sg,xI}^O \end{bmatrix} \begin{bmatrix} 0 & -J_{gaba} \\ J_{nmda} & 0 \end{bmatrix} \quad (59)$$

$$= \bar{\mathbb{I}} - \begin{bmatrix} 0 & -J_{gaba} \mathcal{A}_{sn,xE}^O \\ J_{nmda} \mathcal{A}_{sg,xI}^O & 0 \end{bmatrix} \quad (60)$$

$$= \begin{bmatrix} 1 & J_{gaba} \mathcal{A}_{sn,xE}^O \\ -J_{nmda} \mathcal{A}_{sg,xI}^O & 1 \end{bmatrix} \quad (61)$$

Inverting:

$$(\bar{\mathbb{I}} - \bar{S} \frac{\overline{\overline{\partial w_{inter}}}}{\partial s})^{-1} = \frac{1}{1 + J_{gaba} \mathcal{A}_{sn,xE}^O J_{nmda} \mathcal{A}_{sg,xI}^O} \begin{bmatrix} 1 & -J_{gaba} \mathcal{A}_{sn,xE}^O \\ J_{nmda} \mathcal{A}_{sg,xI}^O & 1 \end{bmatrix} \quad (62)$$

$$(63)$$

Finally,

$$\begin{aligned} \begin{pmatrix} \mathcal{A}_{sn,xE} & \mathcal{A}_{sn,xI} \\ \mathcal{A}_{sg,xE} & \mathcal{A}_{sg,xI} \end{pmatrix} &= (\bar{\mathbb{I}} - \bar{S} \frac{\overline{\overline{\partial w_{inter}}}}{\partial s})^{-1} \bar{S} \\ &= \frac{1}{1 + J_{nmda} J_{gaba} \mathcal{A}_{sn,xE}^O \mathcal{A}_{sg,xI}^O} \\ &\quad \times \begin{bmatrix} \mathcal{A}_{sn,xE}^O & -J_{gaba} \mathcal{A}_{sn,xE}^O \mathcal{A}_{sg,xI}^O \\ J_{nmda} \mathcal{A}_{sn,xE}^O \mathcal{A}_{sg,xI}^O & \mathcal{A}_{sg,xI}^O \end{bmatrix} \end{aligned} \quad (64)$$

which expresses closed loop sensitivities (sensitivities for the pools when they are coupled) as functions of open loop sensitivities (sensitivities for the pools with only their recurrent coupling).

## 2.4 Functional expressions of closed loop sensitivities

We define the functions  $\Phi_{n_E}$ ,  $\Phi_{g_E}$ ,  $\Phi_{n_I}$  and  $\Phi_{g_I}$  to explicitly express the dependencies of the closed loop sensitivities for an isolated region:

$$\left\{ \begin{array}{l}
\mathcal{A}_{sn,x_E} = \Phi n_E(\vec{s}^*, x_E, x_I) = \frac{\mathcal{A}_{sn,x_E}^O}{1 + J_{nmda} J_{gaba} \mathcal{A}_{sn,x_E}^O \mathcal{A}_{sg,x_I}^O} \\
\mathcal{A}_{sg,x_E} = \Phi g_E(\vec{s}^*, x_E, x_I) = \frac{J_{nmda} \mathcal{A}_{sn,x_E}^O \mathcal{A}_{sg,x_I}^O}{1 + J_{nmda} J_{gaba} \mathcal{A}_{sn,x_E}^O \mathcal{A}_{sg,x_I}^O} \\
\mathcal{A}_{sn,x_I} = \Phi n_I(\vec{s}^*, x_E, x_I) = \frac{-J_{gaba} \mathcal{A}_{sn,x_E}^O \mathcal{A}_{sg,x_I}^O}{1 + J_{nmda} J_{gaba} \mathcal{A}_{sn,x_E}^O \mathcal{A}_{sg,x_I}^O} \\
\mathcal{A}_{sg,x_I} = \Phi g_I(\vec{s}^*, x_E, x_I) = \frac{\mathcal{A}_{sg,x_I}^O}{1 + J_{nmda} J_{gaba} \mathcal{A}_{sn,x_E}^O \mathcal{A}_{sg,x_I}^O}
\end{array} \right. \quad (65)$$

where the open loop sensitivities are given in Eq 57.

108

### 3 Two coupled area system

109

We now turn to the coupling between two areas. We seek the expression of their closed loop sensitivities to perturbation as functions of their sensitivities when isolated, as it has been expressed in the previous sections, hence, corresponding to their response to perturbation when the feedback loops are opened at the system scale.

110

111

112

113

#### 3.1 Splitting $x_{inter}$ from $x_{intra}$

114

At fixed point, the system can be expressed as:

115

$$\begin{cases} 0 = -\beta^E sn_i^* + \alpha^E T_{glu}(1 - sn_i^*)rn_i^* & \equiv fn_i(sn_i^*, rn_i^*) \\ 0 = -\beta^I sg_i^* + \alpha^I T_{gaba}(1 - sg_i^*)rg_i^* & \equiv fg_i(sn_i^*, rn_i^*) \end{cases} \quad (66)$$

with

116

$$\begin{cases} rn_i^* = \frac{a_E xn_i^* - b_E}{1 - e^{-d_E(a_E xn_i^* - b_E)}} & \equiv hn_i(xn_i^*) \\ rg_i^* = \frac{a_I xg_i^* - b_I}{1 - e^{-d_I(a_I xg_i^* - b_I)}} & \equiv hg_i(xg_i^*) \end{cases} \quad (67)$$

where  $i \in \{1, 2\}$ .

117

In Eq 67,  $xn_i^*$  and  $xg_i^*$  represent the respective total input current to area  $i$ . In order to express closed loop sensitivities of the areas as a function of their open loop sensitivities, we split explicitly this total input current between an internal current within an area, due to intra-pool recurrence, and coupling between pools:

118

119

120

121

$$\begin{cases} xn_{intra,i}^* = W_+ J_{nmda} sn_i^* - J_{gaba,i} sg_i^* & \equiv wn_{intra,i}(sn_i^*, sg_i^*) \\ xg_{intra,i}^* = J_{nmda} sn_i^* - J_- sg_i^* & \equiv wg_{intra,i}(sn_i^*, sg_i^*) \end{cases} \quad (68)$$

and external current due to the external inputs and the coupling between the two areas:

122

123

$$\begin{cases} xn_{inter,i}^* = k_{E_{ij}} \kappa_{ij} sn_j^* + B_{E_i} & \equiv wn_{inter,i}(sn_j^*) \quad j \neq i \\ xg_{inter,i}^* = (1 - k_{E_{ij}}) \kappa_{ij} sn_j^* + B_{I_i} & \equiv wg_{inter,i}(sn_j^*) \quad j \neq i \end{cases} \quad (69)$$

so that total input currents reads:

124

$$\begin{cases} xn_i^* = xn_{intra,i}^* + xn_{inter,i}^* = wn_i(xn_{intra,i}^*, xn_{inter,i}^*) \\ xg_i^* = xg_{intra,i}^* + xg_{inter,i}^* = wg_i(xg_{intra,i}^*, xg_{inter,i}^*) \end{cases} \quad (70)$$

#### 3.2 Intermediate variables elimination

125

We define

126

$$\vec{s} = \begin{pmatrix} sn_1 \\ sn_2 \\ sg_1 \\ sg_2 \end{pmatrix}, \quad \vec{r} = \begin{pmatrix} rn_1 \\ rn_2 \\ rg_1 \\ rg_2 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} xn_1 \\ xn_2 \\ xg_1 \\ xg_2 \end{pmatrix}, \quad \vec{x}_{intra} = \begin{pmatrix} xn_{intra,1} \\ xn_{intra,2} \\ xg_{intra,1} \\ xg_{intra,2} \end{pmatrix}, \quad \vec{x}_{inter} = \begin{pmatrix} xn_{inter,1} \\ xn_{inter,2} \\ xg_{inter,1} \\ xg_{inter,2} \end{pmatrix} \quad (71)$$

and

127

$$\vec{B}_E = \begin{pmatrix} B_{E_1} \\ B_{E_2} \end{pmatrix}, \quad \vec{B}_I = \begin{pmatrix} B_{I_1} \\ B_{I_2} \end{pmatrix} \quad (72)$$

so that we can summarize the fixed point as:

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{x}_{intra}, \vec{x}_{inter}) \\ \vec{x}_{intra} = \vec{w}_{intra}(\vec{s}) \\ \vec{x}_{inter} = \vec{w}_{inter}(\vec{s}, \vec{B}_E, \vec{B}_I) \end{cases} \quad (73)$$

Plugging intermediate variables  $\vec{r}$ ,  $\vec{x}$ ,  $\vec{x}_{intra}$  into the first equation, Eq 73 can be rewritten as:

$$\begin{cases} \vec{F}(\vec{s}, \vec{x}_{inter}) = \vec{0} \\ \vec{x}_{inter} = \vec{w}_{inter}(\vec{s}, \vec{B}_E, \vec{B}_I) \end{cases} \quad (74)$$

where

$$\begin{cases} Fn_i(sn_i^*, sg_i^*, xn_{inter,i}^*) = -\beta^E sn_i^* \\ \quad + \alpha^E T_{glu}(1 - sn_i^*) \frac{a_E(W_+ J_{nmda} sn_i^* - J_{gaba_i} sg_i^* + xn_{inter,i}^*) - b_E}{1 - e^{-d_E[a_E(W_+ J_{nmda} sn_i^* - J_{gaba_i} sg_i^* + xn_{inter,i}^*) - b_E]}} \\ Fg_i(sn_i^*, sg_i^*, xg_{inter,i}^*) = -\beta^I sg_i^* \\ \quad + \alpha^I T_{gaba}(1 - sg_i^*) \frac{a_I(J_{nmda} sn_i^* - J_- sg_i^* + xg_{inter,i}^*) - b_I}{1 - e^{-d_I[a_I(J_{nmda} sn_i^* - J_- sg_i^* + xg_{inter,i}^*) - b_I]}} \end{cases} \quad (75)$$

and where  $\vec{x}_{inter}$ , given by Eq 69, will be the transfer variables.

### 3.3 Expressing sensitivities w.r.t. forcings

#### 3.3.1 Perturbation of one forcing $B$

Here, we build the general expression of the sensitivity, for any given forcing  $B$  among  $B_{E_1}$ ,  $B_{E_2}$ ,  $B_{I_1}$  or  $B_{I_2}$ .

The perturbed form of system 74 reads:

$$\begin{cases} \vec{F}(\vec{s}, \vec{x}_{inter}) = \vec{0} \\ \vec{x}_{inter} = \vec{w}_{inter}(\vec{s}, B + \delta B) \end{cases} \quad (76)$$

By linearization at fixed point, we get:

$$\begin{cases} \frac{\partial \vec{F}}{\partial \vec{s}} \delta \vec{s}_B + \frac{\partial \vec{F}}{\partial \vec{x}_{inter}} \delta \vec{x}_{inter B} = \vec{0} \\ \delta \vec{x}_{inter B} = \frac{\partial \vec{w}_{inter}}{\partial \vec{s}} \delta \vec{s}_B + \frac{\partial \vec{w}_{inter}}{\partial B} \delta B \end{cases} \quad (77)$$

Plugging the second equation into the first, we have:

$$\left( \frac{\partial \vec{F}}{\partial \vec{s}} + \frac{\partial \vec{F}}{\partial \vec{x}_{inter}} \frac{\partial \vec{w}_{inter}}{\partial \vec{s}} \right) \delta \vec{s}_B = - \frac{\partial \vec{F}}{\partial \vec{x}_{inter}} \frac{\partial \vec{w}_{inter}}{\partial B} \delta B \quad (78)$$

Denoting

$$\bar{S} = -\frac{\overline{\frac{\partial F}{\partial s}}^{-1} \overline{\frac{\partial F}{\partial x_{inter}}}}{\overline{\frac{\partial F}{\partial s}}} \quad (79)$$

expression 78 can be written under the canonical form as:

$$(\bar{\mathbb{I}} - \bar{S} \overline{\frac{\partial w_{inter}}{\partial s}}) \overrightarrow{\delta s_B} = \bar{S} \overrightarrow{\frac{\partial w_{inter}}{\partial B}} \delta B \quad (80)$$

In this expression  $\overline{\frac{\partial w_{inter}}{\partial s}}$  represents the coupling between the two areas, and is:

$$\overline{\frac{\partial w_{inter}}{\partial s}} = \begin{pmatrix} 0 & \frac{\partial w_{inter,1}}{\partial s n_2} & 0 & 0 \\ \frac{\partial w_{inter,2}}{\partial s n_1} & 0 & 0 & 0 \\ 0 & \frac{\partial w_{inter,1}}{\partial s n_2} & 0 & 0 \\ \frac{\partial w_{inter,2}}{\partial s n_1} & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & k_{E_{12}} \kappa_{12} & 0 & 0 \\ k_{E_{21}} \kappa_{21} & 0 & 0 & 0 \\ 0 & (1 - k_{E_{12}}) \kappa_{12} & 0 & 0 \\ (1 - k_{E_{21}}) \kappa_{21} & 0 & 0 & 0 \end{pmatrix} \quad (81)$$

If we set  $\overline{\frac{\partial w_{inter}}{\partial s}} = \bar{0}$  while the r.h.s. remains untouched, we then obtain the effect of perturbation in the open loop case:

$$\overrightarrow{\delta s_B^O} = \bar{S} \overrightarrow{\frac{\partial w_{inter}}{\partial B}} \delta B \quad (82)$$

The canonical expression then reads:

$$(\bar{\mathbb{I}} - \bar{G}) \overrightarrow{\delta s_B} = \overrightarrow{\delta s_B^O} \quad (83)$$

where

$$\bar{G} = -\frac{\overline{\frac{\partial F}{\partial s}}^{-1} \overline{\frac{\partial F}{\partial x_{inter}}}}{\overline{\frac{\partial F}{\partial s}}} \overline{\frac{\partial w_{inter}}{\partial s}} = \bar{S} \overline{\frac{\partial w_{inter}}{\partial s}} \quad (84)$$

is the feedback gain matrix.

### 3.3.2 Expressing Open Loop Sensitivities

Perturbations are operated upon transfer variables, given by  $\overrightarrow{w_{inter}}$ .

Let  $\overrightarrow{x_{inter}^O}$  denote the transfer variables in open loop condition when  $\overline{\frac{\partial w_{inter}}{\partial s}} = \bar{0}$  (ie  $\kappa_{12} = \kappa_{21} = 0$ ):

$$\overrightarrow{x_{inter}^O} = \begin{pmatrix} B_{E_1} \\ B_{E_2} \\ B_{I_1} \\ B_{I_2} \end{pmatrix} \quad (85)$$

Since forcing parameters are independent, we have:  $\overline{\frac{\partial w_{inter}}{\partial x_{inter}^O}} = \bar{\mathbb{I}}$ .

Hence considering each perturbation one by one, the operator  $\overrightarrow{\frac{\partial w_{inter}}{\partial B}}$  acts as a selector of a column of  $\bar{S}$ , so that we can express the sensitivities in closed loop condition as function of sensitivities in open loop condition.

For instance, perturbing  $B_{E_1}$ , we get:

$$\vec{\delta s}_{B_{E_1}}^O = \overline{\overline{S}} \frac{\overrightarrow{\partial w_{inter}}}{\partial B_{E_1}} \delta B_{E_1} \iff \frac{\vec{\delta s}_{B_{E_1}}^O}{\delta B_{E_1}} = \underbrace{\overline{\overline{S}} \frac{\overrightarrow{\partial w_{inter}}}{\partial B_{E_1}}}_{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}} \quad (86)$$

hence the open loop sensitivities of both areas to a perturbation upon the excitatory pool of the first area yields the first column of  $\overline{\overline{S}}$ .

Obviously, in the open loop condition, the second area is not perturbed at all.

Furthermore, considering that perturbing here  $B_{E_1}$  is the same thing as perturbing  $x_{E_1}$  in the first area considered as an isolated area, we have, by definition given in Eq 38, that:

$$\vec{\delta s}_{B_{E_1}}^O = \begin{pmatrix} \delta sn_1 \\ \delta sn_2 \\ \delta sg_1 \\ \delta sg_2 \end{pmatrix}_{E_1, O} = \begin{pmatrix} \mathcal{A}_{sn_1, x_{E_1}} \delta B_{E_1} \\ 0 \\ \mathcal{A}_{sg_1, x_{E_1}} \delta B_{E_1} \\ 0 \end{pmatrix} \quad (87)$$

hence the first column of  $\overline{\overline{S}}$  is given by:

$$\frac{\vec{\delta s}_{B_{E_1}}^O}{\delta B_{E_1}} = \begin{pmatrix} \mathcal{A}_{sn_1, x_{E_1}} \\ 0 \\ \mathcal{A}_{sg_1, x_{E_1}} \\ 0 \end{pmatrix} \quad (88)$$

Following the same lines of reasoning for the three other perturbations, we finally obtain:

$$\overline{\overline{S}} = \begin{pmatrix} \overline{\overline{\mathcal{A}_{sn, x_E}}} & \overline{\overline{\mathcal{A}_{sn, x_I}}} \\ \overline{\overline{\mathcal{A}_{sg, x_E}}} & \overline{\overline{\mathcal{A}_{sg, x_I}}} \end{pmatrix} \quad (89)$$

with

$$\overline{\overline{\mathcal{A}_{sn, x_E}}} = \begin{pmatrix} \mathcal{A}_{sn_1, x_{E_1}} & 0 \\ 0 & \mathcal{A}_{sn_2, x_{E_2}} \end{pmatrix}, \quad \overline{\overline{\mathcal{A}_{sn, x_I}}} = \begin{pmatrix} \mathcal{A}_{sn_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sn_2, x_{I_2}} \end{pmatrix} \quad (90)$$

$$\overline{\overline{\mathcal{A}_{sg, x_E}}} = \begin{pmatrix} \mathcal{A}_{sg_1, x_{E_1}} & 0 \\ 0 & \mathcal{A}_{sg_2, x_{E_2}} \end{pmatrix}, \quad \overline{\overline{\mathcal{A}_{sg, x_I}}} = \begin{pmatrix} \mathcal{A}_{sg_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sg_2, x_{I_2}} \end{pmatrix} \quad (91)$$

### 3.3.3 Using Isolated Area Sensitivities

In the same spirit as in Sec 2.3.2, these open loop sensitivities of areas can be expressed by the analytical expression of their sensitivities when considered isolated, yet to be evaluated respectively at fixed points  $\vec{s}_1^*$  and  $\vec{s}_2^*$  yielded by the closed loop system and taking into account the total amount of external forcing, so we write:

$$\begin{cases} \mathcal{A}_{sn_i, x_{E_i}} = \Phi n_E(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \\ \mathcal{A}_{sg_i, x_{E_i}} = \Phi g_E(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \\ \mathcal{A}_{sn_i, x_{I_i}} = \Phi n_I(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \\ \mathcal{A}_{sg_i, x_{I_i}} = \Phi g_I(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \end{cases} \quad (92)$$

where  $xn_{inter,i}^*$  and  $x_I = xg_{inter,i}^*$  are defined in Eq 69. Analytical expression for the functions  $\Phi n_E$ ,  $\Phi g_E$ ,  $\Phi n_I$  and  $\Phi g_I$  are explicitly given in section 2.4. 171  
172

### 3.3.4 Closed loop sensitivities as functions of Open Loop Sensitivities 173

From definition 84 for  $\overline{\overline{G}}$ , we have: 174

$$\overline{\overline{G}} = \begin{pmatrix} \mathcal{A}_{sn_1, x_{E_1}} & 0 & \mathcal{A}_{sn_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sn_2, x_{E_2}} & 0 & \mathcal{A}_{sn_2, x_{I_2}} \\ \mathcal{A}_{sg_1, x_{E_1}} & 0 & \mathcal{A}_{sg_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sg_2, x_{E_2}} & 0 & \mathcal{A}_{sg_2, x_{I_2}} \end{pmatrix} \begin{bmatrix} 0 & k_{E_{12}}\kappa_{12} & 0 & 0 \\ k_{E_{21}}\kappa_{21} & 0 & 0 & 0 \\ 0 & (1 - k_{E_{12}})\kappa_{12} & 0 & 0 \\ (1 - k_{E_{21}})\kappa_{21} & 0 & 0 & 0 \end{bmatrix} \quad (93)$$

that we will write as:

$$\overline{\overline{G}} = \begin{bmatrix} 0 & G_{12} & 0 & 0 \\ G_{21} & 0 & 0 & 0 \\ 0 & G_{32} & 0 & 0 \\ G_{41} & 0 & 0 & 0 \end{bmatrix} \quad (94)$$

with 175

$$\begin{cases} G_{12} = \mathcal{A}_{sn_1, x_{E_1}} k_{E_{12}}\kappa_{12} + \mathcal{A}_{sn_1, x_{I_1}} (1 - k_{E_{12}})\kappa_{12} \\ G_{21} = \mathcal{A}_{sn_2, x_{E_2}} k_{E_{21}}\kappa_{21} + \mathcal{A}_{sn_2, x_{I_2}} (1 - k_{E_{21}})\kappa_{21} \\ G_{32} = \mathcal{A}_{sg_1, x_{E_1}} k_{E_{12}}\kappa_{12} + \mathcal{A}_{sg_1, x_{I_1}} (1 - k_{E_{12}})\kappa_{12} \\ G_{41} = \mathcal{A}_{sg_2, x_{E_2}} k_{E_{21}}\kappa_{21} + \mathcal{A}_{sg_2, x_{I_2}} (1 - k_{E_{21}})\kappa_{21} \end{cases} \quad (95)$$

For the l.h.s. term in expression 83, we then have: 176

$$\overline{\overline{\mathbb{I}}} - \overline{\overline{G}} = \left( \begin{array}{cc|cc} 1 & -G_{12} & 0 & 0 \\ -G_{21} & 1 & 0 & 0 \\ 0 & -G_{32} & 1 & 0 \\ -G_{41} & 0 & 0 & 1 \end{array} \right) \quad (96)$$

that we will write as: 177

$$\overline{\overline{\mathbb{I}}} - \overline{\overline{G}} = \left( \begin{array}{c|c} \overline{\overline{B_1}} & \overline{\overline{B_2}} \\ \hline \overline{\overline{B_3}} & \overline{\overline{B_4}} \end{array} \right) \quad (97)$$

Considering the property that 178

If  $\overline{\overline{M}} = \begin{pmatrix} \overline{\overline{A}} & \overline{\overline{B}} \\ \overline{\overline{C}} & \overline{\overline{D}} \end{pmatrix}$  with  $\overline{\overline{D}}$  invertible

$$\text{then } \overline{\overline{M}}^{-1} = \begin{pmatrix} \overline{\overline{R}} & \overline{\overline{S}} \\ \overline{\overline{T}} & \overline{\overline{U}} \end{pmatrix} \text{ with } \begin{cases} \overline{\overline{R}} = (\overline{\overline{A}} - \overline{\overline{B}}\overline{\overline{D}}^{-1}\overline{\overline{C}})^{-1} \\ \overline{\overline{S}} = -\overline{\overline{R}}\overline{\overline{B}}\overline{\overline{D}}^{-1} \\ \overline{\overline{T}} = -\overline{\overline{D}}^{-1}\overline{\overline{C}}\overline{\overline{R}} \\ \overline{\overline{U}} = \overline{\overline{D}}^{-1}(\overline{\overline{I}} - \overline{\overline{C}}\overline{\overline{S}}) \end{cases}$$

we obtain

179

$$\begin{aligned} (\overline{\overline{I}} - \overline{\overline{G}})^{-1} &= \left( \overline{\overline{C}}_1 \middle| \overline{\overline{C}}_2 \right) \text{ where } \begin{cases} \overline{\overline{C}}_1 = (\overline{\overline{B}}_1 - \overline{\overline{B}}_2(\overline{\overline{B}}_4^{-1})\overline{\overline{B}}_3)^{-1} = \overline{\overline{B}}_1^{-1} \\ \overline{\overline{C}}_2 = -\overline{\overline{C}}_1\overline{\overline{B}}_2(\overline{\overline{B}}_4^{-1}) = \overline{\overline{0}} \\ \overline{\overline{C}}_3 = -\overline{\overline{B}}_4^{-1}\overline{\overline{B}}_3\overline{\overline{C}}_1 = -\overline{\overline{B}}_3\overline{\overline{C}}_1 = -\overline{\overline{B}}_3\overline{\overline{B}}_1^{-1} \\ \overline{\overline{C}}_4 = \overline{\overline{B}}_4^{-1}(\overline{\overline{I}} - \overline{\overline{B}}_3\overline{\overline{C}}_2) = \overline{\overline{I}} \end{cases} \quad (98) \\ &= \left( \overline{\overline{B}}_1^{-1} \middle| \overline{\overline{0}} \right) \quad (99) \\ &\quad \left( -\overline{\overline{B}}_3\overline{\overline{B}}_1^{-1} \middle| \overline{\overline{I}} \right) \end{aligned}$$

To recover sensitivities for the closed loop condition, we then consider:

180

$$(\overline{\overline{I}} - \overline{\overline{G}})^{-1}\overline{\overline{S}} = \left( \overline{\overline{B}}_1^{-1} \middle| \overline{\overline{0}} \right) \left( \overline{\overline{\mathcal{A}}}_{sn,x_E} \middle| \overline{\overline{\mathcal{A}}}_{sn,x_I} \right) \quad (100)$$

$$= \left( \overline{\overline{B}}_1^{-1}\overline{\overline{\mathcal{A}}}_{sn,x_E} \middle| \overline{\overline{B}}_1^{-1}\overline{\overline{\mathcal{A}}}_{sn,x_I} \right) \quad (101)$$

$$= \left( \overline{\overline{B}}_1^{-1}\overline{\overline{\mathcal{A}}}_{sn,x_E} + \overline{\overline{\mathcal{A}}}_{sg,x_E} \middle| \overline{\overline{B}}_1^{-1}\overline{\overline{\mathcal{A}}}_{sn,x_I} + \overline{\overline{\mathcal{A}}}_{sg,x_I} \right)$$

so that

181

$$\boxed{\frac{\vec{\delta s}_B}{\delta B} = (\overline{\overline{I}} - \overline{\overline{G}})^{-1}\overline{\overline{S}} \frac{\partial w_{inter}}{\partial B}} \quad (102)$$

expresses, in full generality for the two-area system, the matrix of sensitivities to a perturbation upon either forcing  $B$ , and they are expressed as functions of the sensitivities in single-area system, which are in turn expressed as functions of the sensitivities in the single-pool system.

182

183

184

185

### 3.4 Sensitivities in a Target-Control system

186

#### 3.4.1 Extracting the case of interest

187

We now focus on the question of how sensitivities would drive the response of the excitatory pool of one area to the activation of the excitatory pool of the other one,

188

189



depending on the connectivity between the two areas. Hence, we attribute role to each area: the area 1 which excitatory pool is positively perturbed will be called "Control" area (denoted by C), and the area 2 will be called "Target area" (denoted by T).

From now on, the observable will then be denoted as:

$$\vec{s} = \begin{pmatrix} sn_C \\ sn_T \\ sg_C \\ sg_T \end{pmatrix} \quad (103)$$

and the transfer variables as:

$$\vec{x} = \begin{pmatrix} xn_{intra,C} = B_{E_C} + k_{E_{CT}}\kappa_{CT}sn_T \\ xn_{intra,T} = B_{E_T} + k_{E_{TC}}\kappa_{TC}sn_C \\ xg_{intra,C} = B_{I_C} + (1 - k_{E_{CT}})\kappa_{CT}sn_T \\ xg_{intra,T} = B_{I_T} + (1 - k_{E_{TC}})\kappa_{TC}sn_C \end{pmatrix} \quad (104)$$

and we focus on:

$$\begin{pmatrix} \delta sn_T \\ \delta sn_C \end{pmatrix} \quad (105)$$

in response to  $\delta B_{E_C}$ .

To extract the situation of interest, from the general result above, we then pick the case:

$$\vec{\delta s}_{B_{E_C}} = (\bar{\mathbb{I}} - \bar{G})^{-1} \bar{S} \frac{\partial w_{inter}}{\partial B_{E_C}} \delta B_{E_C} \quad (106)$$

$$= \left( \begin{array}{c|c} \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_E} & \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_I} \\ \hline -\bar{B}_3 \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_E} + \bar{\mathcal{A}}_{sg,x_E} & -\bar{B}_3 \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_I} + \bar{\mathcal{A}}_{sg,x_I} \end{array} \right) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \delta B_{E_C} \quad (107)$$

$$= \left( \begin{array}{c} \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_E} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \hline -\bar{B}_3 \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_E} + \bar{\mathcal{A}}_{sg,x_E} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{array} \right) \delta B_{E_C} \quad (108)$$

$$\equiv \begin{pmatrix} \delta sn_C \\ \delta sn_T \\ \delta sg_C \\ \delta sg_T \end{pmatrix} \quad (109)$$

where the perturbations of interest are in the upper part, and we have:

$$\begin{pmatrix} \delta sn_C \\ \delta sn_T \end{pmatrix} = \bar{B}_1^{-1} \bar{\mathcal{A}}_{sn,x_E} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \delta B_{E_C} \quad (110)$$

From definition 97, we get

$$\overline{\overline{B}}_1^{-1} = \begin{pmatrix} 1 & -G_{12} \\ -G_{21} & 1 \end{pmatrix}^{-1} = \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} 1 & G_{12} \\ G_{21} & 1 \end{pmatrix} \quad (111)$$

hence

202

$$\begin{pmatrix} \delta sn_C \\ \delta sn_T \end{pmatrix} = \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} 1 & G_{12} \\ G_{21} & 1 \end{pmatrix} \begin{pmatrix} \mathcal{A}_{sn_C, x_{E_C}} & 0 \\ 0 & \mathcal{A}_{sn_T, x_{E_T}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \delta B_{E_C} \quad (112)$$

$$= \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} \mathcal{A}_{sn_C, x_{E_C}} & G_{12}\mathcal{A}_{sn_T, x_{E_T}} \\ G_{21}\mathcal{A}_{sn_C, x_{E_C}} & \mathcal{A}_{sn_T, x_{E_T}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \delta B_{E_C} \quad (113)$$

$$= \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} \mathcal{A}_{sn_C, x_{E_C}} \\ G_{21}\mathcal{A}_{sn_C, x_{E_C}} \end{pmatrix} \delta B_{E_C} \quad (114)$$

with

203

$$\begin{cases} G_{12} = \mathcal{A}_{sn_C, x_{E_C}} k_{E_{CT}} \kappa_{CT} + \mathcal{A}_{sn_C, x_{I_C}} (1 - k_{E_{CT}}) \kappa_{CT} \\ G_{21} = \mathcal{A}_{sn_T, x_{E_T}} k_{E_{TC}} \kappa_{TC} + \mathcal{A}_{sn_T, x_{I_T}} (1 - k_{E_{TC}}) \kappa_{TC} \end{cases} \quad (115)$$

For the sake of clarity,  $G_{12}$  and  $G_{21}$  are denoted  $G_{CT}$  and  $G_{TC}$  in the main text.

204

Table 1: Set of parameters.

| Parameter    | Value       | Unit            |
|--------------|-------------|-----------------|
| $\beta^E$    | 6.6         | $s^{-1}$        |
| $\alpha^E$   | 72          | $s^{-1}mM^{-1}$ |
| $T_{glu}$    | 0.008       | $mM\ s$         |
| $\beta^I$    | 180         | $s^{-1}$        |
| $\alpha^I$   | 530         | $s^{-1}mM^{-1}$ |
| $T_{gaba}$   | 0.003       | $mM\ s$         |
| $a_E$        | 310         | $nC^{-1}$       |
| $b_E$        | 125         | $Hz$            |
| $d_E$        | 0.16        | $s$             |
| $a_I$        | 615         | $nC^{-1}$       |
| $b_I$        | 177         | $Hz$            |
| $d_I$        | 0.087       | $s$             |
| $W_E$        | 1           | —               |
| $W_I$        | 0.7         | —               |
| $I_0$        | 0.382       | $nA$            |
| $W_+$        | 1.4         | —               |
| $J_{nmda}$   | 0.15        | $nA$            |
| $J_{gaba,i}$ | [0.25;2.75] | $nA$            |
| $G$          | 0.69        | —               |
| $J_-$        | 1           | $nA$            |

## Quatrième partie

### Inversion

Dans cette partie, je présente au chapitre 10 la procédure d'inversion qui permet d'estimer les paramètres neuraux à partir de données BOLD obtenues pendant une tâche de contrôle inhibiteur. Celle-ci bénéficie de l'éclairage apporté par le travail analytique des chapitres 7 et 9. Le chapitre 11 expose ensuite sa mise en oeuvre sur les données d'un sujet effectuant, plus particulièrement, une tâche de contrôle inhibiteur des intrusions mnésiques, avec le paradigme *Think/No Think*. Ce sujet est un participant sain non exposé de l'étude REMEMBER.

Dans la partie III, nous avons travaillé avec deux régions : une région de contrôle et une région cible. Nous étendons ici le nombre de régions qui composent le système, et nous nous donnons la possibilité de travailler avec un groupe de régions de contrôle et un groupe de régions cibles. L'étude de la connectivité, présentée au chapitre 9, permet en effet de choisir une architecture de connectivité dans ce nouveau système étendu, pour abaisser l'activité des régions cibles quand les régions de contrôle sont stimulées.

# Chapitre 10

## Construction de la procédure d'inversion

*Le travail analytique mené au chapitre 7 a permis d'énoncer la possibilité de construire une procédure d'inversion en deux étapes. Je présente à la section 10.1 le modèle direct sous l'hypothèse duquel nous menons une estimation de paramètres neuraux  $J_{gaba}$  et  $I_{stim}$  dans chaque région, à partir de signaux BOLD. A la section 10.2, je présente les objectifs de chacune des deux étapes.*

*La première étape, qui repose sur une inversion bayésienne variationnelle des modèles hémodynamique et BOLD, sera présentée en détails au 10.3, puis validée avec des signaux BOLD simulés au 10.4. Dans chaque région, elle permet de remonter des décours temporels du signal BOLD aux activités neurales  $sn^*$  au repos et en période de stimulation. Ce problème d'inversion est bien posé puisqu'il existe une bijection entre  $sn^*$  et signal BOLD (montrée à la section 7.1.2). Cette étape ne nécessite pas de disposer d'une matrice de connectivité structurale ( $\bar{\bar{C}}$ ).*

*La deuxième étape, qui repose sur une inversion algébrique-numérique du modèle neural au régime stationnaire, est détaillée à la section 10.6. Elle permet de remonter des activités neurales  $sn^*$  au repos et en période de stimulation, aux paramètres neuraux  $J_{gaba}$  et  $I_{stim}$ , en tenant compte d'une matrice de connectivité structurale ( $\bar{\bar{C}}$ ). Elle nécessite également de fixer une architecture de connectivité ( $\bar{\bar{k}}_E$ ) qui permette d'abaisser l'activité des régions cibles quand l'activité des régions de contrôle augmente. Le travail du chapitre 9 permet de choisir une telle architecture de connectivité.*

*La méthode d'analyse des sensibilités neurales, présentée au chapitre 9, a permis de construire un emboîtement de sensibilités, qui part du pool inhibiteur et du pool excitateur de chaque région (plus petit niveau d'emboîtement), pour exprimer les sensibilités dans le système couplé. Dans le présent chapitre, l'étape d'inversion du modèle neural est menée de manière découplée en revenant au pool inhibiteur puis au pool excitateur de chaque région.*

## 10.1 Le modèle direct

Nous nous donnons la possibilité de travailler avec plusieurs régions de contrôle et plusieurs régions cibles couplées entre elles.

### 10.1.1 Deux paramètres d'intérêt par région : $J_{gaba_i}$ et $I_{stim_i}$

Jusqu'ici, nous avons modélisé une stimulation par une augmentation du forçage sur le pool excitateur uniquement dans les régions de contrôle (cf section 6.2.4). Ceci se traduit mathématiquement par un paramètre  $I_{stim_i}$  dans l'équation qui donne le courant  $xn_i$  d'une région de contrôle  $i$ . Ce modèle ne prévoit cependant pas de  $I_{stim_i}$  dans une région cible.

Nous le modifions pour ajouter  $I_{stim_i}$  dans chaque région cible et pouvoir en estimer la valeur. Ceci nous permet de vérifier à partir des données, la pertinence du modèle de stimulation décrit ci-dessus. Si  $I_{stim_i} = 0$  dans les régions cibles et  $I_{stim_i} \neq 0$  dans les régions de contrôle, alors ce modèle de stimulation permet de rendre compte des données.

Finalement, pour chaque région  $i$ , nous allons inverser deux paramètres du modèle neural :  $J_{gaba_i}$  et  $I_{stim_i}$ . Cette région peut être une région de contrôle ou une région cible.

### 10.1.2 Formulation mathématique du modèle complet

Le modèle direct complet utilisé pour analyser les données BOLD s'écrit alors comme suit. En distinguant les blocs, nous avons :

1. La modèle neural au régime stationnaire :

$$\begin{cases} 0 = -\beta^E sn_i^* + \alpha^E T_{glu} rn_i^* (1 - sn_i^*) \\ 0 = -\beta^I sg_i^* + \alpha^I T_{gaba} rg_i^* (1 - sg_i^*) \end{cases} \quad (10.1)$$

$$\begin{cases} rn_i^* = \frac{a_E xn_i^* - b_E}{1 - e^{-d_E(a_E xn_i^* - b_E)}} \\ rg_i^* = \frac{a_I xg_i^* - b_I}{1 - e^{-d_I(a_I xg_i^* - b_I)}} \end{cases} \quad (10.2)$$

$$\begin{cases} xn_i^* = W_+ J_{nmda} sn_i^* - J_{gaba,i} sg_i^* + W_E I_0 + I_{stim_i} u(t) + G J_{nmda} \sum_{j \neq i} k_{Eij} C_{ij} sn_j^* \\ xg_i^* = J_{nmda} sn_i^* - W_- sg_i^* + W_I I_0 + G J_{nmda} \sum_{j \neq i} (1 - k_{Eij}) C_{ij} sn_j^* \end{cases} \quad (10.3)$$

La fonction créneau  $u$  permet de faire varier le forçage sur les pools excitateurs en période de stimulation. Elle est définie telle que :

$$u(t) = \begin{cases} 1 & \text{en période de stimulation} \\ 0 & \text{hors période de stimulation} \end{cases} \quad (10.4)$$

La matrice  $\overline{\overline{k_E}}$  permet de fixer une architecture de connectivité. Nous la choisissons de manière à pouvoir obtenir une baisse d'activité dans les cibles quand les régions de contrôle sont stimulées. Les valeurs des éléments de la matrice sont reportées dans le tableau 10.1. En cohérence avec les résultats de l'article du chapitre 9, une région de contrôle et une région cible sont connectées entre elles par une connexion de type E-I. Deux régions de contrôle ou deux régions cibles sont connectées entre elles par une connexion E-E.

| Connexion qui arrive vers  | Connexion qui part de      |                      |
|----------------------------|----------------------------|----------------------|
|                            | Une région de contrôle $j$ | Une région cible $j$ |
| Une région de contrôle $i$ | $k_{E_{ij}} = 1$           | $k_{E_{ij}} = 1$     |
| Une région cible $i$       | $k_{E_{ij}} = 0$           | $k_{E_{ij}} = 1$     |

**Table 10.1** – Elements de la matrice  $\overline{\overline{k_E}}$  qui fixe l'architecture de connectivité pour un groupe de régions de contrôle et un groupe de régions cibles

2. Le modèle hémodynamique et BOLD :

$$\begin{cases} \dot{z}_i(t) = sn_i^* - \kappa z_i(t) - \gamma(f_i(t) - 1) \\ \dot{f}_i(t) = z_i(t) \\ \dot{v}_i(t) = \frac{1}{\tau} \left( f_i(t) - v_i(t)^{\frac{1}{\alpha}} \right) \\ \dot{q}_i(t) = \frac{f_i(t)}{\tau \rho} \left( 1 - (1 - \rho)^{\frac{1}{f_i(t)}} \right) - \frac{1}{\tau} q_i(t) v_i(t)^{\frac{1}{\alpha} - 1} \end{cases} \quad (10.5)$$

$$y_i(t) = V_0 \left( k_1(1 - q_i(t)) + k_2(1 - q_i(t)v_i(t)^{-1}) + k_3(1 - v_i(t)) \right) \quad (10.6)$$



## 10.2 Présentation du découplage en 2 étapes

La figure 10.1 présente les deux étapes de la procédure d'inversion, que nous construisons plus en détails dans la suite de ce chapitre.

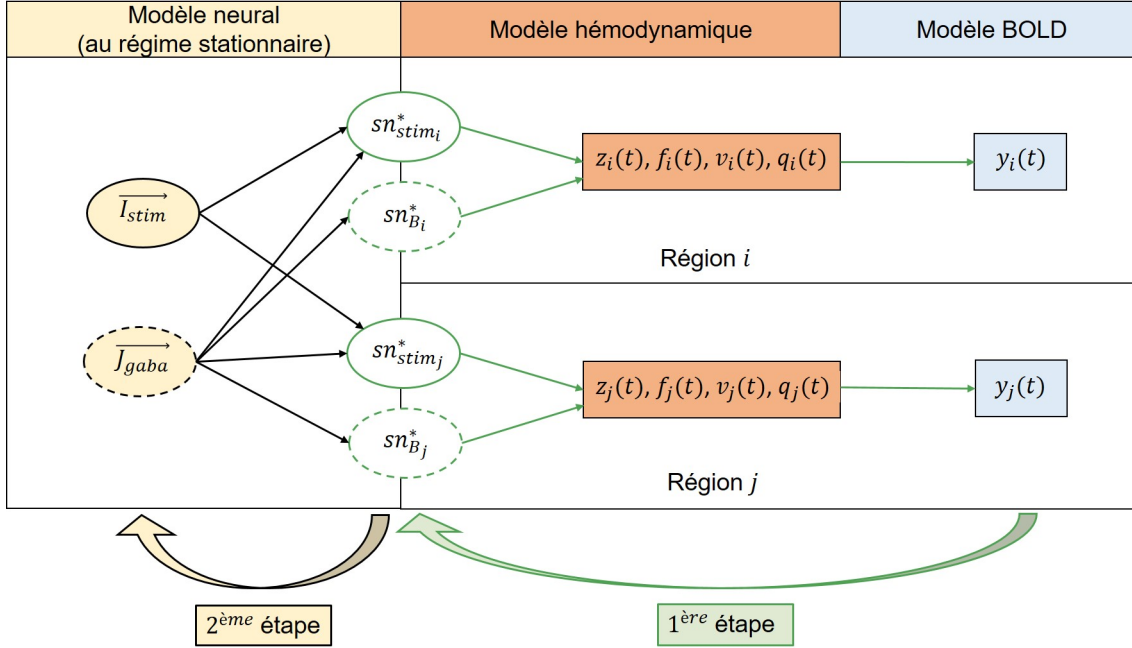


**Figure 10.1** – Présentation schématique des deux étapes de l'inversion. Les flèches de gauche à droite représentent des dépendances du modèle direct. Les flèches de droite à gauche représentent la première et deuxième étape de l'inversion, en indiquant les éléments en entrée et en sortie.  
 $\vec{J}_{gaba} = (J_{gaba_i}, J_{gaba_j})$  et  $\vec{I}_{stim} = (I_{stim_i}, I_{stim_j})$

### 10.2.1 Première étape

Pour chaque région, nous estimons les couples des valeurs basale et stimulée  $(sn_B^*, sn_{stim}^*)$ . Ces points fixes (stables) sont traités comme des paramètres du modèle hémodynamique. Il n'existe pas de couplages entre régions au niveau des modèles hémodynamiques et BOLD : on parle de modèles découplés entre régions. L'inversion peut alors être menée région par région : la méthode d'inversion est dite découplée.

Pour estimer les couples  $(sn_B^*, sn_{stim}^*)$ , nous utilisons une approche d'inversion bayésienne variationnelle ([Daunizeau et al., 2009]). Elle permet de prendre en compte la dynamique du modèle direct (i.e. modèles hémodynamique et BOLD) pour inverser les paramètres à partir des décours temporels mesurés du signal BOLD. Pour sa mise en oeuvre numérique, nous avons travaillé avec des codes de la toolbox Variational Bayesian Analysis (VBA, <https://mbb-team.github.io/VBA-toolbox/>, [Daunizeau et al., 2014]).

### 10.2.2 Deuxième étape

A partir des valeurs de  $sn_B^*$ ,  $sn_{stim}^*$  de toutes les régions, nous inversons les paramètres neuraux d'intérêt  $I_{stim}$  et  $J_{gaba}$  dans chaque région (comme représenté sur la figure 10.1). Il existe des couplages entre régions au niveau du modèle neural : le modèle est couplé entre régions. Néanmoins, nous construisons une procédure d'inversion découplée dans laquelle l'inversion des paramètres neuraux peut se faire région par région, en considérant successivement dans chaque région le pool inhibiteur puis le pool excitateur. Toutes les valeurs de  $sn_B^*$  sont utilisées pour inverser  $J_{gaba}$  dans chacune des régions. Toutes les valeurs des  $sn_{stim}^*$  permettent d'estimer  $I_{stim}$  dans chacune des régions.

## 10.3 Première étape de l'inversion

### 10.3.1 Principe de l'approche bayésienne

#### 10.3.1.1 Exemple de l'estimation de paramètres

Soit un jeu de données observées ( $y$ ). Nous souhaitons analyser ces données à la lumière d'un modèle statistique ( $m$ ). Pour cela, nous cherchons à estimer certains paramètres ( $\theta$ ) du modèle. L'approche bayésienne permet d'estimer les paramètres ( $\theta$ ) du modèle statistique ( $m$ ) à partir de nos connaissances a priori sur les valeurs probables des paramètres et du jeu de données observées ( $y$ ).

Nous notons  $p(\theta|y, m)$  la densité de probabilité des paramètres  $\theta$ , pour le modèle statistique  $m$ , sachant les données  $y$ . Le théorème de Bayes permet d'écrire :

$$p(\theta|y, m) = \frac{p(y|\theta, m)p(\theta|m)}{p(y|m)} \quad (10.7)$$

$p(\theta|y, m)$  est nommée densité de probabilité a posteriori, pour le modèle statistique  $m$ , sachant les données  $y$ . Elle dépend de :

- $p(\theta|m)$  : la densité de probabilité a priori sur le paramètre  $\theta$ . Cette distribution exprime les connaissances a priori sur les valeurs probables des paramètres.

Elles peuvent être plus ou moins informatives (i.e. a priori certaines valeurs sont rendues plus probables que d'autres, ou non). Plus la densité de probabilité a priori est informative et plus elle influence la distribution a posteriori  $p(\theta|y, m)$ <sup>1</sup>. Le choix des a priori pertinents apparaît ainsi comme une question importante dans un problème d'inversion bayésienne.

Différentes manières permettent de fixer ces distributions a priori<sup>2</sup> :

- méthode empirique, qui utilise par exemple des mesures indépendantes de travaux précédents
  - analyse du comportement du modèle pour sélectionner des intervalles de paramètres qui permettent de produire des observables intéressantes pour la question de recherche
  - le domaine de définition des paramètres conduit à exclure certains intervalles impossibles par construction (par exemple, certains paramètres ne peuvent être négatifs ou supérieurs à 1)
- $p(y|\theta, m)$  : la vraisemblance (ou likelihood en anglais)<sup>3</sup>. C'est la densité de probabilité des données conditionnellement à une valeur des paramètres. Autrement dit, si on fixe les paramètres, on se pose la question de la probabilité d'obtenir le jeu de données  $y$ .
  - $p(y|m)$  : évidence du modèle (ou loi marginale ou fonction de partition), marginal likelihood en anglais. C'est la densité de probabilité d'observer ces données étant donné le modèle statistique que l'on s'est donné.

Cette distribution est notamment utile pour la comparaison de modèles : le modèle statistique qui permet de maximiser  $p(y|m)$  est considéré comme le plus probable.

De plus, elle permet de calculer la densité de probabilité des données pour plusieurs modèles possibles

$$p(y) = \sum_m p(y|m)p(m) \quad (10.8)$$

avec  $p(m)$  : densité de probabilité a priori du modèle statistique  $m$ . En général, des densités de probabilité a priori non informatives sont utilisées; et  $p(y)$  : densité de probabilité des données avec tous ces modèles possibles.

Dans la suite de cette thèse, nous ne travaillerons qu'avec un seul modèle statistique. Puisqu'il s'agira toujours du même, nous pouvons rendre implicite la notation " $|m$ " pour alléger les écritures. Nous noterons alors plus simplement :

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (10.9)$$

---

1. <https://mbb-team.github.io/VBA-toolbox/wiki/Structure-of-VBA's-generative-model/>

2. <https://mbb-team.github.io/VBA-toolbox/wiki/Bayesian-modelling-introduction/>

3. "statistical construct that is common to both frequentist and bayesian inference approach" <https://mbb-team.github.io/VBA-toolbox/wiki/Bayesian-modelling-introduction/>

L'évidence du modèle se calcule en sommant les densités de probabilité jointes  $p(\theta, y)$  pour toutes les valeurs de  $\theta$  possibles. Cela revient à intégrer le numérateur de (10.9) :

$$p(y) = \int_{\theta} p(\theta, y) d\theta = \int_{\theta} p(y|\theta) p(\theta) d\theta \quad (10.10)$$

Cette intégrale est souvent difficile à calculer et elle est nécessaire pour le calcul de la constante de normalisation de la loi a posteriori. Nous pourrions l'approcher par une méthode variationnelle. Ce n'est plus la grandeur exacte d'intérêt que nous allons calculer (comme l'évidence du modèle par ex.) mais une valeur approchée qui la minore. Plus précisément, les méthodes variationnelles maximisent un minorant de la grandeur exacte à déterminer ([Keribin, 2009]).

### 10.3.1.2 Exemple de l'estimation de paramètres et d'états cachés

Le modèle statistique peut s'exprimer en fonction d'états dits cachés ( $x$ ), c'est-à-dire non mesurables directement, mais qui permettent de générer les données observées. C'est par exemple le cas des variables d'état hémodynamiques qui permettent de générer le signal BOLD mesurable. Nous notons  $p(x, \theta|y)$  la densité de probabilité jointe a posteriori (pour les états cachés  $x$  et les paramètres  $\theta$ ), sachant les données. Dans ce cas-ci, la densité de probabilité jointe a posteriori est un peu plus complexe à écrire :

$$p(x, \theta|y) = \frac{p(y|x, \theta) p(x, \theta)}{p(y)} \quad (10.11)$$

$$= \frac{p(y|x, \theta) p(x|\theta) p(\theta)}{p(y)} \quad (10.12)$$

Les difficultés d'intégration peuvent survenir à deux niveaux. D'abord le calcul de l'évidence consiste en une double intégrale sur les états cachés et les paramètres :

$$p(y) = \int_x \int_{\theta} p(x, \theta, y) dx d\theta = \int_x \int_{\theta} p(y|x, \theta) p(x|\theta) p(\theta) dx d\theta \quad (10.13)$$

Ensuite, il est intéressant de pouvoir calculer les lois marginales a posteriori  $p(x|y)$  et  $p(\theta|y)$ . En particulier,  $p(\theta|y)$  est la densité de probabilité des paramètres  $\theta$  que nous souhaitons inverser à partir du jeu de données  $y$ . Elles s'écrivent comme une intégrale de la loi jointe a posteriori (respectivement sur  $\theta$  et  $x$ ) :

$$p(x|y) = \int_{\theta} p(x, \theta|y) d\theta \quad (10.14)$$

$$p(\theta|y) = \int_x p(x, \theta|y) dx \quad (10.15)$$

Le nombre d'intégrales à calculer augmente avec le nombre de variables aléatoires impliquées dans le modèle statistique. A nouveau, ces intégrales sont souvent difficiles à calculer et nous allons pouvoir les approximer en passant par du calcul variationnel.

L'approche bayésienne variationnelle (avec approximation de Laplace) va nous permettre d'approximer la log évidence  $\ln(p(y))$  par une borne inférieure et d'estimer une densité de probabilité approchante de  $p(x|y)$  et  $p(\theta|y)$  et de la loi jointe a posteriori  $p(x, \theta|y)$ . Nous verrons en détails comment à la section 10.3.5.

### 10.3.2 Réécriture des modèles hémodynamique et BOLD préalable à l'utilisation de la toolbox VBA

La toolbox VBA regroupe un ensemble de codes en langage Matlab qui permettent d'estimer par une approche bayésienne variationnelle des paramètres dans des modèles dynamiques non linéaires ([Daunizeau et al., 2014]).

Comme nous le verrons dans le paragraphe 10.3.5.2.2, la méthode d'inversion repose sur une approximation de Laplace qui contraint la forme des distributions sur les paramètres inversés, et également sur les variables d'état. Ce sont nécessairement des gaussiennes. Les paramètres que l'on cherche à estimer et les variables d'état du modèle dynamique doivent donc tous être définis sur  $\mathbb{R}$ .

Dans les modèles hémodynamique et BOLD présentés à la section 10.1.2, les paramètres  $sn_{B_i}^*$  et  $sn_{stim_i}^*$  ainsi que la plupart des variables hémodynamiques ( $f_i, v_i$  et  $q_i$ ) ne sont pas définies sur  $\mathbb{R}$ . Dans cette partie, nous présentons un travail de réécriture des équations des deux modèles. Nous détaillons d'abord le changement d'écriture des paramètres  $sn_{B_i}^*$  et  $sn_{stim_i}^*$  (10.3.2.2), puis celui des variables d'état hémodynamiques (10.3.2.3). Nous aboutissons alors à un modèle écrit avec des paramètres et variables que nous dirons "transformés", définis sur  $\mathbb{R}$ .

Ce travail préalable nous permettra ensuite de définir un modèle statistique/probabiliste de la mesure (que nous détaillerons à la section 10.3.4). L'objectif principal sera alors d'estimer les distributions de probabilité a posteriori sur les paramètres transformés.

#### 10.3.2.1 Modèle hémodynamique et BOLD non transformés

Pour commencer, nous ré-écrivons le modèle hémodynamique pour rendre explicites  $sn_{B_i}^*$  et  $sn_{stim_i}^*$ , les deux paramètres que l'on souhaite inverser, dans l'équation différentielle qui régit l'évolution temporelle du signal de vasodilatation.

Pour chaque région  $i$  :

$$\begin{cases} \dot{z}_i(t) = sn_{B_i}^* \delta_{u(t),0} + sn_{stim_i}^* \delta_{u(t),1} - \kappa z_i(t) - \gamma(f_i(t) - 1) \equiv h z_i(sn_{B_i}^*, sn_{stim_i}^*, z_i, f_i) \\ \dot{f}_i(t) = z_i(t) \equiv h f_i(z_i) \\ \dot{v}_i(t) = \frac{1}{\tau} (f_i(t) - v_i(t)^{\frac{1}{\alpha}}) \equiv h v_i(f_i, v_i) \\ \dot{q}_i(t) = \frac{f_i(t)}{\tau \rho} \left(1 - (1 - \rho)^{\frac{1}{f_i(t)}}\right) - \frac{1}{\tau} q_i(t) v_i(t)^{\frac{1}{\alpha}-1} \equiv h q_i(f_i, q_i, v_i) \end{cases} \quad (10.16)$$

$$\text{Avec } \begin{cases} u(t) = 0 \text{ en période de non stimulation} \\ u(t) = 1 \text{ en période de stimulation} \end{cases} \quad (10.17)$$

Le symbole  $\delta_{u(t),0}$  (resp.  $\delta_{u(t),1}$ ) est le delta de Kronecker qui est égal à 1 si  $u(t) = 0$  (resp.  $u(t) = 1$ ).  $sn_{B_i}^*$  et  $sn_{stim_i}^*$  permettent d'appliquer des valeurs différentes du forçage (toujours comprises entre 0 et 1) sur le signal de vasodilatation de chaque région, en fonction

des périodes de stimulation ou de non stimulation.

Le signal BOLD est une fonction de  $v_i$  et  $q_i$  :

$$y_i = V_0 \left( k_1(1 - q_i) + k_2(1 - q_i v_i^{-1}) + k_3(1 - v_i) \right) \equiv B_i(v_i, q_i) \quad (10.18)$$

$$\text{Avec les paramètres : } \begin{cases} k1 = 4.3\theta_0\rho TE \\ k2 = \epsilon r_0\rho TE \\ k3 = 1 - \epsilon \end{cases} \quad (10.19)$$

Nous précisons les domaines de définition des variables d'état hémodynamiques :

- Signal de vasodilatation :  $z_i \in \mathbb{R}$
- Flux :  $f_i \in \mathbb{R}^+$
- Volume :  $v_i \in \mathbb{R}^+$
- Concentration en déoxyhémoglobine :  $q_i \in \mathbb{R}^+$

### 10.3.2.2 Changement d'écriture des paramètres à inverser

Les deux paramètres  $sn_{B_i}^*$  et  $sn_{stim_i}^*$  appartiennent à l'intervalle  $[0; 1]$ . Nous avons besoin de définir une bijection de  $\mathbb{R}$  dans  $[0; 1]$  pour pouvoir écrire les paramètres "transformés" sur lesquels porteront l'inversion. Ceux-ci seront respectivement notés  $sn_{B_i}$  et  $sn_{stim_i}$ .

Soit  $\Phi$  une fonction sigmoïde bijective de  $\mathbb{R}$  dans  $[0; 1]$  :

$$\Phi : \mathbb{R} \mapsto [0; 1] \quad (10.20)$$

$$\tilde{sn} \mapsto \frac{1}{1 + \exp(-(\tilde{sn} + \ln(\frac{s_0}{1-s_0})))} \quad (10.21)$$

$s_0$  est la valeur de  $\Phi(\tilde{sn})$  lorsque  $\tilde{sn} = 0$ .

Nous pouvons écrire  $sn_{B_i}$  (resp.  $sn_{stim_i}$ ) comme une fonction sigmoïde de  $sn_{B_i}$  (resp.  $sn_{stim_i}$ ) définie sur  $\mathbb{R}$ <sup>4</sup> :

$$\begin{cases} sn_{B_i} \equiv \Phi(sn_{B_i}) = \frac{1}{1 + \exp(-(\tilde{sn}_{B_i} + \ln(\frac{s_0}{1-s_0})))} \\ sn_{stim_i} \equiv \Phi(sn_{stim_i}) = \frac{1}{1 + \exp(-(\tilde{sn}_{stim_i} + \ln(\frac{s_0}{1-s_0})))} \end{cases} \quad (10.22)$$

$s_0$  est la valeur de  $sn_{B_i}$  (resp.  $sn_{stim_i}$ ) lorsque  $\tilde{sn}_{B_i} = 0$  (resp.  $\tilde{sn}_{stim_i} = 0$ ).

---

4. Cette transformation sigmoïdale est identique à celle proposée dans les codes de démonstration de la toolbox VBA (<https://mbb-team.github.io/VBA-toolbox/>) pour le paramètre  $\rho$  du modèle hémodynamique (qui représente une fraction) et sa version transformée  $\tilde{\rho}$ .

Nous allons travailler avec des densités de probabilité a priori gaussiennes de ces paramètres transformés (cf 10.3.4.1.2), identiques dans toutes les régions et indépendantes entre elles. Le modèle initial, à partir duquel nous savons réfléchir, s'exprime avec des paramètres non transformés, il est donc utile de calculer les lois correspondantes des paramètres non transformés pour savoir fixer les densités a priori.

Pour toute région  $i$ , supposons  $\tilde{sn}_{B_i}$  et  $\tilde{sn}_{stim_i}$  deux variables aléatoires à valeurs dans  $\mathbb{R}$ .  $sn_{B_i}$  et  $sn_{stim_i}$  sont deux variables aléatoires à valeurs dans  $[0; 1]$ .

Nous définissons  $sn_{B_i}$  et  $sn_{stim_i}$  respectivement en fonction de  $\tilde{sn}_{B_i}$  et  $\tilde{sn}_{stim_i}$  :

$$\begin{cases} sn_{B_i} = \Phi(\tilde{sn}_{B_i}) \\ sn_{stim_i} = \Phi(\tilde{sn}_{stim_i}) \end{cases} \quad (10.23)$$

$\tilde{sn}_{B_i}$  et  $\tilde{sn}_{stim_i}$  sont deux variables aléatoires indépendantes qui suivent respectivement une loi normale centrée en 0, de variance  $\sigma_\theta^2$  :  $\mathcal{N}(\mu_\theta = 0, \sigma_\theta^2)$ . Nous choisissons une loi a priori identique pour ne privilégier aucune relation d'ordre entre les valeurs de  $\tilde{sn}_{B_i}$  et  $\tilde{sn}_{stim_i}$ , et donc de  $sn_{B_i}$  et  $sn_{stim_i}$  dans chaque région et entre les régions. Ceci nous permettra de vérifier si, à partir de données BOLD, nous trouvons bien une augmentation d'activité neurale dans les régions de contrôle et une baisse dans les cibles (cf section 11.1).

D'après l'annexe 17.1.2, nous pouvons déduire la loi correspondante de  $sn_{B_i}$  et  $sn_{stim_i}$ , pour toute région  $i$ .

Avec  $\mu_\theta = 0$ ,  $sn_{B_i}$  suit la densité de probabilité définie par :

$$\forall sn_{B_i} \in [0, 1], p(sn_{B_i}) = \frac{1}{(1 - sn_{B_i})sn_{B_i}\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{[\ln(\frac{sn_{B_i}(1-s_0)}{(1-sn_{B_i})s_0})]^2}{2\sigma_\theta^2}\right) \quad (10.24)$$

$sn_{stim_i}$  suit la même densité de probabilité qui peut s'écrire :

$$\forall sn_{stim_i} \in [0, 1], p(sn_{stim_i}) = \frac{1}{(1 - sn_{stim_i})sn_{stim_i}\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{[\ln(\frac{sn_{stim_i}(1-s_0)}{(1-sn_{stim_i})s_0})]^2}{2\sigma_\theta^2}\right) \quad (10.25)$$

$s_0$  est la valeur de  $sn_{B_i}$  et  $sn_{stim_i}$  lorsque  $\tilde{sn}_{B_i} = 0$  et  $\tilde{sn}_{stim_i} = 0$ . Nous fixons,  $s_0 = 0.5$ , milieu de l'intervalle des valeurs  $sn_{B_i}$  et  $sn_{stim_i}$  possibles (entre 0 et 1).

Nous trouverons à la section 10.4, figure 10.7, une représentation graphique des lois de probabilités que suivent les variables aléatoires transformées et non transformées avec une variance  $\sigma_\theta^2 = 3$ .

### 10.3.2.3 Changement d'écriture des variables d'état

#### 10.3.2.3.1 Définition des changement de variable

Les variables d'état  $f_i$ ,  $v_i$  et  $q_i$  sont définies sur  $\mathbb{R}^+$ . Il est nécessaire de définir une bijection de  $\mathbb{R}$  dans  $\mathbb{R}^+$  pour nous permettre de travailler avec des variables d'état "transformées" sur lesquelles porteront l'inversion.

Soit la bijection :

$$\mathcal{C} : \mathbb{R} \mapsto \mathbb{R}^+ \quad (10.26)$$

$$\tilde{x} \mapsto e^{\tilde{x}} \quad (10.27)$$

Nous définissons les variables d'état  $f_i$ ,  $v_i$  et  $q_i$  comme fonction des variables transformées  $\tilde{f}_i$ ,  $\tilde{v}_i$  et  $\tilde{q}_i$  définies sur  $\mathbb{R}$  :

$$\begin{cases} f_i \equiv \mathcal{C}(\tilde{f}_i) = e^{\tilde{f}_i} \\ v_i \equiv \mathcal{C}(\tilde{v}_i) = e^{\tilde{v}_i} \\ q_i \equiv \mathcal{C}(\tilde{q}_i) = e^{\tilde{q}_i} \end{cases} \quad (10.28)$$

Nous allons travailler avec des densités de probabilité a priori gaussiennes de ces variables transformées (a priori sur la condition initiale, cf 10.3.4.2.2.2), identiques dans toutes les régions et indépendantes entre elles. Le modèle initial s'exprimant avec des variables non transformées, il est utile de calculer les lois correspondantes des variables non transformées pour savoir fixer les densités a priori.

Pour toute région  $i$ , soient  $\tilde{F}_i$ ,  $\tilde{V}_i$  et  $\tilde{Q}_i$  trois variables aléatoires indépendantes qui suivent une loi normale  $\mathcal{N}(\mu_0 = 0, \sigma_0^2)$ . Nous définissons  $F_i$ ,  $V_i$  et  $Q_i$  trois autres variables aléatoires à valeurs dans  $[0; 1]$ , chacune définies en fonction respectivement de  $\tilde{F}_i$ ,  $\tilde{V}_i$  et  $\tilde{Q}_i$ .

$$\begin{cases} F_i = \mathcal{C}(\tilde{F}_i) = e^{\tilde{F}_i} \\ V_i = \mathcal{C}(\tilde{V}_i) = e^{\tilde{V}_i} \\ Q_i = \mathcal{C}(\tilde{Q}_i) = e^{\tilde{Q}_i} \end{cases} \quad (10.29)$$

Nous pouvons alors déduire la densité de probabilité correspondante que suivent respectivement  $F_i$ ,  $V_i$  et  $Q_i$ , pour toute région  $i$ . D'après l'annexe 17.1.3, avec  $\mu_0 = 0$ , elle est définie par la loi lognormale :

$$\forall x \in \mathbb{R}^+, p(x) = \frac{1}{x\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\ln(x))^2}{2\sigma_0^2}\right) \quad (10.30)$$

Nous trouverons à la section 10.4, figure 10.9, une représentation graphique de ces densités de probabilités, avec une variance égale à  $\sigma_0^2 = 10^{-4}$ .

### 10.3.2.3.2 Modèle hémodynamique transformé

Nous allons avoir besoin de la dynamique temporelle des variables d'états transformées. Pour cela, nous avons besoin de réécrire le modèle de la section 10.3.2.1.

1) Equation différentielle sur le signal de vasodilatation :

$$\dot{z}_i(t) = sn_{B_i}\delta_{u(t),0} + sn_{stim_i}\delta_{u(t),1} - \kappa z_i(t) - \gamma(f_i(t) - 1) = h z_i(sn_{B_i}, sn_{stim_i}, z_i, f_i) \quad (10.31)$$

Avec les paramètres et variables transformés, nous avons :



$$\dot{z}_i(t) = \Phi(s\tilde{n}_{B_i})\delta_{u(t),0} + \Phi(s\tilde{n}_{stim_i})\delta_{u(t),1} - \kappa z_i(t) - \gamma(e^{\tilde{f}_i(t)} - 1) \equiv h\tilde{z}_i(s\tilde{n}_{B_i}, s\tilde{n}_{stim_i}, z_i, \tilde{f}_i) \quad (10.32)$$

2) Equation différentielle sur le flux :

Nous partons de :

$$\dot{f}_i(t) = z_i(t) \equiv h f_i(z_i) \quad (10.33)$$

Or

$$\dot{f}_i = \frac{df_i}{dt} = \frac{d(e^{\tilde{f}_i})}{dt} = \frac{d\tilde{f}_i}{dt} e^{\tilde{f}_i} = \dot{\tilde{f}_i} e^{\tilde{f}_i} \quad (10.34)$$

Donc l'équation différentielle sur le flux peut se réécrire comme :

$$\dot{\tilde{f}_i}(t) = z_i(t) e^{\tilde{f}_i(t)} \equiv h\tilde{f}_i(z_i, \tilde{f}_i) \quad (10.35)$$

3) Equation différentielle sur le volume :

Nous partons de :

$$\dot{v}_i(t) = \frac{1}{\tau} \left( f_i(t) - v_i(t)^{\frac{1}{\alpha}} \right) \equiv h v_i(f_i, v_i) \quad (10.36)$$

En écrivant :

$$\dot{v}_i = \dot{\tilde{v}_i} e^{\tilde{v}_i} \quad (10.37)$$

Nous pouvons en déduire :

$$\dot{\tilde{v}_i}(t) = \frac{1}{\tau} \left( e^{\tilde{f}_i(t) - \tilde{v}_i(t)} - e^{(\frac{1}{\alpha} - 1)\tilde{v}_i(t)} \right) \equiv h\tilde{v}_i(\tilde{f}_i, \tilde{v}_i) \quad (10.38)$$

4) Equation différentielle sur la quantité de déoxyhémoglobine :

Nous partons de :

$$\dot{q}_i(t) = \frac{f_i(t)}{\tau\rho} \left( 1 - (1 - \rho)^{\frac{1}{\tilde{f}_i(t)}} \right) - \frac{1}{\tau} q_i(t) v_i(t)^{\frac{1}{\alpha} - 1} \equiv h q_i(f_i, q_i, v_i) \quad (10.39)$$

En écrivant :

$$\dot{q}_i = \dot{\tilde{q}_i} e^{\tilde{q}_i} \quad (10.40)$$

Nous pouvons en déduire :

$$\dot{\tilde{q}_i}(t) = \frac{1}{\tau\rho} e^{\tilde{f}_i(t) - \tilde{q}_i(t)} \left( 1 - (1 - \rho)^{e^{-\tilde{f}_i(t)}} \right) - \frac{1}{\tau} e^{(\frac{1}{\alpha} - 1)\tilde{v}_i(t)} \equiv h\tilde{q}_i(\tilde{f}_i, \tilde{q}_i, \tilde{v}_i) \quad (10.41)$$

En conclusion, pour toute région  $i$  le **modèle hémodynamique transformé** s'écrit :

$$\begin{cases} s\tilde{n}_{B_i} = \Phi(s\tilde{n}_{B_i}) = \frac{1}{1 + \exp(-(s\tilde{n}_{B_i} + \ln(\frac{s_0}{1-s_0})))} \\ s\tilde{n}_{stim_i} = \Phi(s\tilde{n}_{stim_i}) = \frac{1}{1 + \exp(-(s\tilde{n}_{stim_i} + \ln(\frac{s_0}{1-s_0})))} \end{cases} \quad (10.42)$$

$$\begin{cases} \dot{z}_i(t) = \Phi(s\tilde{n}_{B_i})\delta_{u(t),0} + \Phi(s\tilde{n}_{stim_i})\delta_{u(t),1} - \kappa z_i(t) - \gamma(e^{\tilde{f}_i(t)} - 1) = \tilde{h}z_i(s\tilde{n}_{B_i}, s\tilde{n}_{stim_i}, z_i, \tilde{f}_i) \\ \dot{\tilde{f}}_i(t) = z_i(t)e^{\tilde{f}_i(t)} = \tilde{h}f_i(z_i, \tilde{f}_i) \\ \dot{\tilde{v}}_i(t) = \frac{1}{\tau} \left( e^{\tilde{f}_i(t) - \tilde{v}_i(t)} - e^{(\frac{1}{\alpha} - 1)\tilde{v}_i(t)} \right) = \tilde{h}v_i(\tilde{f}_i, \tilde{v}_i) \\ \dot{\tilde{q}}_i(t) = \frac{1}{\tau\rho} e^{\tilde{f}_i(t) - \tilde{q}_i(t)} \left( 1 - (1 - \rho)e^{-\tilde{f}_i(t)} \right) - \frac{1}{\tau} e^{(\frac{1}{\alpha} - 1)\tilde{v}_i(t)} = \tilde{h}q_i(\tilde{f}_i, \tilde{q}_i, \tilde{v}_i) \end{cases} \quad (10.43)$$

$$\text{Avec } \begin{cases} u(t) = 0 \text{ en période de non stimulation} \\ u(t) = 1 \text{ en période de stimulation} \end{cases} \quad (10.44)$$

Pour la suite, nous posons  $\vec{x} = (\vec{z}, \vec{f}, \vec{v}, \vec{q})$  le vecteur des variables d'état et  $\vec{\theta} = (s\tilde{n}_B, s\tilde{n}_{stim})$  le vecteur des paramètres hémodynamiques inconnus que nous souhaitons estimer.

$u(t)$  est l'input extérieur connu à l'instant  $t$ , qui permet d'appliquer des forçages  $\vec{s\tilde{n}_B}$  ou  $\vec{s\tilde{n}_{stim}}$  sur les signaux de vasodilatation  $\vec{z}$  en fonction des périodes de repos ou de stimulation.

La fonction vectorielle  $\vec{H}$  nous permettra d'écrire le modèle hémodynamique à tout instant  $t$  :

$$\boxed{\dot{\vec{x}}(t) = \vec{H}(\vec{x}(t), \vec{\theta}, u(t))} \quad (10.45)$$

### 10.3.2.3.3 Modèle BOLD transformé

Nous avons :

$$y_i = V_0 \left( k_1(1 - q_i) + k_2(1 - q_i v_i^{-1}) + k_3(1 - v_i) \right) = B_i(v_i, q_i) \quad (10.46)$$

qui peut se réécrire en :

$$y_i = V_0 \left( k_1(1 - e^{\ln(q_i)}) + k_2(1 - \frac{e^{\ln(q_i)}}{e^{\ln(v_i)}}) + k_3(1 - e^{\ln(v_i)}) \right) \quad (10.47)$$

En introduisant les changements de variables définis dans l'équation (10.28), nous obtenons la fonction algébrique qui donne le signal BOLD en fonction de  $\tilde{v}_i$  et  $\tilde{q}_i$  :

$$y_i = V_0 \left( k_1(1 - e^{\tilde{q}_i}) + k_2(1 - e^{\tilde{q}_i - \tilde{v}_i}) + k_3(1 - e^{\tilde{v}_i}) \right) \equiv \tilde{B}_i(\tilde{v}_i, \tilde{q}_i) \quad (10.48)$$

La fonction vectorielle  $\vec{B}$  nous permettra d'écrire le modèle BOLD transformé à tout instant  $t$  :

$$\boxed{\vec{y}(t) = \vec{B}(\vec{x}(t))} \quad (10.49)$$

### 10.3.3 Définition des fonctions d'évolution et d'observation

Les mesures BOLD réalisées en IRMf sont séparées d'un intervalle de temps de durée  $TR$ , dans notre cas  $TR = 2.05$  s. Nous notons  $t_{m-1}$  l'instant de la mesure BOLD numéro  $m - 1$ . Nous avons ainsi  $t_m = t_{m-1} + TR$ .

Dans ce paragraphe, nous définissons deux fonctions pour nous permettre d'écrire l'état du système à chaque instant de mesure ([Daunizeau et al., 2009], [Daunizeau et al., 2014]). La fonction d'évolution  $\vec{f}$  va nous permettre d'écrire l'état du système hémodynamique à un instant  $t_m$  (i.e.  $\vec{x}(t_m)$ ) en fonction de l'état à  $t_{m-1}$  (i.e.  $\vec{x}(t_{m-1})$ ). La fonction d'observation  $\vec{g}$  donnera à chaque instant de mesure  $t_m$  les valeurs du signal BOLD  $\vec{y}(t_m)$  en fonction des valeurs des variables hémodynamiques  $\vec{x}(t_m)$ .

#### 10.3.3.1 Fonction d'évolution $\vec{f}$

Entre deux instants de mesure du signal BOLD,  $t_{m-1}$  et  $t_m$ , l'équation différentielle (10.45) peut se réécrire sous la forme intégrale suivante :

$$\vec{x}(t_{m-1} + TR) = \vec{x}(t_{m-1}) + \int_{t_{m-1}}^{t_{m-1} + TR} \vec{H}(\vec{x}(t), \vec{\theta}, u(t)) dt \quad (10.50)$$

Pour pouvoir approximer numériquement cette intégrale, nous utilisons un schéma d'Euler. Pour cela, nous divisons l'intervalle de temps  $[t_{m-1}; t_{m-1} + TR]$  en  $n$  échantillons de durée  $\Delta t$ , de sorte que  $TR = n\Delta t$ .  $\Delta t$  est le pas d'intégration numérique de l'équation différentielle (10.45) entre  $t_{m-1}$  et  $t_m = t_{m-1} + TR$ .

Nous pouvons dans un premier temps calculer  $\vec{x}(t_{m-1} + \Delta t)$  en fonction de  $\vec{x}(t_{m-1})$  :

$$\vec{x}(t_{m-1} + \Delta t) \approx \vec{x}(t_{m-1}) + \Delta t \vec{H}(\vec{x}(t_{m-1}), \vec{\theta}, u(t_{m-1})) \equiv \vec{e}(\vec{x}(t_{m-1}), \vec{\theta}, u(t_{m-1})) \quad (10.51)$$

C'est cette fonction  $\vec{e}$  que nous devons coder dans VBA.

Soit  $j \in [1...n]$ . Pour chaque pas de temps  $\Delta t$  dans l'intervalle  $[t_{m-1}; t_{m-1} + TR]$ , nous pouvons écrire l'état du système à  $t_{m-1} + j\Delta t$  en fonction de l'état au pas de temps précédent  $t_{m-1} + (j - 1)\Delta t$  :

$$\vec{x}(t_{m-1} + j\Delta t) \approx \vec{x}(t_{m-1} + (j - 1)\Delta t) + \Delta t \vec{H}(\vec{x}(t_{m-1} + (j - 1)\Delta t), \vec{\theta}, u(t_{m-1} + (j - 1)\Delta t)) \quad (10.52)$$

$$= \vec{e}(\vec{x}(t_{m-1} + (j - 1)\Delta t), \vec{\theta}, u(t_{m-1} + (j - 1)\Delta t)) \quad (10.53)$$

En calculant tous les  $\overrightarrow{x}(t_{m-1} + j\Delta t)$  de proche en proche, nous arrivons à  $\overrightarrow{x}(t_{m-1} + n\Delta t) = \overrightarrow{x}(t_{m-1} + TR)$ . Nous définissons la fonction  $\overrightarrow{f}$  telle que :

$$\overrightarrow{x}(t_m) = \overrightarrow{f}\left(\overrightarrow{x}(t_{m-1}), \overrightarrow{\theta}\right) \quad (10.54)$$

En posant :

$$\begin{cases} \overrightarrow{x_{m-1}} \equiv \overrightarrow{x}(t_{m-1}) \\ \overrightarrow{x_m} \equiv \overrightarrow{x}(t_m = t_{m-1} + TR) \end{cases} \quad (10.55)$$

Nous pouvons écrire :

$$\boxed{\overrightarrow{x_m} = \overrightarrow{f}(\overrightarrow{x_{m-1}}, \overrightarrow{\theta})} \quad (10.56)$$

La fonction  $\overrightarrow{f}$  est la fonction dite d'évolution dans la toolbox VBA. Elle donne l'état du système hémodynamique à un instant  $t_m = t_{m-1} + TR$  en fonction de l'état à  $t_{m-1}$ . Elle dépend de tous les forçages  $\left(u(t_{m-1} + (j-1)\Delta t)\right)_{j \in [1, n]}$  le long de la trajectoire entre deux  $TR$  consécutifs, et qui sont connus. Nous rendrons implicite cette dépendance, contrairement aux paramètres  $\overrightarrow{\theta}$  que nous voulons estimer<sup>5</sup>.

### 10.3.3.2 Fonction d'observation $\overrightarrow{g}$

La fonction qui donne le signal BOLD en fonction des variables d'état hémodynamique à l'instant  $t_m$  s'écrit :

$$\overrightarrow{y}(t_m) = \overrightarrow{B}(\overrightarrow{x}(t_m)) \quad (10.58)$$

Nous notons  $\overrightarrow{y_m}$  la mesure numéro  $m$  de la série temporelle des observations, acquise à l'instant  $t_m$  :  $\overrightarrow{y_m} \equiv \overrightarrow{y}(t_m)$ . De la même manière, nous écrivons  $\overrightarrow{x_m} \equiv \overrightarrow{x}(t_m)$ . Nous en déduisons :

$$\boxed{\overrightarrow{y_m} = \overrightarrow{B}(\overrightarrow{x_m}) \equiv \overrightarrow{g}(\overrightarrow{x_m})} \quad (10.59)$$

---

5. [Beal, 1998], [Daunizeau et al., 2009] travaillent avec des forçages extérieurs appliqués sur les états cachés uniquement aux instants de mesure :  $u_m \equiv u(t_m) = 0$  ou  $1$ . Ils définissent  $(U_m)_{m \in 1:M}$  une suite de v.a. dont les réalisations  $u_m$  sont connues. Nous nous donnons en plus la possibilité d'avoir des forçages tout le long de la trajectoire. Toutes les densités de probabilités qu'ils expriment sont des densités de probabilités conditionnelles  $|u_{1:M}$ , qui sont rendues implicites. Nous adoptons nous aussi cette convention d'écriture. Ils écrivent par exemple la loi jointe du graphe bayésien pour les paramètres, les états cachés et les observables :

$$p(\overrightarrow{\theta}, \sigma, \overrightarrow{x_{0:M}}, \overrightarrow{y_{1:M}}) \equiv p(\overrightarrow{\theta}, \sigma, \overrightarrow{x_{0:M}}, \overrightarrow{y_{1:M}} | u_{1:M}) \quad (10.57)$$

La fonction  $\overrightarrow{g}$  est la fonction dite d'observation dans la toolbox VBA. C'est celle que nous devons coder. Elle donne à chaque instant de mesure  $t_m$  les valeurs du signal BOLD  $\overrightarrow{y}(t_m)$  en fonction des valeurs des variables hémodynamiques  $\overrightarrow{x}(t_m)$ . En particulier, nous avons aussi  $\overrightarrow{y}_{m-1} = \overrightarrow{g}(\overrightarrow{x}_{m-1})$ .

Nous représentons à la figure [10.2](#) un schéma qui récapitule les dépendances entre variables, entre chaque pas de temps  $\Delta t$  et entre chaque  $TR$ .

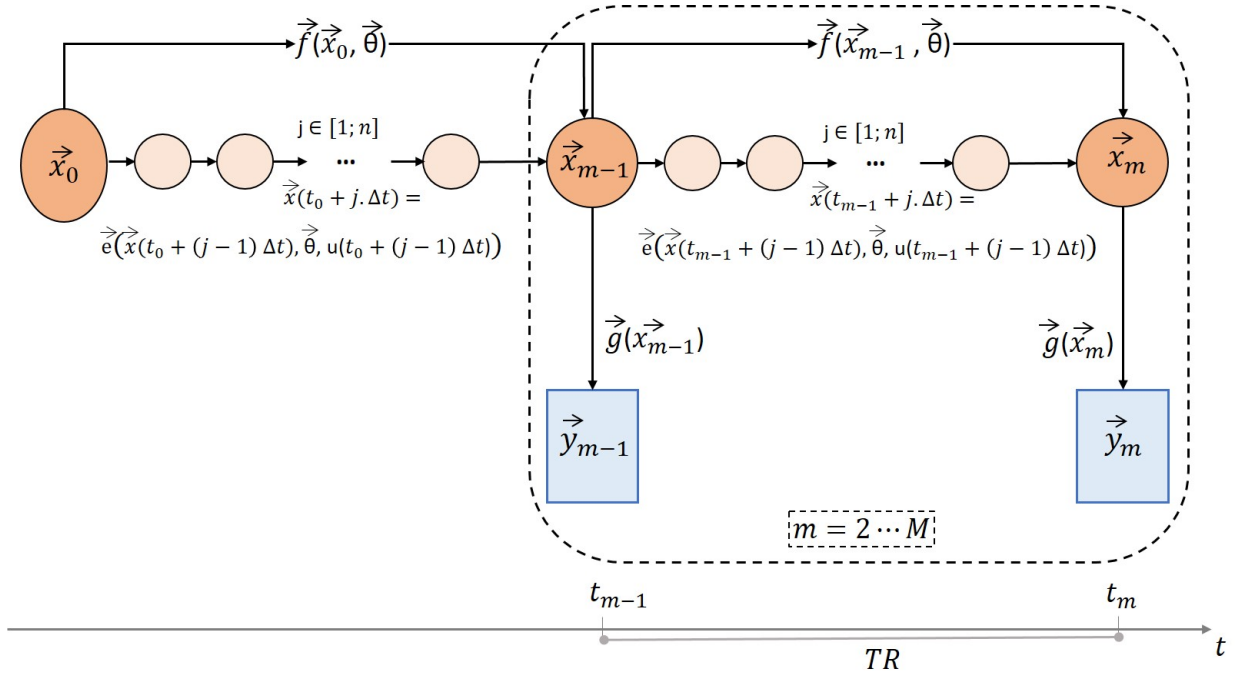


**Figure 10.2** – Schéma des dépendances entre variables d'état hémodynamiques  $(\vec{x}_j)_{j \in [0, M]}$  et BOLD  $(\vec{y}_j)_{j \in [1, M]}$ .

L'intervalle de temps entre deux mesures en IRMf est noté  $TR$ . Nous représentons une séquence temporelle de  $M$  mesures simulées, de durée totale  $M \times TR$ . Chaque  $TR$  est divisé en  $n$  intervalles de durée  $\Delta t$ .

$\vec{\theta}$  donne les paramètres hémodynamiques  $s\tilde{n}_B$  et  $s\tilde{n}_{stim}$  de toutes les régions.  $u(t)$  permet de modifier le forçage appliqué sur le signal de vasodilatation en fonction des périodes de stimulation ou de repos pendant la tâche de contrôle inhibiteur.

Nous partons d'une condition initiale hémodynamique  $\vec{x}_0$ . La fonction  $\vec{e}$  permet de calculer les variables hémodynamiques  $\vec{x}$  tous les pas de temps  $\Delta t$ . La fonction  $\vec{f}$  permet de calculer les variables hémodynamiques à chaque  $TR$  ( $\vec{x}_m$ ) en fonction de leurs valeurs au  $TR$  précédent ( $\vec{x}_{m-1}$ ). La trajectoire hémodynamique dépend des paramètres hémodynamiques  $\vec{\theta}$ .

Tous les  $TR$ , la fonction  $\vec{g}$  permet de calculer le signal BOLD  $\vec{y}_m$ , en fonction des variables hémodynamiques  $\vec{x}_m$ .

Inspiré de [Daunizeau et al., 2009].

### 10.3.4 Définition du modèle statistique direct génératif de la mesure BOLD

Un modèle génératif est un modèle statistique qui décrit les mécanismes par lesquels les données sont produites ([Daunizeau et al., 2014]). Dans cette partie, nous contruisons pas à pas le modèle génératif des mesures BOLD à partir du modèle dynamique transformé que nous avons écrit à la section 10.3.2.

Nous travaillons avec les mesures BOLD de  $N$  régions cérébrales, acquises au cours du temps avec une IRMf, pendant  $M \times TR$  (dans notre cas,  $N = 15$  et  $M = 940$ ). Nous disposons ainsi d'une série temporelle de  $M$  mesures BOLD successives, notée  $(\vec{y}_1, \dots, \vec{y}_M)$ . L'entier naturel  $m$  désignera l'indice d'un numéro de mesure et  $t_m$  l'instant de cette mesure ( $t_m = m \times TR$ ). Chaque vecteur  $\vec{y}_m$  contient les mesures des  $N$  régions cérébrales, numérotées  $i$ .

#### 10.3.4.1 Paramètres hémodynamiques : $\tilde{s}n_{B_i}$ et $\tilde{s}n_{stim_i}$

##### 10.3.4.1.1 Définition des variables aléatoires

Pour chaque région  $i$ , nous cherchons à estimer les paramètres  $\tilde{s}n_{B_i}$  et  $\tilde{s}n_{stim_i}$  du modèle hémodynamique direct, à partir de la série temporelle de mesures  $(\vec{y}_1, \dots, \vec{y}_M)$ . Ces paramètres seront appelés paramètres hémodynamiques (ou paramètres d'évolution dans VBA). Par définition, ils sont fixés et ne varient pas au cours du temps. Nous définissons un vecteur de variables aléatoires de dimension  $n_\theta = 2N$ , non indicé sur le temps (ou plus précisément sur le numéro  $m$  de la mesure), noté  $\vec{\theta}^{v.a.}$ . Une réalisation de cette variable aléatoire sera notée  $\vec{\theta}$ .

De manière plus précise, l'estimation des paramètres hémodynamiques revient à inférer la distribution de probabilité que suit la v.a.  $\vec{\theta}^{v.a.}$ , et qui permet de générer la série temporelle  $(\vec{y}_1, \dots, \vec{y}_M)$  avec la plus grande probabilité.

##### 10.3.4.1.2 Distributions de probabilités marginales ou a priori

La loi de probabilité de  $\vec{\theta}^{v.a.}$  est la distribution a priori sur les paramètres hémodynamiques. Elle aura la forme d'une loi gaussienne multidimensionnelle, paramétrée par sa moyenne a priori  $\vec{\mu}_\theta^0$  et sa matrice de covariance a priori  $\vec{\Sigma}_\theta^0$ . Elle porte sur les paramètres transformés (cf annexe 17.1.2 pour avoir une expression analytique de la densité de probabilité a priori correspondante sur les paramètres non transformés).

$$\forall \vec{\theta} \in \mathbb{R}^{n_\theta} \quad p(\vec{\theta}^{v.a.} = \vec{\theta}) \equiv p(\vec{\theta}) = \mathcal{N}(\vec{\mu}_\theta^0, \vec{\Sigma}_\theta^0) \quad (10.60)$$

#### 10.3.4.2 Processus stochastiques d'états cachés et d'états observables

##### 10.3.4.2.1 Définition des variables aléatoires

La série temporelle des mesures  $(\vec{y}_1, \dots, \vec{y}_M)$  peut être vue comme une réalisation d'un processus stochastique à temps discret (i.e. une suite de variables aléatoires indicées dans un ensemble dénombrable), que nous notons  $(\vec{Y}_m)_{m \in [1:M]}$ . Chaque vecteur  $\vec{Y}_m$  contient  $N$  variables aléatoires définies sur  $\mathbb{R}$  (une v.a. par région).

D'après notre modèle, le signal BOLD est généré à partir des valeurs de plusieurs autres variables d'état hémodynamiques. Celles-ci ne sont pas directement mesurées avec une IRMf : on les appelle les "états cachés" (ou hidden states en anglais).

Soit  $(\vec{X}_m)_{m \in [0:M]}$  une autre suite de variables aléatoires à valeurs dans  $\mathbb{R}$ , de dimension  $4N$ . Une réalisation de ce processus stochastique est noté  $\vec{x}_m$  et regroupe pour chacune des  $N$  régions les 4 valeurs des variables d'état hémodynamiques non mesurées. D'après la section 10.3.2.3, nous travaillons avec les variables transformées pour le flux ( $\vec{f}_m$ ), le volume ( $\vec{v}_m$ ) et la concentration en déoxyhémoglobine ( $\vec{q}_m$ ), mais pas pour le signal de vasodilatation ( $\vec{z}_m$ , sans tilde).  $\vec{Z}_m$ ,  $\vec{F}_m$ ,  $\vec{V}_m$  et  $\vec{Q}_m$  désignent les v.a. à valeurs dans  $\mathbb{R}$  respectivement associées aux grandeurs précédentes (chacune de dimension  $N$ ). La variable aléatoire  $\vec{X}_m$  et une réalisation  $\vec{x}_m$  se notent comme suit :

$$\vec{X}_m = \begin{pmatrix} \vec{Z}_m \\ \vec{F}_m \\ \vec{V}_m \\ \vec{Q}_m \end{pmatrix}, \quad \vec{x}_m = \begin{pmatrix} \vec{z}_m \\ \vec{f}_m \\ \vec{v}_m \\ \vec{q}_m \end{pmatrix} \quad (10.61)$$

#### 10.3.4.2.2 Etats cachés : Définition des probabilités conditionnelles et des probabilités marginales/a priori

##### 10.3.4.2.2.1 Probabilités de transition entre états cachés

Supposons un jeu de paramètres hémodynamiques  $\vec{\theta}$  connu et l'ensemble des états  $(\vec{x}_0, \dots, \vec{x}_{m-1})$  réalisés. Notons  $p(\vec{X}_m = \vec{x}_m | \vec{X}_{m-1} = \vec{x}_{m-1}, \dots, \vec{X}_0 = \vec{x}_0, \theta^{v.a.} = \vec{\theta})$  la densité de probabilité de transition pour passer de  $(\vec{x}_0, \dots, \vec{x}_{m-1})$  à  $\vec{x}_m$ , sachant  $\vec{\theta}$ .

La probabilité de transition pour passer dans un état  $\vec{x}_m$  à  $d\vec{x}_m$  près s'écrit donc :

$$p(\vec{X}_m = \vec{x}_m | \vec{X}_{m-1} = \vec{x}_{m-1}, \dots, \vec{X}_0 = \vec{x}_0, \theta^{v.a.} = \vec{\theta}) d\vec{x}_m \quad (10.62)$$

L'intégration des équations différentielles hémodynamiques entre les instants des mesures  $m-1$  et  $m$  permet de connaître  $\vec{x}_m$  à partir de  $\vec{x}_{m-1}$  et  $\vec{\theta}$ . C'est ce que traduit l'équation (10.56). La réalisation de  $\vec{x}_m$  à  $d\vec{x}_m$  près dépend alors uniquement de la réalisation précédente



$\overrightarrow{x_{m-1}}$  et de  $\overrightarrow{\theta}$ .

$$\begin{aligned} p(\overrightarrow{X_m = x_m} | \overrightarrow{X_{m-1} = x_{m-1}}, \dots, \overrightarrow{X_0 = x_0}, \overrightarrow{\theta^{v.a.}} = \overrightarrow{\theta}) d\overrightarrow{x_m} \\ = p(\overrightarrow{X_m = x_m} | \overrightarrow{X_{m-1} = x_{m-1}}, \overrightarrow{\theta^{v.a.}} = \overrightarrow{\theta}) d\overrightarrow{x_m} \end{aligned} \quad (10.63)$$

Sachant  $\overrightarrow{\theta}$ , la suite  $(\overrightarrow{X_m})_{m \in [0:M]}$  respecte la propriété de Markov d'ordre 1 et nous pouvons écrire la densité de probabilité de transition :

$$\begin{aligned} \forall \overrightarrow{x_m} \in \mathbb{R}^{4N}, p(\overrightarrow{X_m = x_m} | \overrightarrow{X_{m-1} = x_{m-1}}, \dots, \overrightarrow{X_0 = x_0}, \overrightarrow{\theta^{v.a.}} = \overrightarrow{\theta}) \\ = p(\overrightarrow{X_m = x_m} | \overrightarrow{X_{m-1} = x_{m-1}}, \overrightarrow{\theta^{v.a.}} = \overrightarrow{\theta}) \end{aligned} \quad (10.64)$$

$$\equiv p(\overrightarrow{x_m} | \overrightarrow{x_{m-1}}, \overrightarrow{\theta}) \quad (10.65)$$

Sans bruit, d'après l'équation (10.56), cette densité de probabilité de transition est donnée par :

$$p(\overrightarrow{x_m} | \overrightarrow{x_{m-1}}, \overrightarrow{\theta}) = \delta(\overrightarrow{x_m} - f(\overrightarrow{x_{m-1}}, \overrightarrow{\theta})) \quad (10.66)$$

#### 10.3.4.2.2.2 Loi de probabilité de $\overrightarrow{X_0}$

La densité de probabilité que suit  $\overrightarrow{X_0}$  est la distribution initiale de la chaîne de Markov  $(\overrightarrow{X_m})_{m \in [0:M]}$ . Il s'agit plus précisément de la distribution a priori sur la condition initiale des états cachés  $(\overrightarrow{Z_0}, \overrightarrow{F_0}, \overrightarrow{V_0}, \overrightarrow{Q_0})$ .

Elle sera de la forme d'une loi gaussienne multidimensionnelle ( $4N$ ). Nous devons spécifier la moyenne  $\overrightarrow{\mu_{x_0}^0}$  et la matrice de covariance  $\overrightarrow{\Sigma_{x_0}^0}$  a priori.

$$\forall \overrightarrow{x_0} \in \mathbb{R}^{4N}, p(\overrightarrow{X_0 = x_0}) \equiv p(\overrightarrow{x_0}) = \mathcal{N}(\overrightarrow{\mu_{x_0}^0}, \overrightarrow{\Sigma_{x_0}^0}) \quad (10.67)$$

#### 10.3.4.2.2.3 Représentation de la chaîne de Markov associée aux états cachés

Un graphe de causalité peut être associé au modèle probabiliste des états cachés et des paramètres hémodynamiques, représentant les relations entre les différentes variables, i.e. dépendance causale ou indépendance (thèse [Beal, 1998]; [Patrick, 2007]). Les variables aléatoires sont contenues dans des noeuds. Il existe une équivalence entre la représentation graphique et la loi de probabilité sous-jacente. Ainsi une flèche de A vers B représente un lien causal de A vers B et permet de définir la probabilité conditionnelle de B sachant A, notée  $p(B|A)$ <sup>6</sup>. On dit que A est le noeud parent de B<sup>7</sup>, A cause un effet sur B (notion de temporalité). L'absence de flèche entre A et B représente l'indépendance entre A et B.

6. mais ne veut pas dire que  $p(A|B) = 0$

7. Dans un réseau bayésien, un noeud ne dépend conditionnellement que de ses parents.

La figure 10.3 donne le graphe associé au modèle probabiliste des états cachés et des paramètres hémodynamiques. L'ensemble des  $(\vec{X}_m)_{m \in [0:M]}$  forme une chaîne de Markov d'ordre 1 dont les densités de probabilité de transition sont définies sachant  $\vec{\theta}^{v.a} = \vec{\theta}$ .

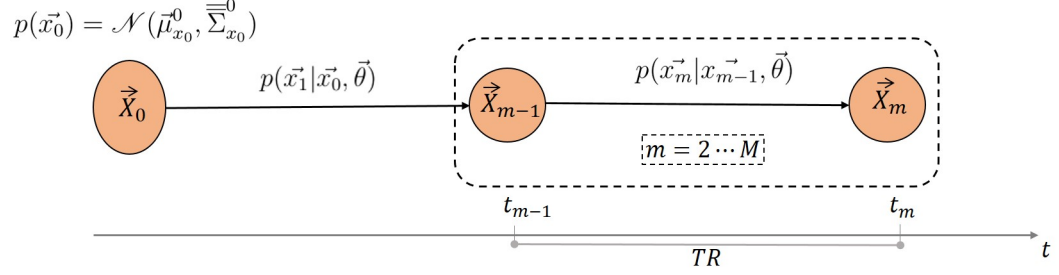


**Figure 10.3** – Chaîne de Markov des états cachés d'ordre 1 (tous les  $TR$ ), notée  $(\vec{X}_m)_{m \in [0:M]}$ . La densité de probabilité initiale est une loi gaussienne. Nous avons représenté les densités de probabilité de transition pour des paramètres hémodynamiques  $\vec{\theta}$  connus, i.e. pour une réalisation de la variable aléatoire des paramètres hémodynamiques  $\vec{\theta}^{v.a}$ .

#### 10.3.4.2.2.4 Densité de probabilité d'émission d'une séquence d'états cachés

La densité de probabilité jointe de toute la chaîne de Markov pour  $\vec{\theta}$  fixé s'écrit :

$$\forall \vec{x}_0, \dots, \vec{x}_M \in \mathbb{R}^{4N}, p(\vec{X}_0 = \vec{x}_0, \dots, \vec{X}_M = \vec{x}_M | \vec{\theta}^{v.a} = \vec{\theta}) \quad (10.68)$$

Nous pouvons l'exprimer en fonction des densités de probabilité initiale et de transition entre états cachés. Pour cela, écrivons la probabilité de réalisation de la série  $(\vec{x}_0, \dots, \vec{x}_M)$  à  $d\vec{x}_0 \dots d\vec{x}_M$  près, pour  $\vec{\theta}$  fixé :

$$p(\vec{X}_0 = \vec{x}_0, \dots, \vec{X}_M = \vec{x}_M | \vec{\theta}^{v.a} = \vec{\theta}) d\vec{x}_0 \dots d\vec{x}_M \quad (10.69)$$

Pour  $\vec{\theta}$  fixé, la suite  $(\vec{X}_m)_{m \in [0:M]}$  possède la propriété de Markov d'ordre 1. Par conséquent, nous pouvons calculer la probabilité de réalisation de la chaîne  $(\vec{x}_0, \dots, \vec{x}_M)$ , sachant  $\vec{\theta}$ , en multipliant :

- la probabilité initiale de la chaîne.  
 $p(\vec{X}_0 = \vec{x}_0) d\vec{x}_0$  est la probabilité de réalisation de  $\vec{x}_0$  à  $d\vec{x}_0$  près.
- les probabilités de transition le long de la chaîne.  
 Quelque soit  $m \in [1; M]$ ,  $p(\vec{X}_m = \vec{x}_m | \vec{X}_{m-1} = \vec{x}_{m-1}, \vec{\theta}^{v.a} = \vec{\theta}) d\vec{x}_m$  est la probabilité d'avoir  $\vec{x}_m$  à  $d\vec{x}_m$  près, sachant que  $\vec{x}_{m-1}$  et  $\vec{\theta}$  ont été réalisés.

Nous obtenons alors :

$$\begin{aligned} p(\vec{X}_0 = x_0, \dots, \vec{X}_M = x_M | \vec{\theta}^{va} = \vec{\theta}) dx_0 \dots dx_M \\ = p(\vec{X}_0 = x_0) dx_0 \prod_{m=1}^M p(\vec{X}_m = x_m | \vec{X}_{m-1} = x_{m-1}, \vec{\theta}^{va} = \vec{\theta}) dx_m \end{aligned} \quad (10.70)$$

La densité de probabilité jointe de toute la chaîne de Markov pour  $\vec{\theta}$  fixé vérifie donc :

$$\begin{aligned} \forall x_0, \dots, x_M \in \mathbb{R}^{4N}, p(\vec{X}_0 = x_0, \dots, \vec{X}_M = x_M | \vec{\theta}^{va} = \vec{\theta}) \\ = p(\vec{X}_0 = x_0) \prod_{m=1}^M p(\vec{X}_m = x_m | \vec{X}_{m-1} = x_{m-1}, \vec{\theta}^{va} = \vec{\theta}) \end{aligned} \quad (10.71)$$

Nous écrirons plus simplement :

$$\forall x_0, \dots, x_M \in \mathbb{R}^{4N}, p(x_0, \dots, x_M | \vec{\theta}) = p(x_0) \prod_{m=1}^M p(x_m | x_{m-1}, \vec{\theta}) \quad (10.72)$$

### 10.3.4.2.3 Etats observables

#### 10.3.4.2.3.1 Probabilités d'observation

Le modèle BOLD se traduit par une fonction algébrique (cf section 10.3.2.3.3) qui permet d'obtenir une valeur du signal BOLD à partir des états cachés (ie variables hémodynamiques transformées). Chaque réalisation  $y_m$  dépend de la réalisation  $x_m$  au même instant de mesure, comme traduit par l'équation (10.59).

Sans bruit, nous pouvons écrire la densité de probabilité d'observation de  $y_m$  à partir de  $x_m$  :

$$p(\vec{Y}_m = y_m | \vec{X}_m = x_m) = \delta(y_m - g(x_m)) \quad (10.73)$$

Nous considérerons l'effet de bruit sur cette densité de probabilité dans la section 10.3.4.3.

#### 10.3.4.2.3.2 Représentation du Modèle de Markov Caché

Les suites  $(\vec{X}_m)_{m \in [0:M]}$  et  $(\vec{Y}_m)_{m \in [1:M]}$  définissent un modèle statistique particulier : c'est un réseau bayésien<sup>8</sup> dynamique<sup>9</sup>, de type modèle de Markov caché, ou Hidden Markov Model (HMM) en anglais, de gauche à droite. Nous le représentons à la figure 10.4. Chaque noeud ne dépend conditionnellement que de ses parents.

8. graphe orienté acyclique (i.e. sans boucle ou auto-connexions)

9. Dynamic Bayesian Network (DBN). Les suites de variables aléatoires permettent de générer la série de mesures au cours du temps

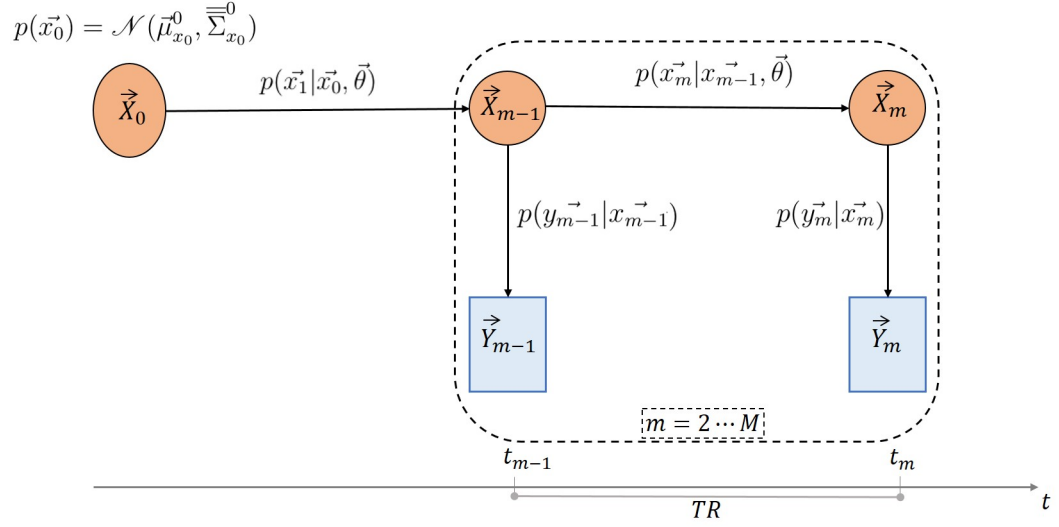


**Figure 10.4** – Modèle de Markov Caché qui permet de générer une série de mesures (tous les  $TR$ ) à partir d'états cachés. Les densités de probabilité initiale, de transition entre états cachés et d'observation définissent le Modèle de Markov Caché

#### 10.3.4.2.3.3 Densité de probabilité d'émission d'une séquence de mesures

Les densités de probabilité initiale, de transition entre états cachés et d'observation définissent le Modèle de Markov Caché. Elles vont nous permettre de calculer d'abord la densité de probabilité jointe de celui-ci sachant  $\vec{\theta}$  ; puis la densité de probabilité d'émission d'une séquence de mesures sachant  $\vec{\theta}$ .

Nous introduisons les notations  $\vec{x_0}, \dots, \vec{x_M} \equiv \vec{x_{0:M}}, y_1, \dots, y_M \equiv y_{1:M}$ ,  
 $\vec{dx_0} \dots \vec{dx_M} \equiv \vec{dx_{0:M}}$  et  $\vec{dy_1} \dots \vec{dy_M} \equiv \vec{dy_{1:M}}$ .

La densité de probabilité jointe du Modèle de Markov Caché sachant  $\vec{\theta}$  s'écrit :

$$\vec{\forall x_{0:M} \in \mathbb{R}^{4N}, \forall y_{1:M} \in \mathbb{R}^N, p(\vec{x_{0:M}}, \vec{y_{1:M}} | \vec{\theta}) \quad (10.74)$$

Pour  $\vec{\theta}$  fixé, la probabilité que le Modèle de Markov Caché ait pu générer la série d'états cachés  $(\vec{x_{0:M}})$  à  $\vec{dx_{0:M}}$  près, et la série de mesures  $(\vec{y_{1:M}})$  à  $\vec{dy_{1:M}}$  s'écrit alors :

$$\vec{p(\vec{x_{0:M}}, \vec{y_{1:M}} | \vec{\theta})} \vec{dx_{0:M}} \vec{dy_{1:M}} \quad (10.75)$$

Elle se calcule en multipliant :

- la probabilité initiale de la chaîne de Markov des états cachés.  
 $\vec{p(x_0)dx_0}$  est la probabilité de réalisation de  $x_0$  à  $dx_0$  près.
- les probabilités de transition le long de la chaîne de Markov des états cachés.  
 Quelque soit  $m \in [1; M]$ ,  $\vec{p(x_m|x_{m-1}, \vec{\theta})dx_m}$  est la probabilité d'avoir  $x_m$  à  $dx_m$  près, sachant que  $x_{m-1}$  et  $\vec{\theta}$  ont été réalisés.

— les probabilités d'observation le long de la chaîne.

Quelque soit  $m \in [1; M]$ ,  $p(\overrightarrow{y_m} | \overrightarrow{x_m}) d\overrightarrow{y_m}$  est la probabilité d'avoir  $\overrightarrow{y_m}$  à  $d\overrightarrow{y_m}$  près, sachant que  $\overrightarrow{x_m}$  a été réalisé.

Nous obtenons ainsi :

$$p(\overrightarrow{x_{0:M}}, \overrightarrow{y_{1:M}} | \overrightarrow{\theta}) d\overrightarrow{x_{0:M}} d\overrightarrow{y_{1:M}} = p(\overrightarrow{x_0}) d\overrightarrow{x_0} \prod_{m=1}^M p(\overrightarrow{x_m} | \overrightarrow{x_{m-1}}, \overrightarrow{\theta}) d\overrightarrow{x_m} \times p(\overrightarrow{y_m} | \overrightarrow{x_m}) d\overrightarrow{y_m} \quad (10.76)$$

La densité de probabilité jointe du Modèle de Markov Caché sachant  $\overrightarrow{\theta}$  est alors donnée par :

$$\forall \overrightarrow{x_{0:M}} \in \mathbb{R}^{4N}, \forall \overrightarrow{y_{1:M}} \in \mathbb{R}^N, p(\overrightarrow{x_{0:M}}, \overrightarrow{y_{1:M}} | \overrightarrow{\theta}) = p(\overrightarrow{x_0}) \prod_{m=1}^M p(\overrightarrow{x_m} | \overrightarrow{x_{m-1}}, \overrightarrow{\theta}) p(\overrightarrow{y_m} | \overrightarrow{x_m}) \quad (10.77)$$

Nous pouvons ensuite calculer la probabilité qu'une séquence particulière de mesures ( $\overrightarrow{y_{1:M}}$ ) ait été émise à  $d\overrightarrow{y_{1:M}}$  près, pour  $\overrightarrow{\theta}$  fixé. Pour cela, nous sommes les probabilités jointes précédentes sur tous les états cachés possibles :

$$\forall \overrightarrow{y_{1:M}} \in \mathbb{R}^N, p(\overrightarrow{y_{1:M}} | \overrightarrow{\theta}) d\overrightarrow{y_{1:M}} = \underbrace{\int_{\overrightarrow{x_{0:M}}} p(\overrightarrow{x_{0:M}}, \overrightarrow{y_{1:M}} | \overrightarrow{\theta}) d\overrightarrow{x_{0:M}}}_{\int_{x_0} \dots \int_{x_M}} \quad (10.78)$$

Nous en déduisons la densité de probabilité d'observation de la séquence ( $\overrightarrow{y_{1:M}}$ ) :

$$p(\overrightarrow{y_{1:M}} | \overrightarrow{\theta}) = \int_{\overrightarrow{x_{0:M}}} p(\overrightarrow{x_{0:M}}, \overrightarrow{y_{1:M}} | \overrightarrow{\theta}) d\overrightarrow{x_{0:M}} \quad (10.79)$$

### 10.3.4.3 Hyperparamètres de bruit sur l'observable

#### 10.3.4.3.1 Variables aléatoires et distributions de probabilité

Nous ajoutons un bruit Gaussien sur l'observable BOLD<sup>10</sup>, de sorte que toute mesure  $\overrightarrow{y_m}$  peut s'écrire :

$$\overrightarrow{y_m} = \overrightarrow{g}(\overrightarrow{x_m}) + \overrightarrow{\epsilon_m} \quad (10.80)$$

$\overrightarrow{\epsilon_m}$  est la réalisation d'une variable aléatoire notée  $\overrightarrow{\epsilon_m^{v.a.}}$ .  
 $(\overrightarrow{\epsilon_m^{v.a.}})_{m \in [1:M]}$  est une suite de vecteurs de variables aléatoires indépendants et identiquement distribués, qui suivent une loi normale de moyenne nulle et de covariance  $\overline{\overline{\Sigma_\epsilon}}$  :

$$\overrightarrow{\epsilon_m^{v.a.}} \sim \mathcal{N}(\overrightarrow{0}, \overline{\overline{\Sigma_\epsilon}}) \quad (10.81)$$

---

10. Principales sources de bruit : bruit lié à la machine et bruit physiologique (cf chapitre 4)

Entre les régions cérébrales, nous supposons le bruit indépendant de sorte que les termes extradiagonaux de la matrice de covariance  $\overline{\overline{\Sigma}}_\epsilon$  sont nuls. Nous noterons  $\sigma$  l'inverse de la variance du bruit pour toutes les régions<sup>11</sup>. La matrice de covariance  $\overline{\overline{\Sigma}}_\epsilon$  s'écrit dans ce cas :

$$\overline{\overline{\Sigma}}_\epsilon = \sigma^{-1} \overline{\overline{\mathbb{I}}}_N \quad (10.82)$$

où  $\overline{\overline{\mathbb{I}}}_N$  est la matrice identité, de dimension  $N \times N$  (i.e. le nombre de régions cérébrales).

$\sigma$  est la réalisation d'une v.a., que nous noterons  $\sigma^{v.a.}$ , à valeurs dans  $\mathbb{R}^+$ .  $\sigma^{v.a.}$  suit une distribution gamma ( $\Gamma$ ) de paramètres  $a_\sigma$  (paramètre de forme ou shape parameter) et  $b_\sigma$  (rate parameter). L'expression de la loi  $\Gamma$ , son espérance et sa variance se trouvent en annexe 17.1.1.

$$\sigma^{v.a.} \sim \Gamma(a_\sigma, b_\sigma) \quad (10.83)$$

On va estimer les paramètres  $a_\sigma$  et  $b_\sigma$  de cette seconde distribution que suit l'inverse de la variance du bruit. Autrement dit, on estime le moment d'une première distribution (la covariance de la loi de probabilité que suit la v.a.  $\overrightarrow{\epsilon}_m^{v.a.}$ ) par une deuxième distribution. Les paramètres de cette seconde distribution (i.e.  $a_\sigma$  et  $b_\sigma$ ) sont appelés les hyperparamètres du bruit.

#### 10.3.4.3.2 Que devient la probabilité d'observation ?

Sans bruit, la densité de probabilité d'observation s'écrivait avec une distribution de Dirac (cf équation (10.73)) Avec ce bruit gaussien, la densité de probabilité d'observation s'écrit maintenant avec une loi normale :

$$\forall \overrightarrow{y}_m \in \mathbb{R}^N, p(\overrightarrow{Y}_m = \overrightarrow{y}_m | \overrightarrow{X}_m = \overrightarrow{x}_m, \sigma^{v.a.} = \sigma) = \mathcal{N}(\overrightarrow{g}(\overrightarrow{x}_m), \sigma^{-1} \overline{\overline{\mathbb{I}}}_N) \quad (10.84)$$

#### 10.3.4.3.3 Loi marginale a priori sur l'inverse de la variance du bruit

Nous noterons  $a_\sigma^0$  et  $b_\sigma^0$  les valeurs a priori des hyperparamètres de bruit sur la mesure BOLD.  $p(\sigma)$  est la distribution a priori que suit l'inverse de la variance du bruit de toutes les régions ([Daunizeau et al., 2014]).

$$\forall \sigma \in \mathbb{R}^+, p(\sigma^{v.a.} = \sigma) = \Gamma(a_\sigma^0, b_\sigma^0) \quad (10.85)$$

---

11. Nous adoptons la notation de [Daunizeau et al., 2009]. Dans ce texte,  $\sigma$  ne désigne pas la racine carrée de la variance du bruit comme habituellement, mais l'inverse de la variance du bruit.

#### 10.3.4.3.4 Représentation du Modèle de Markov Caché avec bruit sur l'observable

Le modèle statistique final est un modèle génératif (c'est-à-dire un modèle statistique/probabiliste qui décrit des mécanismes par lesquels les données sont produites) hiérarchique, dynamique et non linéaire. Nous le représentons à la figure 10.5. La toolbox VBA permet de travailler avec cette catégorie de modèles génératifs ([Daunizeau et al., 2009], [Daunizeau et al., 2014]).

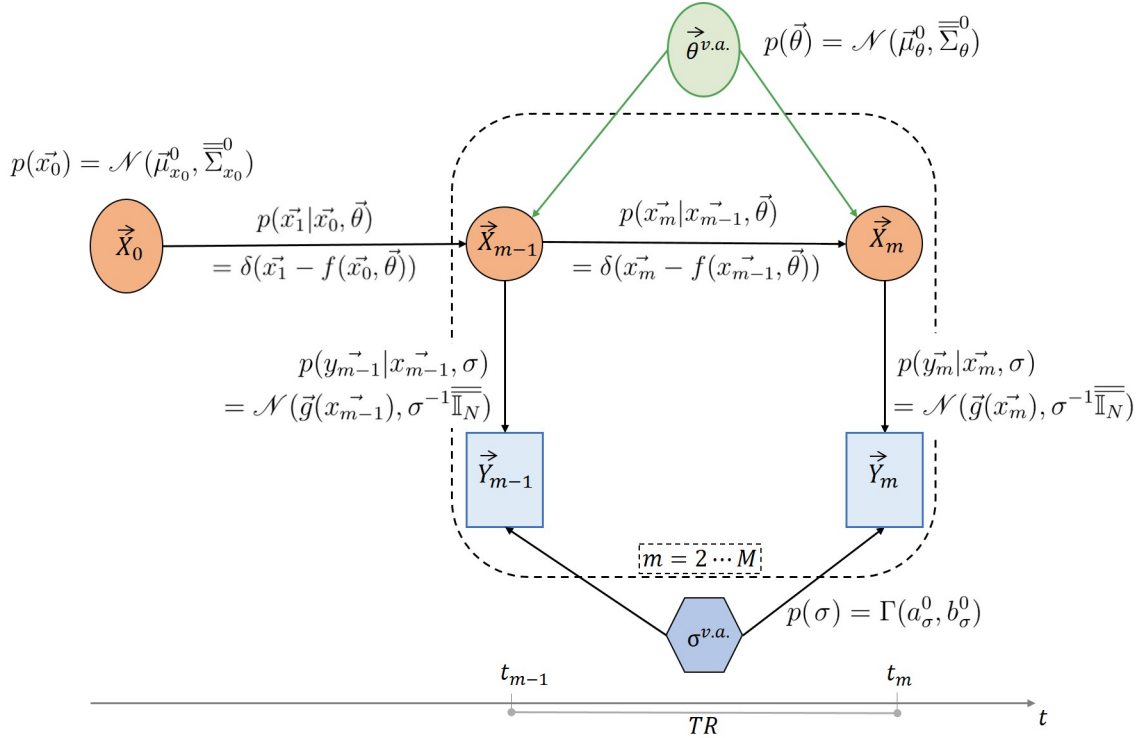


**Figure 10.5** – Graphe représentant le modèle statistique génératif des mesures BOLD.

Il est défini par une suite d'états cachés  $(\vec{X}_m)_{m \in [0:M]}$ , d'observables  $(\vec{Y}_m)_{m \in [1:M]}$ , de paramètres hémodynamiques  $\vec{\theta}^{v.a.}$  et un paramètre de bruit sur les observables  $\sigma^{v.a.}$  (inverse de la variance du bruit gaussien). Nous avons représenté les densités de probabilité a priori sur la condition initiale des états cachés  $p(\vec{x}_0)$ , sur les paramètres hémodynamiques  $p(\vec{\theta})$  et sur l'inverse de la variance du bruit  $p(\sigma)$ . Les deux premières sont des lois normales et la dernière est une loi gamma. Nous avons également représenté les densités de probabilité de transition entre états cachés (qui s'écrivent avec la fonction d'évolution  $\vec{f}$ ) et les densités de probabilité d'observation entre états cachés et observables (qui dépendent de la fonction d'observation  $\vec{g}$  et du paramètre de bruit  $\sigma$ ). Ce schéma est inspiré de [Daunizeau et al., 2009].

### 10.3.5 Méthode d'inversion

Avec cette structure du réseau bayésien, nous recherchons le meilleur jeu de paramètres qui permette de rendre compte des données observées. La procédure d'inversion va permettre de mettre à jour les différentes probabilités conditionnelles du graphe en tenant compte des informations exogènes, c'est-à-dire de la série temporelle des données  $\overrightarrow{y_{1:M}}$  ([Patrick, 2007]). Nous présentons dans cette section le principe de la méthode d'inversion bayésienne variationnelle.

#### 10.3.5.1 Méthode bayésienne variationnelle : un problème d'optimisation fonctionnelle

Les quantités inconnues de ce problème d'inversion sont les paramètres hémodynamiques  $\overrightarrow{\theta}$ , les états cachés en condition initiale  $x_0$  et aux instants de mesures  $x_{1:M}$ , ainsi que l'inverse de la variance du bruit de toutes les régions  $\sigma$ .

Nous cherchons à approximer deux densités de probabilités ([Beal, 1998], [Daunizeau et al., 2009]) :

- la loi jointe a posteriori de toutes ces quantités inconnues sachant le jeu de données  $\overrightarrow{y_{1:M}} : p(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma | \overrightarrow{y_{1:M}})$
- la log évidence du modèle (i.e. le logarithme de l'évidence) :  $\ln(\overrightarrow{p(y_{1:M})})$

Pour cela, nous supposons une distribution arbitraire, notée  $q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma)$ , qui va nous permettre de chercher une loi approchante de la loi jointe a posteriori sachant le jeu de données,  $p(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma | \overrightarrow{y_{1:M}})$ .

Nous pouvons d'abord écrire la log évidence en fonction de la loi jointe à posteriori sachant  $\overrightarrow{y_{1:M}}$  :

$$\ln(\overrightarrow{p(y_{1:M})}) = \ln\left(\frac{\overrightarrow{p(x_0, x_{1:M}, \theta, \sigma | y_{1:M})}}{\overrightarrow{p(x_0, x_{1:M}, \theta, \sigma | y_{1:M})}} \overrightarrow{p(y_{1:M})}\right) \quad (10.86)$$

$$= \ln\left(\frac{\overrightarrow{p(x_0, x_{1:M}, \theta, \sigma, y_{1:M})}}{\overrightarrow{p(x_0, x_{1:M}, \theta, \sigma | y_{1:M})}}\right) \quad (10.87)$$

$$(10.88)$$



En introduisant la distribution arbitraire  $q$ , nous avons :

$$\ln(\overrightarrow{p(y_{1:M})}) = \ln(\overrightarrow{p(y_{1:M})}) \underbrace{\int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma)}_{=1} \quad (10.89)$$

$$= \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \ln(\overrightarrow{p(y_{1:M})}) \quad (10.90)$$

$$= \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \ln\left(\frac{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma, y_{1:M})}}{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma | y_{1:M})}}\right) \quad (10.91)$$

$$= \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \quad (10.92)$$

$$\times \ln\left(\frac{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma, y_{1:M})}}{\overrightarrow{q(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma)}} \frac{\overrightarrow{q(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma)}}{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma | y_{1:M})}}\right) \quad (10.93)$$

Nous écrivons le logarithme du produit comme une somme :

$$\ln(\overrightarrow{p(y_{1:M})}) = \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \ln\left(\frac{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma, y_{1:M})}}{\overrightarrow{q(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma)}}\right) \quad (10.94)$$

$$+ \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \ln\left(\frac{\overrightarrow{q(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma)}}{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma | y_{1:M})}}\right) \quad (10.95)$$

Le premier terme de la somme définit une fonctionnelle que nous noterons  $F(q)$  et qui est nommée énergie libre.

$$F(q) \equiv \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \ln\left(\frac{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma, y_{1:M})}}{\overrightarrow{q(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma)}}\right) \quad (10.96)$$

Le second terme de la somme (10.95) est la divergence de Kulback-Leibler entre la distribution arbitraire  $q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma)$  et la distribution conditionnelle a posteriori  $\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma | y_{1:M})}$ .

$$\begin{aligned} & KL[q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) || p(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma | y_{1:M})] \\ & \equiv \int \overrightarrow{dx_0} \overrightarrow{dx_{1:M}} d\overrightarrow{\theta} d\sigma q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) \ln\left(\frac{\overrightarrow{q(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma)}}{\overrightarrow{p(x_0, x_{1:M}, \overrightarrow{\theta}, \sigma | y_{1:M})}}\right) \end{aligned} \quad (10.97)$$

L'équation (10.95) s'écrit donc

$$\ln(\overrightarrow{p(y_{1:M})}) = F(q) + KL[q(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma) || p(\overrightarrow{x_0}, \overrightarrow{x_{1:M}}, \overrightarrow{\theta}, \sigma | y_{1:M})] \quad (10.98)$$

La divergence de Kulback-Leibler  $KL$  est l'erreur d'approximation entre la log évidence  $\ln p(\overrightarrow{y_{1:M}})$  et l'énergie libre  $F(q)$ . Puisque la quantité  $KL$  est strictement positive ([Beal, 1998]),  $F(q)$  est un minorant de  $\ln p(\overrightarrow{y_{1:M}})$ .

En maximisant la fonctionnelle  $F(q)$  par rapport à la distribution  $q$  (i.e.  $\frac{\delta F}{\delta q} = 0$ ), nous réduisons la divergence  $KL$ . Supposons une distribution  $q^{opt}$  qui maximise  $F(q)$  et minimise  $KL$ .

Si  $KL \rightarrow 0$ , d'après l'expression (10.97) nous en déduisons d'abord :

$$q^{opt}(\overrightarrow{x_0, x_{1:M}}, \overrightarrow{\theta}, \sigma) \approx p(\overrightarrow{x_0, x_{1:M}}, \overrightarrow{\theta}, \sigma | \overrightarrow{y_{1:M}}) \quad (10.99)$$

$q^{opt}$  est une approximation de la distribution jointe a posteriori  $p(\overrightarrow{x_0, x_{1:T}}, \overrightarrow{\theta}, \sigma | \overrightarrow{y_{1:M}})$  (sur toutes les quantités inconnues du modèle, sachant un jeu de données  $\overrightarrow{y_{1:M}}$ ).

Ensuite, d'après la formule (10.98), nous obtenons :

$$F(q^{opt}) \approx \ln p(\overrightarrow{y_{1:M}}) \quad (10.100)$$

$F(q^{opt})$  est le plus grand des minorants (i.e. une borne inférieure) de la log évidence  $\ln p(\overrightarrow{y_{1:M}})$  et peut donc servir de loi approchante.

L'objectif principal de la procédure d'inversion est ainsi de chercher une distribution  $q^{opt}$  qui maximise la fonctionnelle  $F(q)$  (i.e.  $\frac{\delta F}{\delta q} = 0$ ). Ceci permettra d'approximer les deux densités de probabilités énoncées au début de cette section :

- $q^{opt}$  servira de densité de probabilité approchante de la loi jointe a posteriori  $p(\overrightarrow{x_0, x_{1:T}}, \overrightarrow{\theta}, \sigma | \overrightarrow{y_{1:M}})$ .
- $F(q^{opt})$  est une approximation de la log évidence  $\ln p(\overrightarrow{y_{1:M}})$ .

Pour rechercher  $q^{opt}$ , deux principales approximations seront nécessaires : l'approximation en champ moyen<sup>12</sup> et l'approximation de Laplace ([Daunizeau et al., 2009]).

### 10.3.5.2 Hypothèses de travail pour chercher une loi approchante optimale $q^{opt}$

#### 10.3.5.2.1 Approximation en champ moyen

L'approximation en champ moyen postule l'indépendance de  $\overrightarrow{\theta}$ ,  $\sigma$ ,  $\overrightarrow{x_0}$  et  $\overrightarrow{x_{1:T}}$  conditionnellement par rapport aux données  $\overrightarrow{y_{1:M}}$ . Elle a pour intérêt de pouvoir séparer les distributions sur les paramètres et sur les états cachés.

La loi jointe a posteriori du réseau bayésien peut ainsi s'écrire :

$$p(\overrightarrow{x_0, x_{1:T}}, \overrightarrow{\theta}, \sigma | \overrightarrow{y_{1:M}}) = p(\overrightarrow{x_0} | \overrightarrow{y_{1:M}}) p(\overrightarrow{x_{1:T}} | \overrightarrow{y_{1:M}}) p(\overrightarrow{\theta} | \overrightarrow{y_{1:M}}) p(\sigma | \overrightarrow{y_{1:M}}) \quad (10.101)$$

---

12. corpus de la physique statistique

Nous pouvons également séparer les distributions  $q_{x_0}, q_{x_{1:T}}, q_\theta$  et  $q_\sigma$  dans la distribution arbitraire  $q$  :

$$q(\overrightarrow{x_0}, \overrightarrow{x_{1:T}}, \overrightarrow{\theta}, \sigma) = q_{x_0}(\overrightarrow{x_0}) q_{x_{1:T}}(\overrightarrow{x_{1:T}}) q_\theta(\overrightarrow{\theta}) q_\sigma(\sigma) \quad (10.102)$$

L'énergie libre peut se réécrire comme une somme en séparant les contributions des différentes densités de probabilité  $q_\square$  ([Beal, 1998], [Daunizeau et al., 2009]), de sorte que :

$$F(q(\overrightarrow{x_0}, \overrightarrow{x_{1:T}}, \overrightarrow{\theta}, \sigma)) = F(q_{x_0}(\overrightarrow{x_0}), q_{x_{1:T}}(\overrightarrow{x_{1:T}}), q_\theta(\overrightarrow{\theta}), q_\sigma(\sigma)) .$$

En particulier, la loi approchante optimale  $q^{opt}$  s'écrit :

$$q^{opt}(\overrightarrow{x_0}, \overrightarrow{x_{1:T}}, \overrightarrow{\theta}, \sigma) = q_{x_0}^{opt}(\overrightarrow{x_0}) q_{x_{1:T}}^{opt}(\overrightarrow{x_{1:T}}) q_\theta^{opt}(\overrightarrow{\theta}) q_\sigma^{opt}(\sigma) \quad (10.103)$$

- $q_{x_0}^{opt}(\overrightarrow{x_0})$  est la distribution approchante de la densité de probabilité a posteriori sur la condition initiale des états cachés  $\overrightarrow{x_0}$ , sachant le jeu de données  $\overrightarrow{y_{1:M}}$  ([Daunizeau et al., 2009]) :

$$q_{x_0}^{opt}(\overrightarrow{x_0}) \approx p(\overrightarrow{x_0} | \overrightarrow{y_{1:M}}) \quad (10.104)$$

- $q_{x_{1:T}}^{opt}(\overrightarrow{x_{1:T}})$  est la distribution approchante de la densité de probabilité a posteriori sur les états cachés aux instants de mesures  $\overrightarrow{x_{1:T}}$ , sachant le jeu de données  $\overrightarrow{y_{1:M}}$  :

$$q_{x_{1:T}}^{opt}(\overrightarrow{x_{1:T}}) \approx p(\overrightarrow{x_{1:T}} | \overrightarrow{y_{1:M}}) \quad (10.105)$$

- $q_\theta^{opt}(\overrightarrow{\theta})$  est la distribution approchante de la densité de probabilité a posteriori des paramètres hémodynamiques  $\overrightarrow{\theta}$  que l'on souhaite inverser, sachant le jeu de données  $\overrightarrow{y_{1:M}}$  :

$$q_\theta^{opt}(\overrightarrow{\theta}) \approx p(\overrightarrow{\theta} | \overrightarrow{y_{1:M}}) \quad (10.106)$$

- $q_\sigma^{opt}(\sigma)$  est la distribution approchante de la densité de probabilité a posteriori de l'inverse de la variance du bruit  $\sigma$ , sachant le jeu de données  $\overrightarrow{y_{1:M}}$  :

$$q_\sigma^{opt}(\sigma) \approx p(\sigma | \overrightarrow{y_{1:M}}) \quad (10.107)$$

L'énergie libre correspondante s'écrit :

$$F^{opt} \equiv F(q^{opt}(\overrightarrow{x_0}, \overrightarrow{x_{1:T}}, \overrightarrow{\theta}, \sigma)) = F(q_{x_0}^{opt}(\overrightarrow{x_0}), q_{x_{1:T}}^{opt}(\overrightarrow{x_{1:T}}), q_\theta^{opt}(\overrightarrow{\theta}), q_\sigma^{opt}(\sigma))$$

La loi approchante optimale  $q^{opt}$  vérifie que  $\frac{\delta F}{\delta q_\square^{opt}} = 0$ . Cette condition donne une solution analytique mais non explicite du problème d'optimisation fonctionnelle. En effet, les statistiques suffisantes des distributions  $q_\square$  dépendent les une des autres ([Daunizeau et al., 2009]). Il est nécessaire de limiter le nombre de statistiques suffisantes des distributions pour pouvoir résoudre ce problème et c'est ce que permet l'approximation de Laplace.

### 10.3.5.2.2 Approximation de Laplace

L'approximation de Laplace permet de chercher les distributions  $q_{x_0}$ ,  $q_{x_{1:T}}$  et  $q_\theta$  de forme gaussienne. Ces distributions sont définies sur  $\mathbb{R}$  et pour rappel, c'est pour cela que nous avons réécrit à la section 10.3.2 le modèle direct avec des variables et des paramètres transformés définis sur  $\mathbb{R}$ .

Ces distributions sont paramétrées par leur moyenne et leur matrice de covariance respectives (i.e. les statistiques suffisantes de chaque distribution).

$$q_{x_0} = \mathcal{N}(\mu_{x_0}, \overrightarrow{\overline{\Sigma_{x_0}}}) \quad (10.108)$$

$$q_{x_{1:T}} = \mathcal{N}(\mu_{x_{1:T}}, \overrightarrow{\overline{\Sigma_{x_{1:T}}}}) \quad (10.109)$$

$$q_\theta = \mathcal{N}(\mu_\theta, \overrightarrow{\overline{\Sigma_\theta}}) \quad (10.110)$$

Indépendamment de cette approximation, la distribution de l'inverse de la variance du bruit  $\sigma$  (ie  $q_\sigma$ ) est une loi Gamma de paramètres  $a_\sigma$  et  $b_\sigma$ .

$$q_\sigma = \Gamma(a_\sigma, b_\sigma) \quad (10.111)$$

Finalement, nous cherchons un maximum de la fonctionnelle  $F(q)$  sous contrainte, i.e. dans un espace de densités de probabilités  $q_\square$  séparables et de forme choisie.

Les distributions  $q_\square$ <sup>13</sup> dépendent les unes des autres. Pour estimer la loi approchante optimale  $q^{opt}$ , il est maintenant possible de mettre à jour à tour de rôle les statistiques suffisantes des  $q_\square$  qui sont en nombre réduit. Pour cela, on passe par un algorithme itératif de montée de gradient  $\frac{\delta F}{\delta q}$  ([Daunizeau et al., 2009]).

### 10.3.5.3 Principe de l'algorithme d'estimation des paramètres

L'algorithme bayésien variationnel de la toolbox VBA permet de maximiser de manière itérative l'énergie libre  $F(q_{x_0}(x_0), q_{x_{1:T}}(x_{1:T}), q_\theta(\vec{\theta}), q_\sigma(\sigma))$ . Lors de chaque itération de l'algorithme, les statistiques suffisantes des distributions libres  $q_\square$  sont mises à jour à tour de rôle. Elles dépendent les unes des autres. Nous n'indiquerons dans ce texte que les dépendances pour mieux comprendre le principe de l'algorithme, sans entrer dans le détail des calculs qui se trouvent dans [Daunizeau et al., 2009].

- Mise à jour des hyperparamètres de bruit  $a_\sigma$  et  $b_\sigma$  pour optimiser  $F$  par rapport à  $q_\sigma(\sigma)$  (i.e.  $\frac{\delta F}{\delta q_\sigma} = 0$ ), les autres  $q_\square$  étant fixées.  
 $a_\sigma$  et  $b_\sigma$  sont mis à jour en utilisant :
  - les moments des distributions sur les états cachés  $q_{x_{1:T}}$
  - les paramètres  $(a_\sigma^0, b_\sigma^0)$  de la distribution a priori sur  $\sigma$
- Mise à jour des moments des distributions gaussiennes sur les paramètres  $q_\theta$ . L'objectif est de maximiser  $F$  par rapport à  $q_\theta$  ( $\frac{\delta F}{\delta q_\theta} = 0$ ), les autres  $q_\square$  étant fixées. On parle de

---

13. i.e. distributions sur les paramètres hémodynamiques ( $q_\theta$ ), les états cachés ( $q_{x_0}$  et  $q_{x_{1:M}}$ ) et l'inverse de la variance du bruit ( $q_\sigma$ )

l'étape de Maximisation Bayésienne Variationnelle (ou VBM de l'anglais Variational Bayesian Maximization). Les étapes VBE et VBM forment l'algorithme VBEM.

Un algorithme de Gauss-Newton permet d'actualiser la moyenne  $\vec{\mu}_\theta$  en l'incrémentant de  $\vec{\Delta\mu}_\theta$ , ainsi que de calculer la nouvelle matrice de covariance  $\vec{\Sigma}_\theta$ .

$\vec{\Delta\mu}_\theta$  et  $\vec{\Sigma}_\theta$  dépendent des :

- moments des distributions sur les états cachés  $q_{x_{1:T}}$  et les paramètres  $q_\theta$
- moments de la distribution a priori  $p(\vec{\theta})$
- Mise à jour des moments des distributions gaussiennes sur les états cachés en condition initiale  $q_{x_0}$  et aux instants de mesures  $q_{x_{1:M}}$ . L'objectif est de maximiser  $F$  par rapport à  $q_{x_0}$  (i.e.  $\frac{\delta F}{\delta q_{x_0}}$ ) les autres  $q_\square$  étant fixées et  $q_{x_{1:M}}$  (i.e.  $\frac{\delta F}{\delta q_{x_{1:M}}} = 0$ ) les autres  $q_\square$  étant fixées. On parle de l'étape d'Espérance Bayésienne Variationnelle (ou VBE de l'anglais Variational Bayesian Expectation).

$q_{x_0}$  et  $q_{x_{1:M}}$  sont traitées de manière différente. La mise à jour des moments de  $q_{x_{1:M}}$  est l'étape qui demande le plus de ressources numériques.

Pour  $q_{x_{1:M}}$ , un algorithme progressif-regressif (ou forward-backward) de type filtre de Kalman-Rauch étendu<sup>14</sup> permet d'estimer à partir des mesures BOLD les états cachés qu'on ne peut pas mesurer directement. Cette étape dépend des moments des distributions  $q_\theta$  et  $q_\sigma$ .

La condition initiale est traitée comme les paramètres  $\vec{\theta}$ . Pour  $q_{x_0}$ , un algorithme de Gauss-Newton permet d'actualiser la moyenne  $\vec{\mu}_{x_0}$  en l'incrémentant de  $\vec{\Delta\mu}_{x_0}$ , ainsi que de calculer la nouvelle matrice de covariance  $\vec{\Sigma}_{x_0}$ .

$\vec{\Delta\mu}_{x_0}$  et  $\vec{\Sigma}_{x_0}$  dépendent des :

- moments de la distribution sur l'inverse de la variance du bruit  $q_\sigma$
- moments de la distribution a priori  $p(x_0)$

Avec la mise à jour à tour de rôle des distributions  $q_\square$ ,  $F$  augmente peu à peu. L'algorithme s'arrête quand un critère d'arrêt choisi arbitrairement est atteint (soit après un nombre maximal d'itérations, soit quand l'augmentation de  $F$  est inférieure à un certain seuil<sup>15</sup>).

#### 10.3.5.4 Entrées et sorties de VBA

Nous nous intéressons particulièrement à la densité de probabilité gaussienne a posteriori  $q_\theta^{opt}(\vec{\theta}) \approx p(\vec{\theta} | y_{1:M})$ . VBA nous permet d'en estimer le vecteur des moyennes  $\vec{\mu}_\theta$  et la matrice de covariance  $\vec{\Sigma}_\theta$ .

$$q_\theta^{opt}(\vec{\theta}) = \mathcal{N}(\vec{\mu}_\theta, \vec{\Sigma}_\theta) \quad (10.112)$$

Nous avons besoin de préciser en entrée de VBA :

14. "étendu" pour prendre en compte la non linéarité des fonctions  $\vec{f}$  et  $\vec{g}$

15. <https://mbb-team.github.io/VBA-toolbox/wiki/Controlling-the-inversion-using-VBA-options/>

- les données expérimentales  $(\vec{y}_1, \dots, \vec{y}_M)$
- les forçages  $u(t)$  (pour chaque pas de temps  $\Delta t$  qui échantillonne les durées entre 2 mesures)
- la fonction  $\vec{e}$ , les matrices jacobiennes  $\overline{\frac{\partial \vec{e}}{\partial x}}$  et  $\overline{\frac{\partial \vec{e}}{\partial \theta}}$  (respectivement par rapport aux états cachés  $\vec{x}$  et aux paramètres  $\vec{\theta}$ , les calculs des gradients se trouvent en annexe 17.1.4)
- la fonction d'observation  $\vec{g}$  et la matrice jacobienne  $\overline{\frac{\partial \vec{g}}{\partial x}}$  (par rapport aux états cachés  $\vec{x}$ , cf annexe 17.1.5).

Dans VBA, il est possible d'inverser des paramètres BOLD  $\vec{\phi}$  ([Daunizeau et al., 2009], [Daunizeau et al., 2014]) et dans ce cas là nous devons calculer la jacobienne de  $\vec{g}$  par rapport à ces paramètres. Dans notre cas, nous n'inversons pas de paramètres BOLD. Nous n'avons donc pas de jacobienne supplémentaire à calculer.

- les moments des gaussiennes a priori (moyenne et matrice de covariance) :
  - sur les paramètres :  $\vec{\mu}_\theta^0$  et  $\overline{\Sigma}_\theta^0$
  - sur la condition initiale :  $\vec{\mu}_{x_0}^0$  et  $\overline{\Sigma}_{x_0}^0$
- les paramètres de la distribution Gamma a priori sur l'inverse de la variance du bruit (i.e. les hyperparamètres de bruit a priori) :  $a_\sigma^0$  et  $b_\sigma^0$ . Ce sont deux scalaires.

Nous indiquerons les valeurs utilisées pour les moments des densités de probabilité a priori dans la section 10.4, dans laquelle nous allons tester la fiabilité de la procédure d'inversion en passant par des simulations numériques. Ce seront les mêmes densités de probabilité a priori lors de la mise en oeuvre de la procédure d'inversion sur un jeu de données au chapitre 11.

En sortie de VBA, nous obtenons :

- les moments des distributions gaussiennes a posteriori sur :
  - sur la condition initiale  $q_{x_0}^{opt} : \vec{\mu}_{x_0}$  et  $\overline{\Sigma}_{x_0}$
  - sur les états cachés  $q_{x_{1:M}}^{opt}$
  - sur les paramètres hémodynamiques  $q_\theta : \vec{\mu}_\theta$  et  $\overline{\Sigma}_\theta$
- les hyperparamètres de bruit a posteriori :  $a_\sigma$  et  $b_\sigma$ . Ce sont deux scalaires.
- la valeur de  $F^{opt} \equiv F(q^{opt})$  (approximation de la log évidence du modèle).

### 10.3.6 Extension à d'autres possibilités d'inversion

Jusqu'ici nous avons divisé le dérouls temporel du signal BOLD pendant la tâche en deux : des périodes de stimulation No Think (qui permettent d'inverser  $\overrightarrow{sn_{stim}} \equiv \overrightarrow{sn_{NT}}$ <sup>16</sup>)

---

16. En toute rigueur, nous inversons des paramètres transformés, mais il existe une relation bijective entre paramètres transformés et non transformés

et des périodes hors stimulation (qui permette d'inverser  $\overrightarrow{sn_B}$ ). Dans ce paragraphe, nous nous donnons la possibilité d'un autre découpage de la série temporelle BOLD qui permet de distinguer les autres conditions expérimentales du paradigme TNT, présentées à la section 2.2.3.2.

### 10.3.6.1 Trois paramètres par région

Les périodes hors stimulation constituent donc un mélange de signaux au repos et de Think. Nous allons distinguer ces deux conditions expérimentales de repos et de Think, en plus du No Think. Pour chaque région, nous allons inverser un paramètre hémodynamique pour le repos (ou niveau de base,  $\overrightarrow{sn_B}$ ) et pour le Think ( $\overrightarrow{sn_T}$ ), en plus de la condition No Think  $\overrightarrow{sn_{NT}}$ . C'est ce que nous avons représenté dans le tableau de la figure 10.6 sur la deuxième ligne.

Pour chaque région  $i$  l'équation sur le signal de vasodilatation (non transformé) peut s'écrire :

$$\dot{z}_i = sn_{B_i}\delta_{u(t),0} + sn_{T_i}\delta_{u(t),1} + sn_{NT_i}\delta_{u(t),2} - \kappa z_i(t) - \gamma(f_i(t) - 1) \quad (10.113)$$

$$\text{Avec } \begin{cases} u(t) = 0 \text{ en période de non stimulation} \\ u(t) = 1 \text{ en période de stimulation Think} \\ u(t) = 2 \text{ en période de stimulation No Think} \end{cases} \quad (10.114)$$

Le vecteur  $\overrightarrow{\theta}$  des paramètres hémodynamiques inconnus que nous souhaitons estimer est de dimension  $n_\theta = 3N$  (3 paramètres par région) et contient les paramètres transformés ( $sn_{B_i}$ ,  $sn_{T_i}$  et  $sn_{NT_i}$ ) pour chaque région  $i$ .

Comme pour le cas de l'inversion à 2 paramètres par région, ces  $3N$  paramètres sont a priori indépendants et leur densité de probabilité a priori sont identiques (section 10.3.2.2). Tout le reste est strictement identique à l'inversion de 2 paramètres.

### 10.3.6.2 Quatre paramètres par région

Avec l'inversion de 3 paramètres par région, le dérouls temporel du signal BOLD pendant la tâche a été divisé en 3 : des périodes de stimulation No Think (qui permettent d'inverser  $\overrightarrow{sn_{NT}}$ ), des périodes de stimulation Think (qui permettent d'inverser  $\overrightarrow{sn_T}$ ) et des périodes hors stimulation (qui permette d'inverser  $\overrightarrow{sn_B}$ ).

La condition No Think constitue un mélange de signaux mesurés quand le participant a eu une intrusion (condition No Think intrusive) ou non (condition No Think non intrusive.). Nous allons diviser les périodes No Think en distinguant ces deux conditions, pour inverser un paramètre hémodynamique No Think intrusif ( $\overrightarrow{sn_{NT,int}}$ ) et No Think non intrusif ( $\overrightarrow{sn_{NT,Nint}}$ ), en plus des paramètres pour le repos ( $\overrightarrow{sn_B}$ ) et pour le Think ( $\overrightarrow{sn_T}$ ). C'est ce que nous avons représenté dans le tableau de la figure 10.6 sur la troisième ligne.

Pour chaque région  $i$  l'équation sur le signal de vasodilatation (non transformé) peut s'écrire :

$$\dot{z}_i = sn_{B_i}\delta_{u(t),0} + sn_{T_i}\delta_{u(t),1} + sn_{NT,int_i}\delta_{u(t),2} + sn_{NT,Nint_i}\delta_{u(t),3} - \kappa z_i(t) - \gamma(f_i(t) - 1) \quad (10.115)$$

$$\text{Avec } \begin{cases} u(t) = 0 \text{ en période de non stimulation} \\ u(t) = 1 \text{ en période de stimulation Think} \\ u(t) = 2 \text{ en période de stimulation No Think intrusive} \\ u(t) = 3 \text{ en période de stimulation No Think non intrusive} \end{cases} \quad (10.116)$$

Le vecteur  $\vec{\theta}$  des paramètres hémodynamiques inconnus que nous souhaitons estimer est de dimension  $n_\theta = 4N$  (4 paramètres par région) et contient les paramètres transformés ( $sn_{B_i}$ ,  $sn_{T_i}$ ,  $sn_{NT,int_i}$  et  $sn_{NT,Nint_i}$ ) pour chaque région  $i$ . Ces  $4N$  paramètres sont a priori indépendants et leur densité de probabilité a priori sont identiques (section 10.3.2.2). Tout le reste est strictement identique à l'inversion de 2 paramètres.

Les sorties d'inversion de cette première étape donnent des moyennes de paramètres transformés (2, 3 ou 4 par région), qu'il est possible d'écrire en paramètres non transformés (2, 3 ou 4 par région) en utilisant la bijection définie en 10.3.2.2.

La seconde étape de l'inversion se base sur un modèle neural du contrôle inhibiteur. Nous avons besoin uniquement de deux jeux de  $sn^*$  stimulés et non stimulés, qui représentent des périodes pendant lesquelles le contrôle inhibiteur s'exerce ( $sn_{stim}$ ) et des périodes où il ne s'exerce pas ( $sn_B$ ).

Parmi les 2, 3 ou 4 paramètres par région inversés à la première étape, nous n'en sélectionnons donc que 2 pour la seconde étape. Ceux-ci sont entourés sur la figure 10.6.



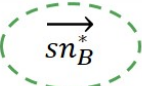
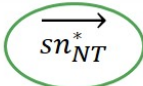
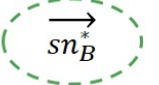
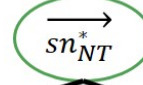
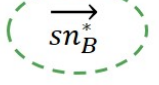
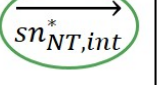
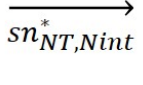
| Scénario d'inversion | Paramètres inversés                                                                                                                                                 |                                                                                                                                                                        |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                    |                                                                                    |                                                                                      |
| 2                    |   |                                                                                      |
| 3                    |   |   |

**Figure 10.6** – Récapitulatif des scénarios d'inversion.

Signification des flèches : Pour passer de l'un à l'autre, le décours temporel qui a servi à l'inversion d'un paramètre (e.g.  $sn_{NT}$ ) peut être divisé pour séparer deux conditions expérimentales (e.g. No Think intrusive et non intrusive) et inverser deux paramètres distincts (e.g.  $sn_{NT,int}$  et  $sn_{NT,Nint}$ ) au scénario suivant.

Les paramètres entourés sont utilisés pour la seconde étape de l'inversion (en pointillés pour les valeurs de  $sn^*$  non stimulés et en trait plein pour des valeurs de  $sn^*$  stimulés).

### 10.3.7 Récapitulatif

Comme dans le chapitre 5, nous pouvons expliciter les quatre points qui caractérisent la première étape de notre procédure d'inversion pour analyser les données d'un participant.

#### Première étape de l'inversion

- le modèle dynamique considéré : modèle du Ballon-Chambre à air étendu et modèle BOLD, avec  $sn_B^*$  et  $sn_{stim}^*$  comme forçages (10.1.2)
- le modèle statistique de la mesure : réseau bayésien dynamique avec une chaîne de Markov cachée
- les paramètres d'intérêt du modèle statistique : paramètres hémodynamiques  $\theta = (sn_B^*, sn_{stim}^*)$ , condition initiale hémodynamique  $x_0$  et inverse de la variance du bruit  $\sigma$  sur l'observable BOLD
- la méthode d'inversion de ces paramètres : méthode d'inversion bayésienne variationnelle (mise en oeuvre numérique avec la toolbox VBA <sup>a</sup>)

a. <https://mbb-team.github.io/VBA-toolbox/>

## 10.4 Test de fiabilité de la première étape : inversion de simulations avec différents niveaux de bruit

### 10.4.1 Simulations

#### 10.4.1.1 Une simulation

Nous simulons une série temporelle de signaux BOLD pour  $N = 15$  régions, sur 940  $TR$ . Pour cela, nous fixons : une condition initiale  $\overrightarrow{x_{0_{simu}}}$ <sup>17</sup>, les paramètres  $\overrightarrow{\theta_{simu}}$  (deux, trois ou quatre  $\tilde{sn}$  par région)<sup>18</sup> et l'inverse de la variance du bruit ( $\sigma_{simu}$ ). Nous lançons une simulation en intégrant numériquement le système d'équations différentielles tous les pas de temps  $\Delta t$ . Ceci nous permet d'obtenir un jeu de mesures tous les  $TR = n\Delta t$ .

#### 10.4.1.2 Tester différents niveaux de bruit

Nous lançons quatre simulations avec quatre niveaux de bruit différents. Pour cela, nous fixons dans le code l'inverse de la variance du bruit  $\sigma_{simu}$  de quatre manière différentes. Du plus petit niveau de bruit au plus grand :

- A) aucun bruit :  $\sigma_{simu} = \infty$
- B)  $\sigma_{simu} = 10^4$
- C)  $\sigma_{simu} = 10^3$
- D)  $\sigma_{simu} = 10^2$ .

La condition initiale  $\overrightarrow{x_{0_{simu}}}$  et les paramètres  $\overrightarrow{\theta_{simu}}$  sont identiques entre les différentes situations. Seul le niveau de bruit varie.

La variance du bruit sur le BOLD vaut  $\frac{1}{\sigma_{simu}}$ . L'écart-type du bruit sur le BOLD vaut alors  $\frac{1}{\sqrt{\sigma_{simu}}}$ , ce qui donne pour :

- B)  $10^{-2}$ ,
- C)  $10^{-3/2} \approx 3 \cdot 10^{-2}$ ,
- D)  $10^{-1}$ .

Au chapitre 7, nous avons vu que sous l'hypothèse de notre modèle l'amplitude maximale du signal BOLD d'une région vaut 0.04<sup>19</sup>. A partir du cas B), nous travaillons donc avec des signaux dont l'écart-type du bruit est du même ordre de grandeur ( $10^{-2}$ ) ou d'un ordre de grandeur supérieur à l'amplitude maximale du signal. Les signaux simulés sont donc très bruités, notamment dans le cas D).

---

17. Nous avons testé avec une condition initiale contenue ou non dans l'a priori

18. tirés aléatoirement dans la gaussienne a priori

19. différences entre les valeurs de BOLD obtenues pour  $sn = 1$  et pour  $sn = 0$

## 10.4.2 Inversion des simulations

Pour chacune des quatre situations précédemment décrites, nous souhaitons obtenir une estimation des paramètres  $\overrightarrow{\theta}_{simu}$ , c'est-à-dire une estimation de la densité de probabilité a posteriori que suit la variable aléatoire  $\overrightarrow{\theta}^{va}$ . Pour cela, nous devons fixer les densités de probabilité a priori suivantes.

### 10.4.2.1 Densités de probabilité a priori

#### 10.4.2.1.1 Paramètres hémodynamiques

Les paramètres sur lesquels portent l'inversion sont les paramètres transformés  $\tilde{sn}_i$ . La densité de probabilité a priori que suit la variable aléatoire  $\overrightarrow{\theta}^{va} \equiv \overrightarrow{\tilde{sn}}$  est une gaussienne multidimensionnelle de moyenne  $\vec{\mu}_\theta = \vec{0}$ . La matrice de covariance  $\overline{\Sigma}_\theta$  est une matrice diagonale (i.e. a priori les paramètres sont indépendants entre eux). Chaque élément de la diagonale vaut 3. Nous représentons la densité de probabilité a priori que suit chaque variable aléatoire  $\theta_i^{va} \equiv \tilde{sn}_i$  dans la figure 10.7(Gauche) : une loi normale de variance égale à 3. En nous rapportant à l'annexe 17.1.2, nous pouvons tracer la densité de probabilité a priori correspondante pour les paramètres non transformés  $sn_i$  (cf figure 10.7(Droite)). La densité de probabilité a priori choisie permet de couvrir tout l'intervalle des valeurs possibles de  $sn_i$  (entre 0 et 1) et est relativement non informatif pour permettre de ne favoriser aucune valeur de  $sn_i$  dans cet intervalle.

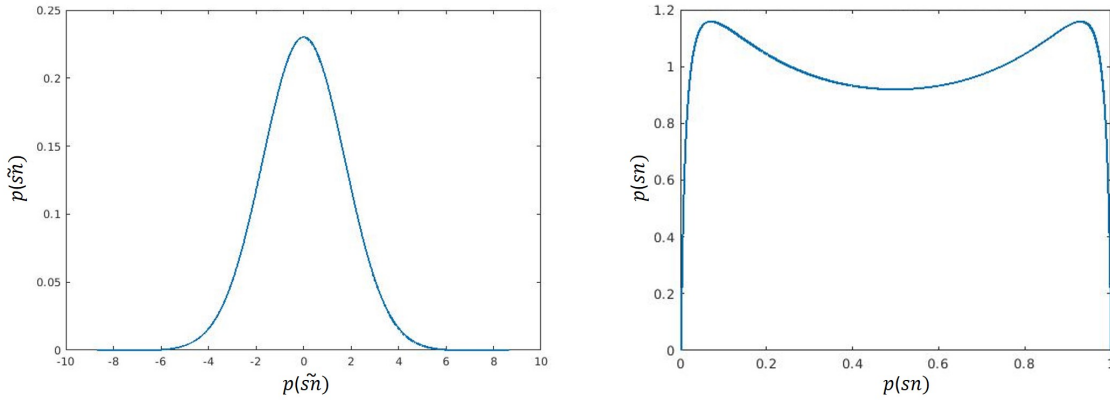


**Figure 10.7** – Gauche : Densité de probabilité a priori que suit chaque paramètre indépendant transformé  $\tilde{sn}_i : \mathcal{N}(0, 3)$ , Droite : Densité de probabilité a priori correspondante que suit chaque paramètre indépendant non transformé  $sn_i$ .

#### 10.4.2.1.2 Paramètres de bruit

La densité de probabilité a priori que suit l'inverse de la variance du bruit est une loi gamma de paramètres  $a_\sigma = 1$  et  $b_\sigma = 10^{-4}$ . Nous la représentons sur la figure 10.8. L'espérance de cette distribution vaut  $E(\sigma^{va}) = \frac{a_\sigma}{b_\sigma} = 10^4$ .

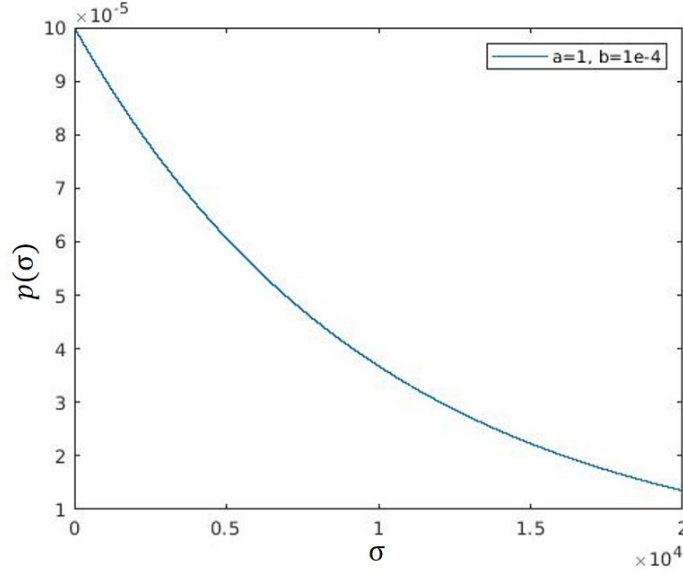


**Figure 10.8** – Distribution de probabilité a priori que suit  $\sigma^{va}$  : loi gamma de paramètres  $a_\sigma = 1$  et  $b_\sigma = 10^{-4}$ .

Nous nous assurons que cet a priori est pertinent pour les inversions des simulations précédemment décrites : les valeurs  $\sigma = 10^4$ ,  $\sigma = 10^3$  et  $\sigma = 10^2$  sont contenues dans l'a priori.

#### 10.4.2.1.3 Condition initiale

Pour rappel, le modèle direct a été écrit en variables transformées pour les flux, volumes et concentrations en déoxyhémoglobine. Le signal de vasodilatation n'a pas été réécrit puisque défini sur  $\mathbb{R}$  (cf section 10.3.2.3). La condition initiale s'écrit  $\vec{x}_0 = (\vec{z}, \vec{f}, \vec{v}, \vec{q})$ .

La densité de probabilité a priori que suit la variable aléatoire  $\vec{X}_0$  est une gaussienne multidimensionnelle, de moyenne  $\vec{\mu}_0 = \vec{0}$  et de matrice de covariance diagonale, avec  $10^{-4}$  comme éléments sur la diagonale et 0 partout ailleurs. Ceci revient à dire que chaque composante est indépendante et que nous fixons pour chaque  $Z_i(0)$ ,  $\tilde{F}_i(0)$ ,  $\tilde{V}_i(0)$  et  $\tilde{Q}_i(0)$  une densité de probabilité a priori gaussienne de moyenne nulle<sup>20</sup> et de variance  $10^{-4}$ , représentée sur la figure 10.9 (Gauche).

20. ces valeurs correspondent à celles données par [Deco et al., 2018] et [Wang et al., 2019] :  $z_i(0) = 0$  et  $f_i(0) = 1$ ,  $v_i(0) = 1$  et  $q_i(0) = 1$ , ce qui correspond pour les variables transformées à  $\tilde{f}_i(0) = 0$ ,  $\tilde{v}_i(0) = 0$  et  $\tilde{q}_i(0) = 0$

En nous servant de l'annexe 17.1.3, nous pouvons visualiser la densité de probabilité correspondante pour les variables non transformées  $F_i(0)$ ,  $V_i(0)$  et  $Q_i(0)$  (cf figure 10.9, Droite).

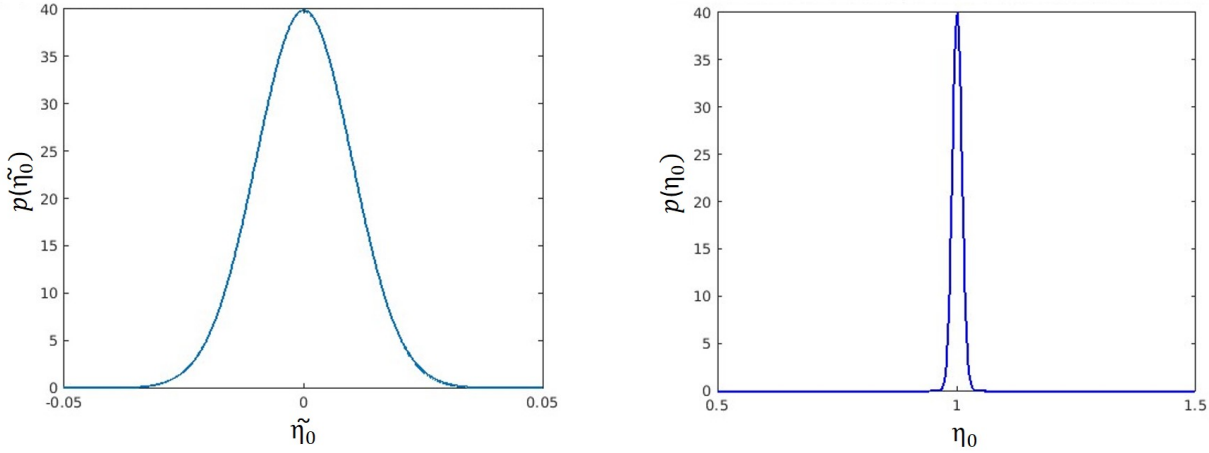


**Figure 10.9** – Gauche : Densité de probabilité a priori normale pour  $Z_i(0)$ ,  $\tilde{F}_i(0)$ ,  $\tilde{V}_i(0)$  et  $\tilde{Q}_i(0)$  :  $\mathcal{N}(0, 10^{-4})$ . Droite : Densité de probabilité lognormale correspondante pour  $F_i(0)$ ,  $V_i(0)$  et  $Q_i(0)$ .

#### 10.4.2.2 Fixer et inverser la condition initiale ?

Nous présentons dans la section suivante les résultats en sortie d'inversion pour deux cas distincts :

- la condition initiale fixée pour la simulation est contenue dans l'a priori
- la condition initiale fixée pour la simulation n'est pas contenue (i.e. est éloignée) de l'a priori.

Pour chaque cas, nous avons inversé ou non la condition initiale. Ne pas inverser la condition initiale signifie que la densité a posteriori de la condition initiale est exactement égale à celle a priori (présenté au paragraphe 10.4.2.1.3). Nous avons remarqué que le fait d'inverser ou non la condition initiale change de manière négligeable l'inversion des paramètres hémodynamiques ( $\vec{\theta}$ ) et des hyperparamètres de bruit ( $a_\sigma$  et  $b_\sigma$ ).

Dans toute la suite, nous présentons en conséquence uniquement les figures de résultats obtenues sans inverser la condition initiale. En sortie d'inversion, nous obtenons 1) les moments de la densité a posteriori sur les paramètres hémodynamiques et 2) les hyperparamètres de bruit a posteriori  $a_\sigma$  et  $b_\sigma$  i.e. les valeurs des paramètres de la loi gamma a posteriori que suit  $\sigma^{va}$ , l'inverse de la variance du bruit.

#### 10.4.2.3 Sorties d'inversion pour une condition initiale contenue dans l'a priori

Nous avons simulé une série temporelle de signaux BOLD pour  $N = 15$  régions, sur 940  $TR$ , avec une condition initiale égale à  $\overrightarrow{x_{0_{simu}}} = \vec{0}$ , qui est la moyenne de l'a priori. La

condition initiale est donc contenue dans l'a priori.

L'inversion est très rapide : elle dure de 2 à 4 minutes en fonction du nombre de paramètres à estimer par région. Nous pouvons trouver en annexe 17.1.6, les résultats d'inversion pour les paramètres hémodynamiques transformés  $\tilde{sn}$ . Pour chaque niveau de bruit, nous avons représenté les valeurs moyennes de la densité a posteriori en fonction des valeurs fixées pour la simulation.

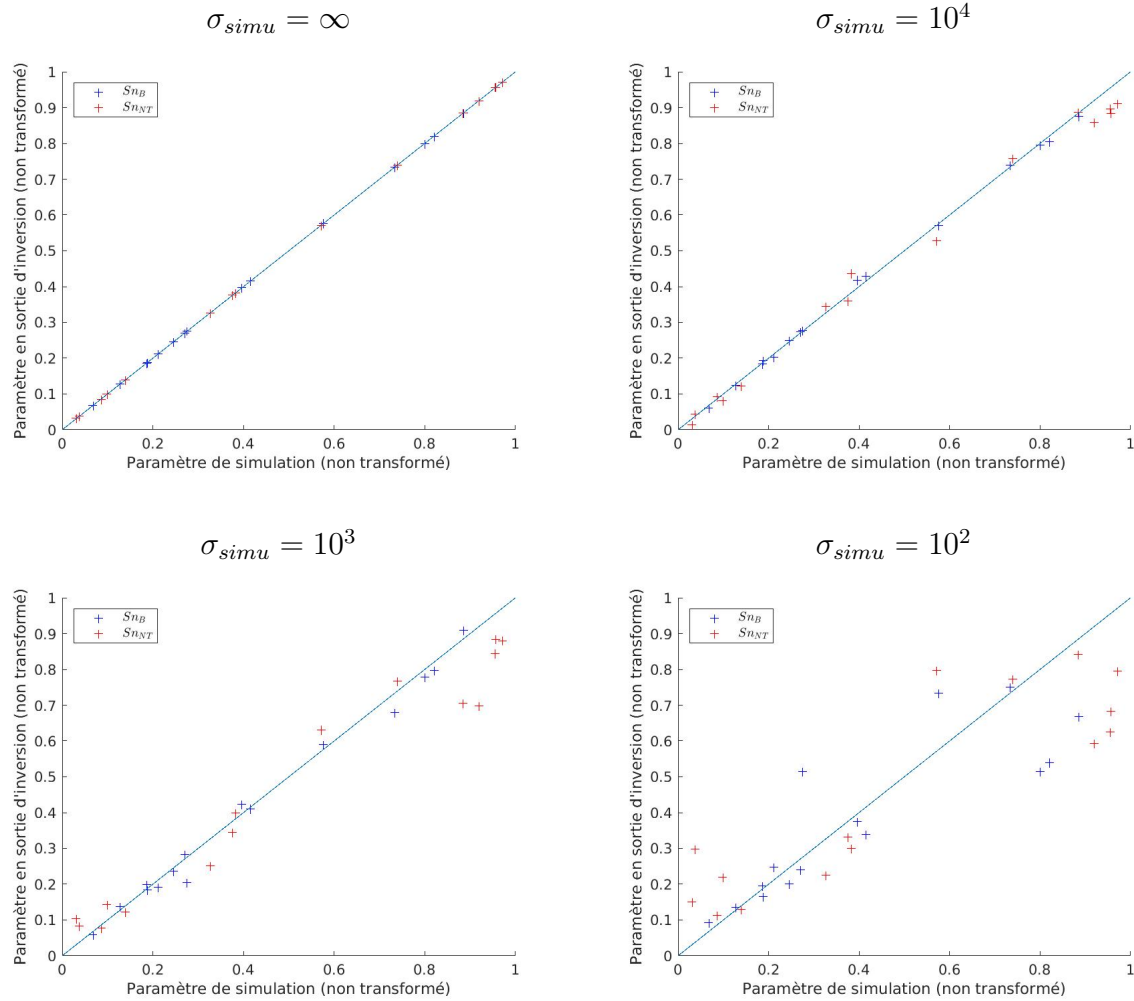
La bijection définie à la section 10.3.2.2 permet d'obtenir à partir des paramètres transformés  $\tilde{sn}$  les valeurs non transformées  $sn$  correspondantes. Nous l'appliquons aux paramètres de simulation  $sn_{\tilde{simu}}$  et aux moyennes estimées en sortie d'inversion  $sn_{\tilde{inv}}$ . Dans ce chapitre, pour un niveau de bruit fixé, nous représentons la valeur estimée pour un paramètre hémodynamique non transformé  $sn_{inv}$ , en fonction de sa valeur fixée pour la simulation  $sn_{simu}$ . Pour chaque région, nous avons estimé deux, trois ou quatre paramètres, comme présenté à la section 10.3.6. Les résultats sont présentés dans les graphes aux sections 10.4.2.3.1, 10.4.2.3.2 et 10.4.2.3.3. Un point du graphe représente un paramètre pour une région.

Dans chaque cas, nous vérifions que nous retrouvons exactement les paramètres de simulation en sortie d'inversion lorsque le signal n'est pas bruité. Cela valide la première étape de la procédure d'inversion, sans bruit. Comme attendu, plus la simulation est bruitée (i.e. plus  $\sigma$ , l'inverse de la variance du bruit, diminue) et moins bien nous récupérons les paramètres de simulation en sortie d'inversion.

Nous avons ensuite vérifié l'estimation des paramètres de bruit. Pour chaque simulation, nous avons fixé  $\sigma_{simu}$  qui est l'inverse de la variance du bruit sur le signal BOLD. La procédure d'inversion estime  $a_\sigma$  et  $b_\sigma$ , les paramètres de la loi gamma a posteriori que suit  $\sigma^{va}$ . A partir des valeurs inversés de  $a_\sigma$  et  $b_\sigma$ , nous calculons l'espérance de la loi gamma a posteriori : elle est égale au rapport  $\frac{a_\sigma}{b_\sigma}$ . L'ensemble des valeurs sont reportées dans les tableaux aux sections 10.4.2.3.1, 10.4.2.3.2 et 10.4.2.3.3. Nous vérifions que l'espérance s'approche bien des valeurs de  $\sigma_{simu}$  fixées pour chaque simulation. L'estimation du bruit est donc cohérente.

### 10.4.2.3.1 2 paramètres par région

#### Inversion des $sn$



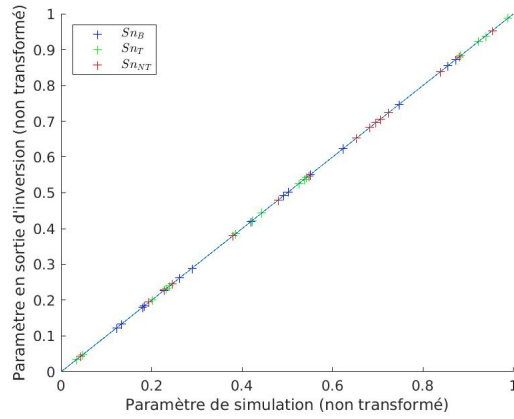
#### Bruit

| Niveau de bruit ( $\sigma_{simu}$ ) | $a_\sigma$ inversé | $b_\sigma$ inversé | Espérance $E(\sigma^{va})$ |
|-------------------------------------|--------------------|--------------------|----------------------------|
| $\infty$                            | 7051               | 0.00010021         | 70359986.8937              |
| $10^4$                              | 7051               | 0.70815            | 9956.9437                  |
| $10^3$                              | 7051               | 7.0549             | 999.4528                   |
| $10^2$                              | 7051               | 69.4685            | 101.4993                   |

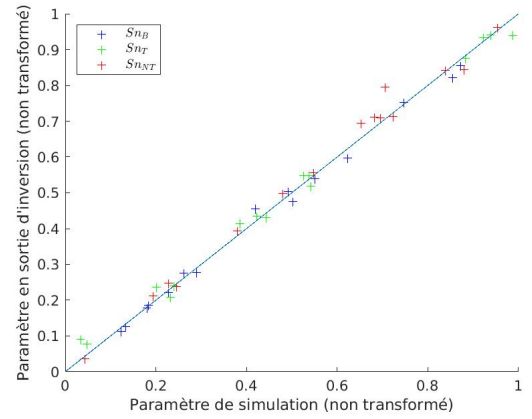
### 10.4.2.3.2 3 paramètres par région

#### Inversion des $sn$

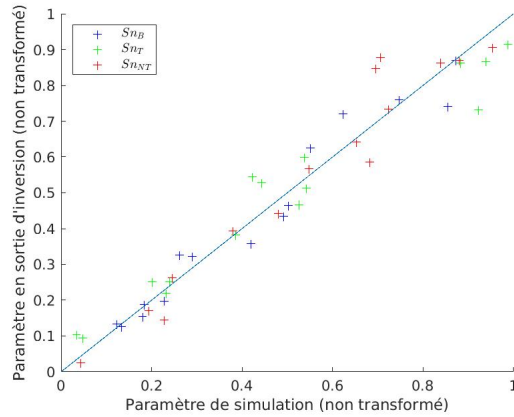
$$\sigma_{simu} = \infty$$



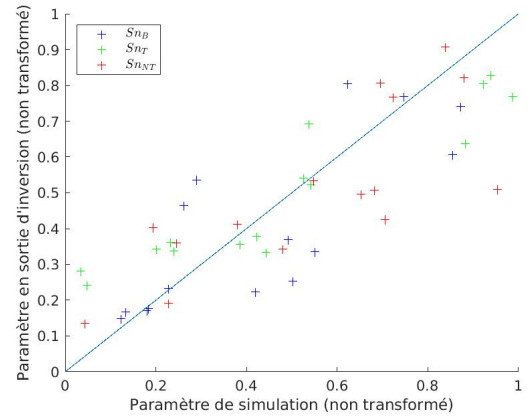
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$



$$\sigma_{simu} = 10^2$$



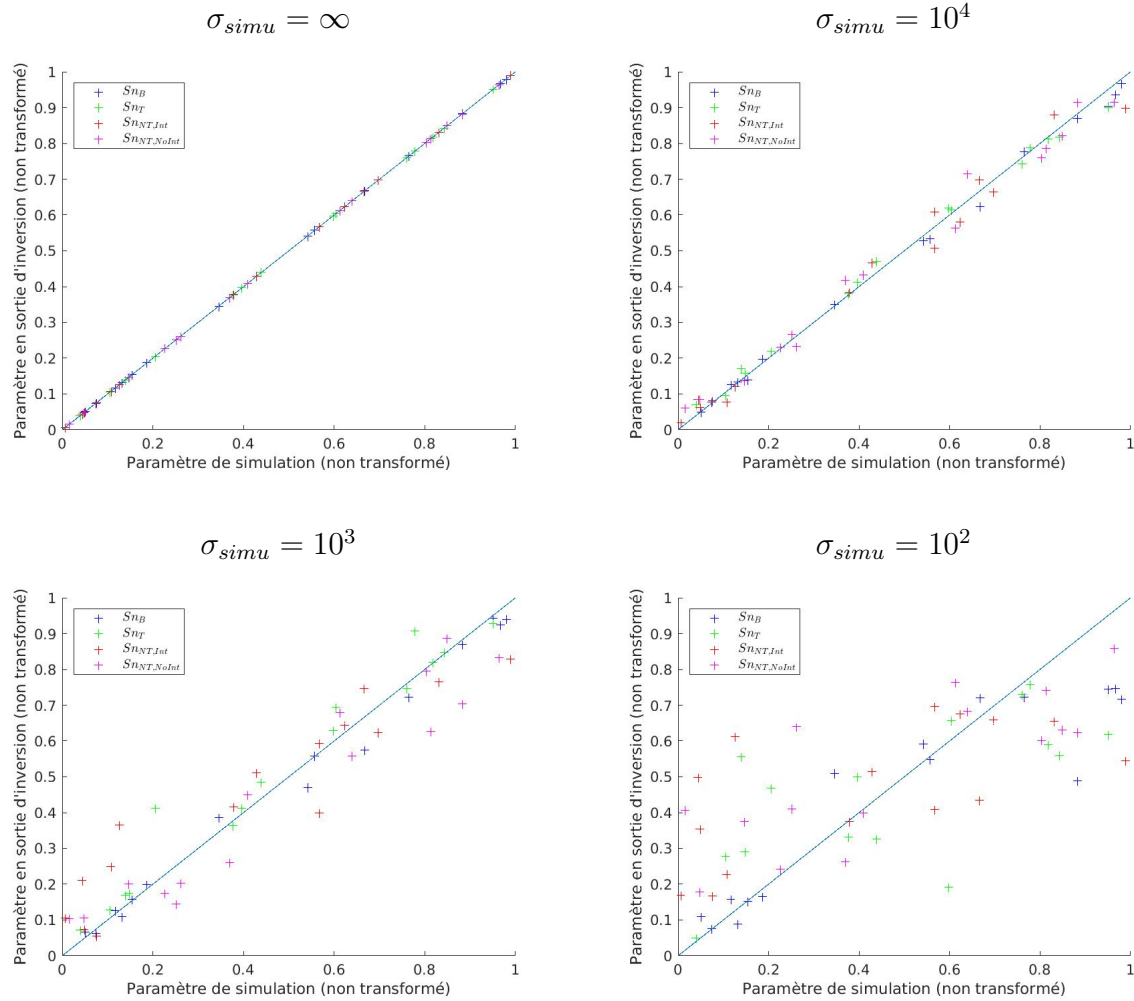
#### Bruit

| Niveau de bruit ( $\sigma_{simu}$ ) | $a_\sigma$ inversé | $b_\sigma$ inversé | Espérance $E(\sigma^{va})$ |
|-------------------------------------|--------------------|--------------------|----------------------------|
| $\infty$                            | 7051               | 0.00010032         | 70284922.7417              |
| $10^4$                              | 7051               | 0.69481            | 10148.1131                 |
| $10^3$                              | 7051               | 7.0767             | 996.3725                   |
| $10^2$                              | 7051               | 70.4455            | 100.0916                   |



### 10.4.2.3.3 4 paramètres par région

#### Inversion des $sn$



#### Bruit

| Niveau de bruit ( $\sigma_{simu}$ ) | $a_\sigma$ inversé | $b_\sigma$ inversé | Espérance $E(\sigma^{va})$ |
|-------------------------------------|--------------------|--------------------|----------------------------|
| $\infty$                            | 7051               | 0.00010043         | 70209691.1279              |
| $10^4$                              | 7051               | 0.69359            | 10165.9082                 |
| $10^3$                              | 7051               | 7.05               | 1000.1463                  |
| $10^2$                              | 7051               | 70.0175            | 100.7033                   |

#### 10.4.2.4 Sorties d'inversion pour une condition initiale non contenue dans l'a priori

Nous avons simulé une série temporelle de signaux BOLD pour  $N = 15$  régions, sur 940  $TR$ , avec une condition initiale en dehors de la densité a priori :  $\overrightarrow{x_{0_{simu}}} = \vec{1}$ .

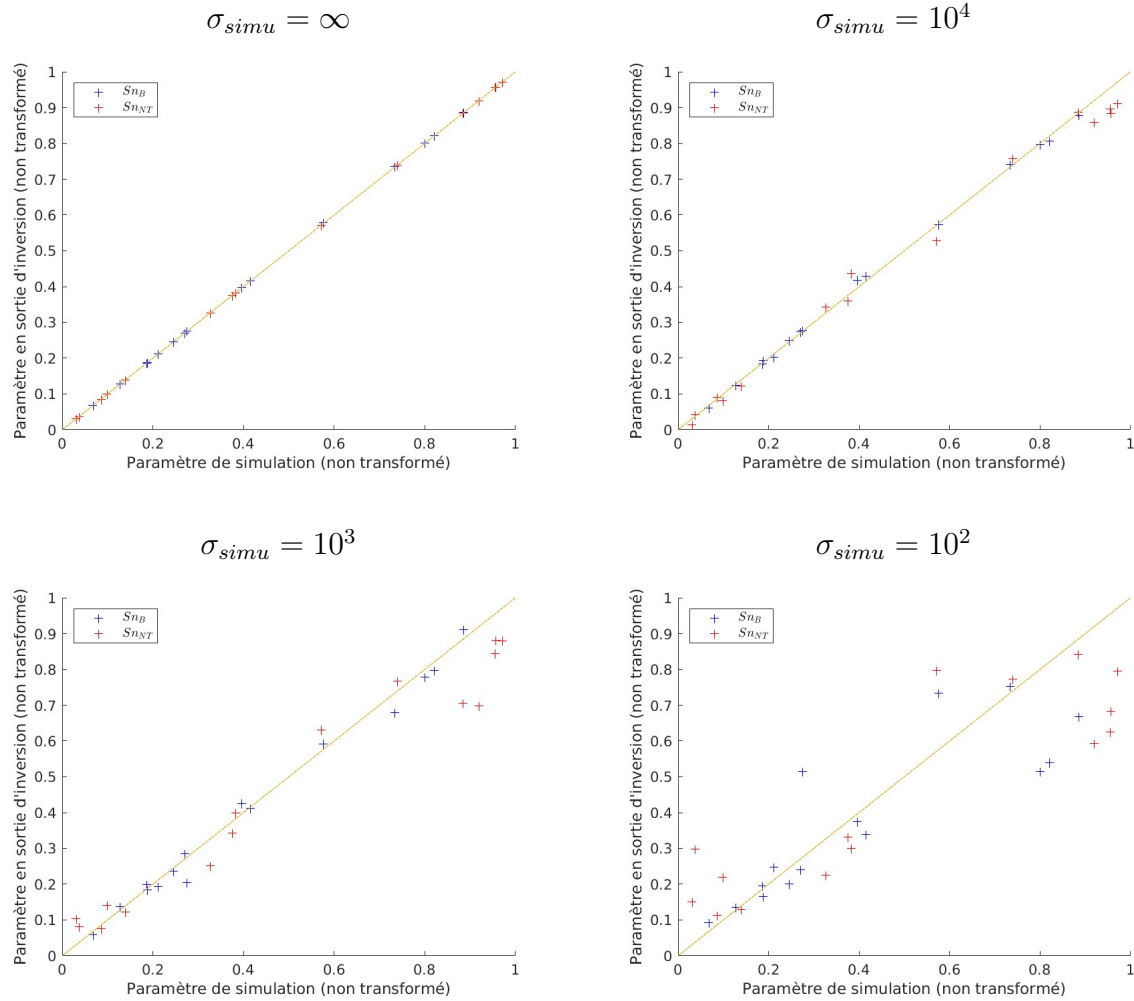
Comme pour le paragraphe précédent, pour un niveau de bruit fixé, nous représentons la valeur estimée pour un paramètre hémodynamique non transformé  $sn_{inv}$ , en fonction de sa valeur fixée pour la simulation  $sn_{simu}$ . Les résultats sont reportés dans les graphes des sections 10.4.2.4.1, 10.4.2.4.2 et 10.4.2.4.3. Les graphes équivalents pour les paramètres transformés se trouvent en annexe 17.1.6.

Les observations sont identiques : 1) nous validons la première étape de la procédure d'inversion sans bruit 2) plus le signal est bruité et moins bien nous récupérons les paramètres de simulation en sortie d'inversion.

Comme précédemment, nous vérifions également l'estimation des paramètres de bruit. Nous reportons dans les tableaux des sections 10.4.2.4.1, 10.4.2.4.2 et 10.4.2.4.3 l'estimation des paramètres  $a_\sigma$  et  $b_\sigma$ , ainsi que l'espérance de la loi gamma a posteriori que suit l'inverse de la variance du bruit  $\sigma^{va}$ . Pour chaque simulation, nous vérifions à nouveau que l'espérance a posteriori est proche de  $\sigma_{simu}$  : l'estimation du bruit est cohérente.

#### 10.4.2.4.1 2 paramètres par région

##### Inversion des $sn$



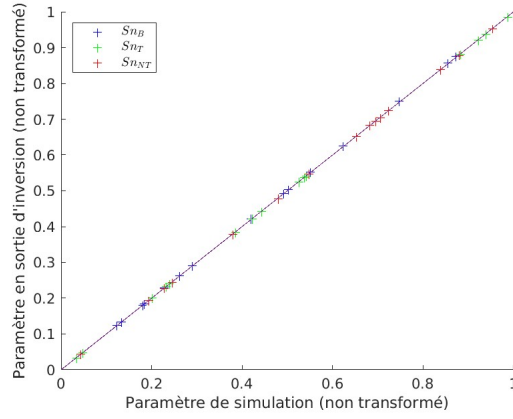
##### Bruit

| Niveau de bruit ( $\sigma_{simu}$ ) | $a_\sigma$ inversé | $b_\sigma$ inversé | Espérance $E(\sigma^{va})$ |
|-------------------------------------|--------------------|--------------------|----------------------------|
| $\infty$                            | 7051               | 0.0095139          | 741127.1018                |
| $10^4$                              | 7051               | 0.7186             | 9812.1121                  |
| $10^3$                              | 7051               | 7.0701             | 997.3012                   |
| $10^2$                              | 7051               | 69.4899            | 101.468                    |

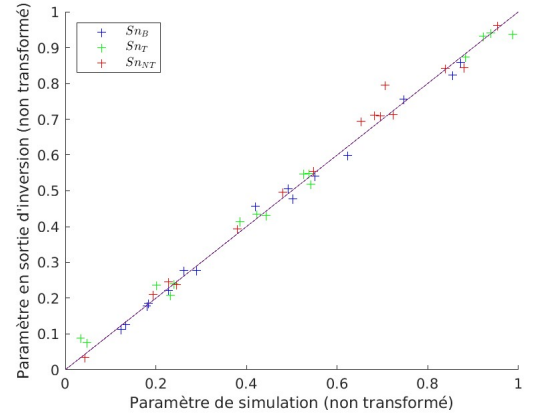
#### 10.4.2.4.2 3 paramètres par région

##### Inversion des $sn$

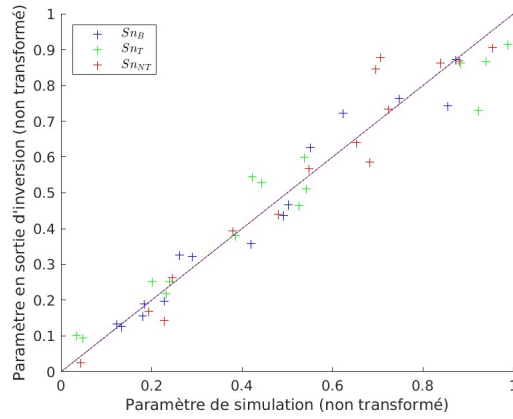
$$\sigma_{simu} = \infty$$



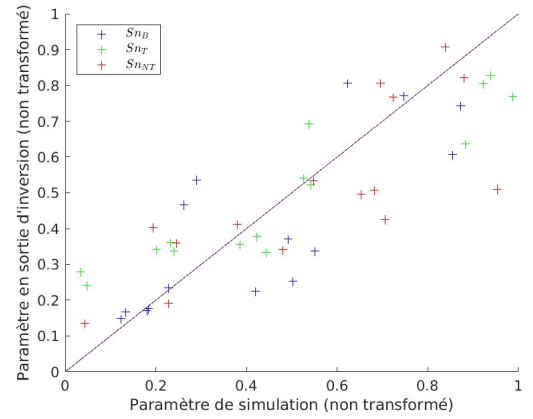
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$



$$\sigma_{simu} = 10^2$$

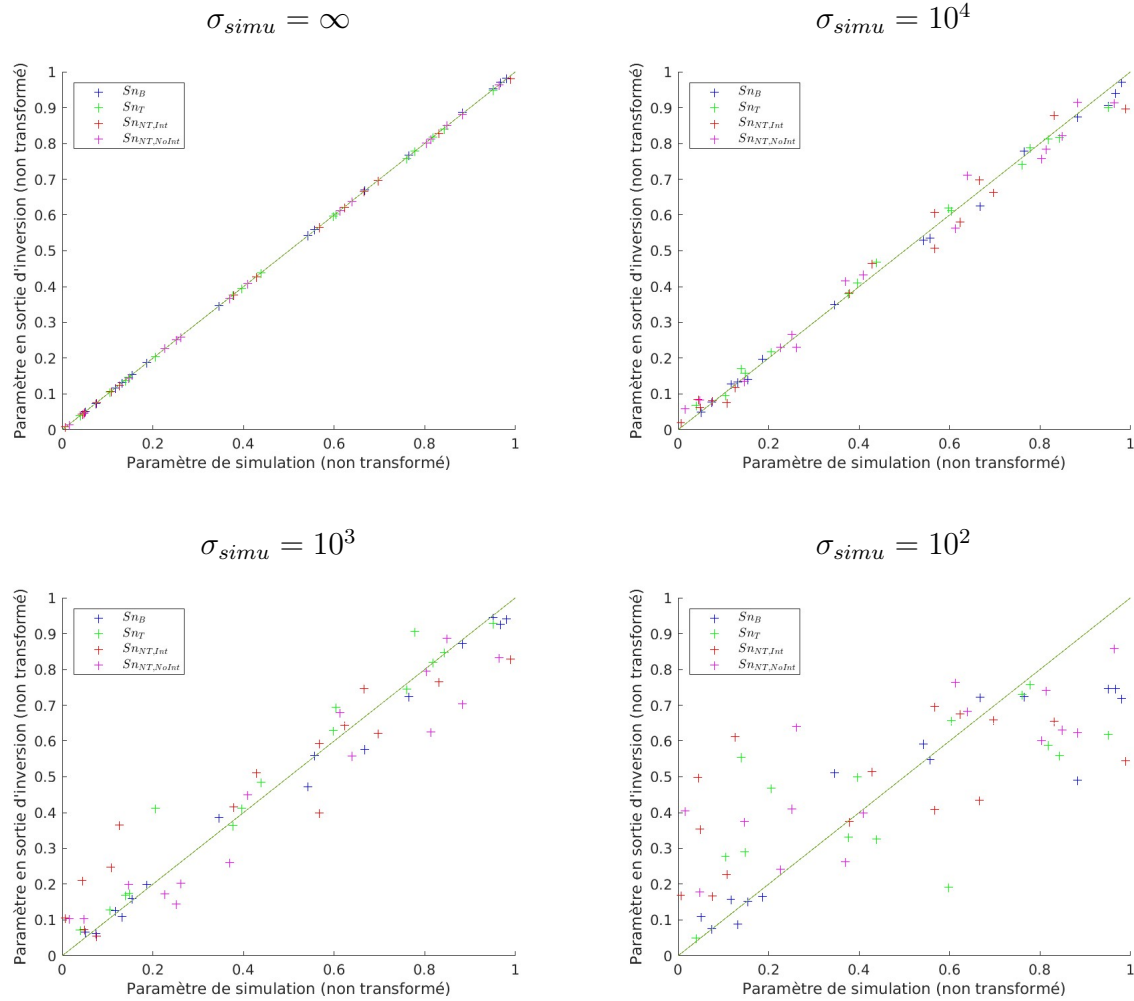


##### Bruit

| Niveau de bruit ( $\sigma_{simu}$ ) | $a_\sigma$ inversé | $b_\sigma$ inversé | Espérance $E(\sigma^{va})$ |
|-------------------------------------|--------------------|--------------------|----------------------------|
| $\infty$                            | 7051               | 0.0093648          | 752922.2074                |
| $10^4$                              | 7051               | 0.70496            | 10001.9601                 |
| $10^3$                              | 7051               | 7.0877             | 994.8288                   |
| $10^2$                              | 7051               | 70.4715            | 100.0547                   |

#### 10.4.2.4.3 4 paramètres par région

##### Inversion des $sn$



##### Bruit

| Niveau de bruit ( $\sigma_{simu}$ ) | $a_\sigma$ inversé | $b_\sigma$ inversé | Espérance $E(\sigma^{va})$ |
|-------------------------------------|--------------------|--------------------|----------------------------|
| $\infty$                            | 7051               | 0.0091522          | 770414.0756                |
| $10^4$                              | 7051               | 0.70262            | 10035.2711                 |
| $10^3$                              | 7051               | 7.055              | 999.4329                   |
| $10^2$                              | 7051               | 70.0465            | 100.6616                   |

## 10.5 Lancer la première étape sur toutes les régions en même temps ?

Nous avons vu que la première étape de l'inversion est découplée entre régions : à partir du décours temporel du signal BOLD d'une région, nous pouvons obtenir une estimation des  $\tilde{s}n$  de cette région, sans avoir besoin des décours BOLD des autres régions.

Dans tout ce qui a précédé, nous avons lancé avec VBA l'inversion du signal des  $N = 15$  régions en même temps : la série des mesures  $(\vec{y}_m)_{m \in [1:940]}$  contient des vecteurs de dimension  $(15, 1)$ . La série temporelle des états cachés  $(x_m)_{m \in [0:940]}$  contient des vecteurs de dimensions  $(4 \times 15, 1)$ . La densité a priori sur les paramètres est une gaussienne multidimensionnelle, avec une moyenne  $\vec{\mu}_\theta$  de dimension  $(15, 1)$  et une matrice de covariance a priori  $\overline{\overline{\Sigma}}_\theta$  de dimension  $(15, 15)$ . Nous avons spécifié qu'a priori tous les paramètres hémodynamiques sont indépendants : les termes hors diagonale de  $\overline{\overline{\Sigma}}_\theta$  sont nuls. A posteriori, les paramètres hémodynamiques entre régions restent indépendants<sup>21</sup> et ceux au sein d'une même région présentent des covariances non nulles. L'inversion des signaux semble donc bien être menée de manière indépendante entre les régions. Nous notons "1x15" une inversion des 15 régions en même temps.

Puisque la première étape de l'inversion est découplée entre régions, nous aurions aussi pu inverser le signal des régions une par une, soit mener 15 inversions de 1 région, notée "15x1". Nous avons voulu préciser les différences entre ces deux manières de procéder, de manière à nous assurer que nous pouvons continuer à travailler en inversant toutes les régions en même temps avec VBA, i.e. "1x15" régions.

Lorsque nous inversons en "1x15", en plus des paramètres hémodynamiques, nous estimons deux hyperparamètres de bruit pour toutes les régions :  $a_\sigma$  et  $b_\sigma$ . Ils définissent la loi gamma a posteriori que suit  $\sigma^{va}$ . Une réalisation  $\sigma$  de cette variable aléatoire est égale à l'inverse de la variance du bruit qui est identique pour toutes les régions. Autrement dit, quand nous inversons en "1x15", nous estimons un niveau de bruit identique pour toutes les régions.

Dans le cas "15x1", nous estimons deux hyperparamètres de bruit pour chaque région :  $a_{\sigma,i}$  et  $b_{\sigma,i}$ . Ces deux paramètres définissent la loi gamma a posteriori que suit  $\sigma_i^{va}$ . Une réalisation  $\sigma_i$  de cette variable aléatoire est égale à l'inverse de la variance du bruit de la région  $i$ . Nous estimons donc un niveau de bruit propre à chaque région. L'estimation du bruit est plus fine. Dans nos simulations précédentes,  $\sigma_{simu}$  est identique pour toutes les régions. Il n'est donc pas nécessaire d'inverser le signal des régions une par une (i.e. en "15x1") : l'inversion "1x15" est cohérente. Cependant, lorsque nous travaillons avec des données, le bruit entre les régions peut être suffisamment différent pour qu'une estimation plus fine du bruit ait une conséquence sur l'estimation des paramètres hémodynamiques. Nous avons donc inversé en "1x15" et "15x1" les données d'un participant (sain non exposé de la cohorte REMEMBER, cf 11) et nous avons comparé les résultats.

Bien que le bruit estimé pour chaque région soit un peu différent entre "15x1" et "1x15", l'estimation des paramètres hémodynamiques n'est pas affectée : les différences entre les moyennes et variances a posteriori obtenues avec les deux méthodes sont négligeables devant

21. la matrice de covariance a posteriori contient des blocs nuls hors diagonale

ces valeurs estimées. Nous notons également que l'inversion des 15 régions une par une, i.e. "15x1", met quinze fois plus de temps que l'inversion "1x15".

Puisque les résultats en sortie d'inversion ne sont pas modifiés pour les paramètres hémodynamiques lorsque nous procédons en "15x1" ou "1x15", nous présentons au chapitre [11](#) les résultats de l'inversion "1x15" menée sur les données d'un participant mesurées pendant la tâche Think/No Think.

## 10.6 Deuxième étape de l'inversion

### 10.6.1 Introduction

Comme nous l'avons vu au début de ce chapitre, la deuxième étape de l'inversion consiste à remonter aux paramètres neuraux à partir des valeurs des couples de points fixes en période non stimulée et stimulée ( $sn_B^*, sn_{stim}^*$ ). Le modèle neural est couplé. Cependant, la résolution algébrique du problème d'inversion des  $J_{gaba}$  et  $I_{stim}$  conduit à une technique d'inversion découplée entre régions, et plus précisément découplée entre chaque pool excitateur et chaque pool inhibiteur de chaque région.

Il va finalement nous suffir de résoudre une équation pour le pool excitateur de chaque région  $i$  afin d'estimer  $J_{gaba_i}$  à partir de toutes les valeurs de  $sn_B^*$ . Puis nous allons résoudre une équation pour le pool excitateur de chaque région  $i$  afin d'obtenir  $I_{stim_i}$  à partir de toutes les valeurs de  $sn_{stim}^*$ .

### 10.6.2 Nouvelle formulation du modèle

Nous reformulons le modèle neural présenté à la section 10.1 en distinguant pour chaque région le pool excitateur et le pool inhibiteur. Chaque pool peut ainsi être décrit par son système d'équations. Ce modèle reformulé est toujours couplé avec des couplages entre les pools d'une même région et entre régions.

Pour le pool excitateur d'une région  $i$ , le modèle au régime stationnaire peut s'écrire :

$$\begin{cases} 0 = f_{n_i}(sn_i^*, rn_i^*) = -\beta^E sn_i^* + \alpha^E T_{glu} rn_i^* (1 - sn_i^*) \\ rn_i^* = h_{n_i}(xn_i^*) = \frac{a_E xn_i^* - b_E}{1 - e^{-d_E(a_E xn_i^* - b_E)}} \\ xn_i^* = w_{n_i}(sn_i^*, z_{E_i}^*) = W_+ J_{nmda} sn_i^* + z_{E_i}^* \\ z_{E_i}^* = p_{n_i}(sg_i^*, sn_j^*, J_{gaba_i}, F_{E_i}) = -J_{gaba_i} sg_i^* + F_{E_i} + \sum_{j \neq i} k_{E_{ij}} \kappa_{ij} sn_j^* \end{cases} \quad (10.117)$$

Au régime stationnaire,  $z_{E_i}^*$  est égal au forçage que reçoit le pool excitateur de la région  $i$  (voir le modèle pour 1 pool excitateur écrit dans l'annexe de l'article au chapitre 9). Nous le représentons sur la figure 10.10.

$z_{E_i}^*$  dépend du couplage avec les autres régions, du couplage avec le pool inhibiteur de la région  $i$  et du forçage  $F_{E_i}$ .

$F_{E_i}$  désigne le forçage sur le pool excitateur et vaut :

- Hors stimulation,  $F_{E_i} = B_{E_i} = W_E I_0$ .
- En période de stimulation,  $F_{E_i} = B_{E_i} + I_{stim_i} = W_E I_0 + I_{stim_i}$

Pour le pool inhibiteur d'une région  $i$ , le modèle au régime stationnaire peut s'écrire :



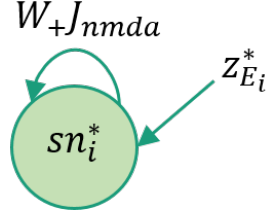


**Figure 10.10** – Représentation d'un pool excitateur, avec un autocouplage excitateur et un forçage.

$$\begin{cases} 0 = f_{g_i}(sg_i^*, rg_i^*) = -\beta^I sg_i^* + \alpha^I T_{gaba_i} rg_i^* (1 - sg_i^*) \\ rg_i^* = h_{g_i}(xg_i^*) = \frac{a_I xg_i^* - b_I}{1 - e^{-d_I(a_I xg_i^* - b_I)}} \\ xg_i^* = w_{g_i}(sg_i^*, z_{I_i}^*) = -W_- sg_i^* + z_{I_i}^* \\ z_{I_i}^* = p_{g_i}(sn_i^*, sn_j^*) = J_{nmda} sn_i^* + B_{I_i} + \sum_{j \neq i} (1 - k_{E_{ij}}) \kappa_{ij} sn_j^* \end{cases} \quad (10.118)$$

Au régime stationnaire,  $z_{I_i}^*$  est égal au forçage que reçoit le pool inhibiteur de la région  $i$  (voir le modèle pour 1 pool inhibiteur écrit dans l'annexe de l'article au chapitre 9). Nous le représentons sur la figure 10.11.

$z_{I_i}^*$  dépend du couplage avec les autres régions, du couplage avec le pool excitateur de la région  $i$  et de son forçage basal  $B_{I_i}$ .

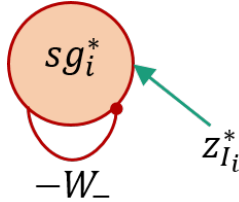


**Figure 10.11** – Représentation d'un pool inhibiteur, avec un autocouplage inhibiteur et un forçage.

### 10.6.3 Principe de la méthode d'inversion algébrique des paramètres neuraux

Avec la première étape de l'inversion, tous les  $sn_i^*$  hors stimulation et en période de stimulation sont connus et donc tous les  $z_{I_i}^*$  correspondant sont fixés.

Le système (10.118) décrit alors le comportement au régime stationnaire du pool inhibiteur de la région  $i$  avec un autocouplage négatif  $-W_-$  et un forçage fixé  $z_{I_i}^*$  (cf figure 10.11). Cela nous permettra d'en déduire la valeur des  $sg_i^*$  hors stimulation et en période de stimulation. Lorsque tous les  $sg_i^*$  sont connus, le système (10.117) décrit le comportement au stationnaire

du pool excitateur de la région  $i$  avec un autocouplage positif  $W_+ J_{nmda}$  et un forçage  $z_{I_i}^*$  (cf figure 10.10), qui dépend uniquement des paramètres que l'on souhaite inverser ( $J_{gaba}$  hors stimulation,  $J_{gaba}$  et  $I_{stim}$  en stimulation).

Nous détaillons dans la suite les étapes de l'inversion algébrique des paramètres neuraux, d'abord hors stimulation pour l'inversion de  $J_{gaba}$  puis en période de stimulation pour l'inversion de  $I_{stim}$ . Dans chacun des cas, nous travaillons région par région, d'abord sur chaque pool inhibiteur puis sur chaque pool excitateur. La méthode d'inversion est alors découplée par région et pour chaque pool de chaque région. Un schéma récapitulatif des étapes d'inversion est présent au paragraphe 10.6.5 et peut être consulté pendant la lecture de la section 10.6.4.

## 10.6.4 Inversion

### 10.6.4.1 Inversion des $J_{gaba}$ hors stimulation

#### 10.6.4.1.1 Modèle hors stimulation

Nous notons de manière générale  $\eta_{B_i}^*$  ou  $\eta_{B_j}^*$  toute variable  $\eta$  décrivant la région  $i$  ou la région  $j$  à son niveau de base i.e. hors période de stimulation.

Pour le pool excitateur de la région  $i$ , nous pouvons écrire le modèle (10.117) hors stimulation :

$$\begin{cases} 0 = f_{n_i}(sn_{B_i}^*, rn_{B_i}^*) \\ rn_{B_i}^* = h_{n_i}(xn_{B_i}^*) \\ xn_{B_i}^* = w_{n_i}(sn_{B_i}^*, z_{E_{B_i}}^*) \\ z_{E_{B_i}}^* = p_{n_i}(sg_{B_i}^*, sn_{B_j}^*, J_{gaba_i}, B_{E_i}) \end{cases} \quad (10.119)$$

Pour le pool inhibiteur de la même région, nous adaptons le modèle (10.118) pour l'écrire aussi hors stimulation :

$$\begin{cases} 0 = f_{g_i}(sg_{B_i}^*, rg_{B_i}^*) \\ rg_{B_i}^* = h_{g_i}(xg_{B_i}^*) \\ xg_{B_i}^* = w_{g_i}(sg_{B_i}^*, z_{I_{B_i}}^*) \\ z_{I_{B_i}}^* = p_{g_i}(sn_{B_i}^*, sn_{B_j}^*) \end{cases} \quad (10.120)$$

#### 10.6.4.1.2 1ère étape : travail sur le pool inhibiteur pour trouver les $sg_{B_i}^*$

Pour commencer, nous travaillons sur le pool inhibiteur de la région  $i$  au régime stationnaire, au niveau de base i.e. hors période de stimulation (modèle (10.120)).

Pour chaque région,  $sn_{B_i}^*$  et  $sn_{B_j}^*$  sont donnés par la première étape de l'inversion du BOLD, ce qui fixe :

$$z_{I_{B_i}}^* = p_{g_i}(sn_{B_i}^*, sn_{B_j}^*) = J_{nmda} sn_{B_i}^* + B_{I_i} + \sum_{j \neq i} (1 - k_{E_{ij}}) \kappa_{ij} sn_{B_j}^* \quad (10.121)$$

Nous cherchons numériquement le point fixe  $sg_{B_i}^*$ , avec la valeur de l'autocouplage et la valeur du forçage. Nous montrons dans l'annexe 17.2.1 que pour un pool inhibiteur avec une valeur fixée du forçage  $z_{I_{B_i}}^*$ , il existe un unique point fixe  $sg_{B_i}^*$  stable. Pour le trouver, deux méthodes sont donc possibles. Dans les deux cas, nous prenons soin de fixer la valeur de  $z_{I_{B_i}}^*$  d'après l'équation (10.121).

1. Nous pouvons lancer une recherche de point fixe en faisant tourner numériquement la dynamique, pour n'importe quelle valeur de condition initiale  $sg(t=0) \in [0, 1]$ . Quand  $sg$  est proche de 0 et que le système s'est stabilisé, nous trouvons la valeur approchée du point fixe  $sg^*$ .
2. Nous pouvons aussi écrire le modèle du pool inhibiteur de la région  $i$  au régime stationnaire sous la forme  $0 = F_g(sg_{B_i}^*, z_{I_{B_i}}^*)$ . Pour la valeur de  $z_{I_{B_i}}^*$  calculée avec (10.121), nous procédons alors par dichotomie pour rechercher une valeur approchée de  $sg$  qui annule la fonction  $F_g$ .

Nous pouvons procéder exactement de la même manière sur le pool inhibiteur d'une autre région  $j$ , pour obtenir  $sg_{B_j}^*$  à partir de  $z_{I_{B_j}}^*$ . Nous travaillons alors sur chaque pool inhibiteur, les uns après les autres (cf figure 10.12).

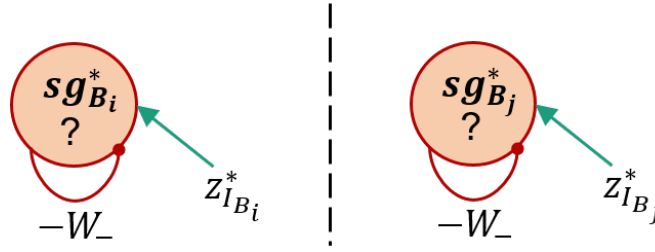


**Figure 10.12** – Inversion des activités  $sg_B^*$  des pools inhibiteurs de chaque région, pour des valeurs fixées des forçages, hors période de stimulation.

#### 10.6.4.1.3 2ème étape : travail sur le pool exciteur pour inverser $J_{gaba_i}$

Nous considérons ensuite le pool exciteur de la région  $i$  au régime stationnaire, au niveau de base i.e. hors période de stimulation (modèle (10.119)). Nous connaissons le point fixe  $sn_{B_i}^*$ , nous allons chercher la valeur du forçage  $z_{E_{B_i}}^*$  correspondant (c'est-à-dire qui permet d'obtenir cette valeur du point fixe).

##### a) Calcul des valeurs des firing rate hors stimulation $rn_{B_i}$

A partir des  $sn_{B_i}^*$  connus, nous pouvons calculer les valeurs de fréquences de potentiels d'action hors stimulation pour chaque région :

$$0 = f_{n_i}(sn_{B_i}^*, rn_{B_i}^*) = -\beta^E sn_{B_i}^* + \alpha^E T_{glu} rn_{B_i}^* (1 - sn_{B_i}^*) \iff rn_{B_i}^* = \frac{\beta^E}{\alpha^E T_{glu}} \frac{sn_{B_i}^*}{(1 - sn_{B_i}^*)} \quad (10.122)$$

**b) Calcul des valeurs de  $xn_{B_i}^*$  qui donnent  $rn_{B_i}^*$**

La fonction  $h_{n_i}$  est une fonction croissante de  $xn_i$  (cf annexe 17.2.2) et donc en particulier de  $xn_{B_i}^*$ . Nous pouvons résoudre numériquement l'équation  $rn_{B_i}^* = h_{n_i}(xn_{B_i}^*)$ , par dichotomie par exemple.

**c) Calcul des valeurs de  $z_{E_{B_i}}^*$  qui donnent  $xn_{B_i}^*$**

Nous avons

$$xn_{B_i}^* = W_+ J_{nmda} sn_{B_i}^* + z_{E_{B_i}}^* \iff z_{E_{B_i}}^* = xn_{B_i}^* - W_+ J_{nmda} sn_{B_i}^* \quad (10.123)$$

**d) Inversion de  $J_{gaba_i}$  à partir de  $z_{E_{B_i}}^*$**

Nous nous servons maintenant de  $z_{E_{B_i}}^*$  pour pouvoir inverser  $J_{gaba_i}$ . En effet, nous avons défini :

$$z_{E_{B_i}}^* = -J_{gaba_i} sg_{B_i}^* + B_{E_i} + \sum_{j \neq i} k_{E_{ij}} \kappa_{ij} sn_{B_j}^* \quad (10.124)$$

Donc on peut en déduire

$$J_{gaba_i} = -(z_{E_{B_i}}^* - B_{E_i} - \sum_{j \neq i} k_{E_{ij}} \kappa_{ij} sn_{B_j}^*) \frac{1}{sg_{B_i}^*} \quad (10.125)$$

Nous pouvons procéder exactement de la même manière sur le pool exciteur d'une autre région  $j$ , pour obtenir  $J_{gaba_j}$  à partir de  $z_{E_{B_j}}^*$ . Nous travaillons alors sur chaque pool exciteur, les uns après les autres (cf figure 10.13).

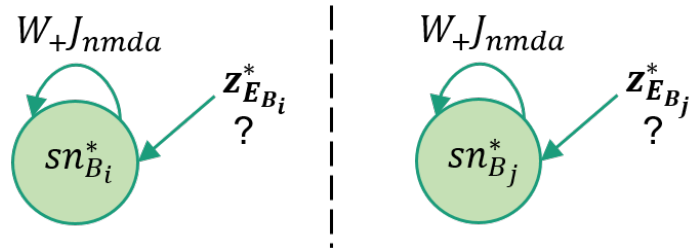


**Figure 10.13** – Inversion des forçages  $z_{E_B}^*$  appliqués sur les pools excitateurs de chaque région, pour des autocouplages excitateurs fixés, hors période de stimulation.

#### 10.6.4.2 Inversion des $I_{stim}$ en période de stimulation

##### 10.6.4.2.1 Modèle en période de stimulation

Nous notons de manière générale  $\eta_{stim_i}^*$  ou  $\eta_{stim_j}^*$  toute variable  $\eta$  décrivant la région  $i$  ou la région  $j$  en période de stimulation.

Nous pouvons écrire le modèle en période de stimulation pour les pools excitateur et inhibiteur de la région  $i$ , en adaptant (10.117) et (10.118) :

Pour le pool excitateur :

$$\begin{cases} 0 = f_{n_i}(sn_{stim_i}^*, rn_{stim_i}^*) \\ rn_{stim_i}^* = h_{n_i}(xn_{stim_i}^*) \\ xn_{stim_i}^* = w_{n_i}(sn_{stim_i}^*, z_{E_{stim_i}}^*) \\ z_{E_{stim_i}}^* = p_{n_i}(sg_{stim_i}^*, sn_{stim_j}^*, J_{gaba_i}, B_{E_i} + I_{stim_i}) \end{cases} \quad (10.126)$$

pour le pool inhibiteur :

$$\begin{cases} 0 = f_{g_i}(sg_{stim_i}^*, rg_{stim_i}^*) \\ rg_{stim_i}^* = h_{g_i}(xg_{stim_i}^*) \\ xg_{stim_i}^* = w_{g_i}(sg_{stim_i}^*, z_{I_{stim_i}}^*) \\ z_{I_{stim_i}}^* = p_{g_i}(sn_{stim_i}^*, sn_{stim_j}^*) \end{cases} \quad (10.127)$$

#### 10.6.4.2.2 1ère étape : travail sur le pool inhibiteur pour trouver $sg_{stim_i}^*$

Comme au paragraphe 10.6.4.1.2, nous travaillons d'abord sur le pool inhibiteur de la région  $i$  au régime stationnaire, en période de stimulation (modèle (10.127)).

Pour chaque région,  $sn_{stim_i}^*$  et  $sn_{stim_j}^*$  sont donnés par la première étape de l'inversion du BOLD, ce qui fixe :

$$z_{I_{stim_i}}^* = p_{g_i}(sn_{stim_i}^*, sn_{stim_j}^*) = J_{nmda} sn_{stim_i}^* + B_{I_i} + \sum_{j \neq i} (1 - k_{E_{ij}}) \kappa_{ij} sn_{stim_j}^* \quad (10.128)$$

Nous cherchons numériquement le point fixe  $sg_{stim_i}^*$ , avec la valeur de l'autocouplage et la valeur du forçage.

Comme au paragraphe 10.6.4.1.2, deux méthodes sont possibles :

1. faire tourner numériquement la dynamique, pour n'importe quelle valeur de condition initiale  $sg(t=0) \in [0, 1]$  ;
2. écrire le modèle du pool inhibiteur de la région  $i$  au régime stationnaire sous la forme  $0 = F_g(sg_{stim_i}^*, z_{I_{stim_i}}^*)$  et faire une recherche de 0 de la fonction  $F_g$  par dichotomie. Dans les deux cas, il est nécessaire de fixer la valeur du forçage  $z_{I_{stim_i}}^*$  en utilisant l'équation 10.128.

Nous pouvons procéder exactement de la même manière sur le pool inhibiteur d'une autre région  $j$ , pour obtenir  $sg_{stim_j}^*$  à partir de  $z_{I_{stim_j}}^*$ . Nous travaillons alors sur chaque pool inhibiteur, les uns après les autres (cf figure 10.14).

#### 10.6.4.2.3 2ème étape : travail sur le pool excitateur pour inverser $I_{stim_i}$

Nous considérons ensuite le pool excitateur de la région  $i$  au régime stationnaire, en période de stimulation (modèle (10.119)). Nous connaissons le point fixe  $sn_{stim_i}^*$ , nous allons chercher

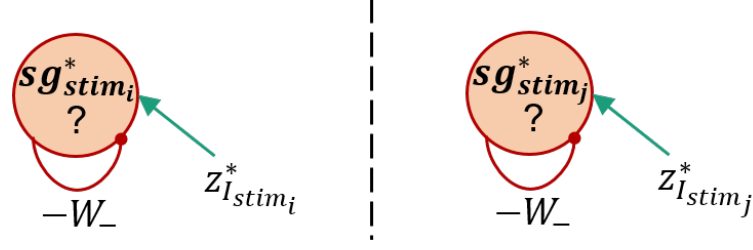


**Figure 10.14** – Inversion des activités  $sg_{stim}^*$  des pools inhibiteurs de chaque région, pour des valeurs fixées des forçages, en période de stimulation.

la valeur du forçage  $z_{E_{stim_i}}^*$  correspondant (c'est-à-dire qui permet d'obtenir cette valeur du point fixe).

**a) Calcul des valeurs des firing rate en période de stimulation  $rn_{stim_i}$**

A partir des  $sn_{stim_i}^*$  connus, nous pouvons calculer les valeurs de fréquences de potentiels d'action en période de stimulation pour chaque région :

$$0 = f_{n_i}(sn_{stim_i}^*, rn_{stim_i}^*) = -\beta^E sn_{stim_i}^* + \alpha^E T_{glu} rn_{stim_i}^* (1 - sn_{stim_i}^*) \iff rn_{stim_i}^* = \frac{\beta^E}{\alpha^E T_{glu}} \frac{sn_{stim_i}^*}{(1 - sn_{stim_i}^*)} \quad (10.129)$$

**b) Calcul des valeurs de  $xn_{stim_i}^*$  qui donnent  $rn_{stim_i}^*$**

La fonction  $h_{n_i}$  est une fonction croissante de  $xn_i$  (cf annexe 17.2.2) et donc en particulier de  $xn_{stim_i}^*$ . Nous pouvons résoudre numériquement l'équation  $rn_{stim_i}^* = h_{n_i}(xn_{stim_i}^*)$ , par dichotomie par exemple.

**c) Calcul des valeurs de  $z_{E_{stim_i}}^*$  qui donnent  $xn_{stim_i}^*$**

$xn_{stim_i}^*$  est égal à :

$$xn_{stim_i}^* = W_+ J_{nmda} sn_{stim_i}^* + z_{E_{stim_i}}^* \iff z_{E_{stim_i}}^* = xn_{stim_i}^* - W_+ J_{nmda} sn_{stim_i}^* \quad (10.130)$$

**d) Inversion de  $I_{stim_i}$  à partir de  $z_{E_{stim_i}}^*$**

Nous pouvons maintenant nous servir de  $z_{E_{stim_i}}^*$  pour pouvoir inverser  $I_{stim_i}$ . En effet, nous avons défini :

$$z_{E_{stim_i}}^* = -J_{gaba_i} sg_{stim_i}^* + B_{E_i} + I_{stim_i} + \sum_{j \neq i} k_{E_{ij}} \kappa_{ij} sn_{stim_j}^* \quad (10.131)$$

Donc on peut en déduire

$$I_{stim_i} = z_{E_{stim_i}}^* + J_{gaba_i} sg_{stim_i}^* - B_{E_i} - \sum_{j \neq i} k_{E_{ij}} \kappa_{ij} sn_{stim_j}^* \quad (10.132)$$

Nous devons utiliser les valeurs de  $J_{gaba_i}$  inversées précédemment (i.e. en se servant des valeurs de  $sn_{B_i}^*$  hors stimulation) pour inverser  $I_{stim_i}$  en période de stimulation.

Nous pouvons procéder exactement de la même manière sur le pool excitateur d'une autre région  $j$ , pour obtenir  $I_{stim_j}$  à partir de  $z_{E_{stim_j}}^*$ . Nous travaillons alors sur chaque pool excitateur, les uns après les autres (cf figure 10.15).

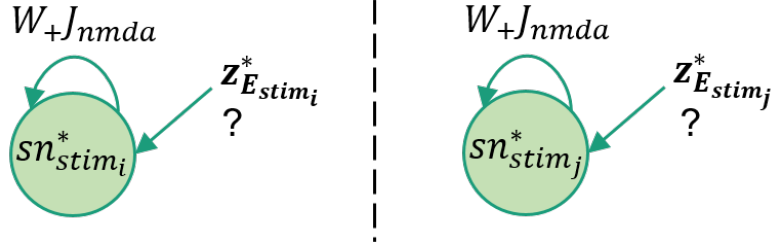
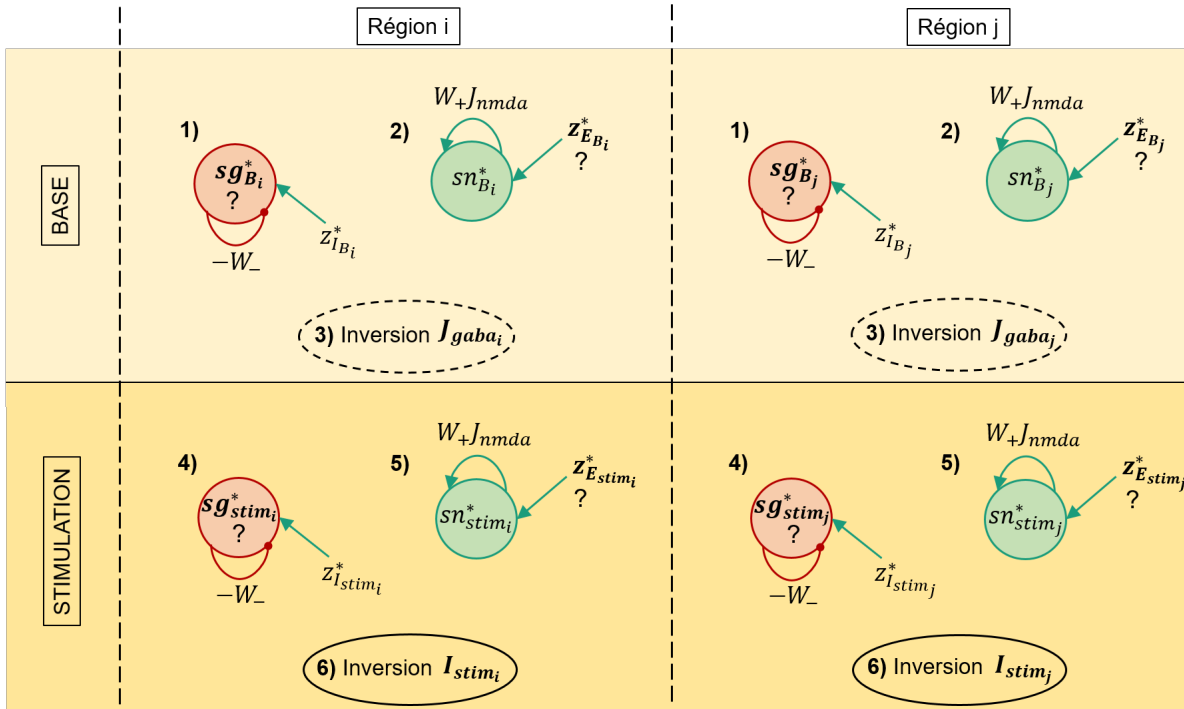


**Figure 10.15** – Inversion des forçages  $z_{E_{stim}}^*$  appliqués sur les pools excitateurs de chaque région, pour des autocouplages excitateurs fixés, en période de stimulation.

### 10.6.5 Schéma récapitulatif de la deuxième étape



La méthode est découpée par pool de régions : un travail sur chaque pool de chaque région respectivement au niveau de base et en période de stimulation permet d'inverser un  $J_{gaba}$  puis un  $I_{stim}$  par région.

Comme ici la formulation de l'inversion est algébrique, nous nous sommes assurés de son implémentation sur un jeu unique de points fixes  $sn^*$ .

# Chapitre 11

## Preuve de concept : Inversion des données d'un participant

*Je présente les résultats d'inversion obtenues à partir de la mise en oeuvre de la procédure d'inversion présentée au chapitre 11, en utilisant les données d'IRMf d'un participant pendant une tâche de contrôle inhibiteur des intrusions mnésiques (paradigme Think/No Think). Ce participant a été choisi dans le groupe de sujets sains non exposés de la cohorte REMEMBER.*

*Dans la continuité des études antérieures menées dans mon laboratoire d'accueil [Mary et al., 2020], le système choisi est composé de 15 régions : 11 régions de contrôle ('SFG2', 'SFG12', 'SFG4', 'MFG16', 'MFG20', 'MFG22', 'MFG24', 'IFG30', 'IFG34', 'IFG38', 'CG178'), et 4 régions cibles ('Hipp216', 'Hipp218', 'PhG112', 'PCun152').*

Rappelons que dans la tâche TNT, nous avons 3 conditions qualifiées en temps de stimulation : la condition Think, la condition NoThink avec succès du contrôle inhibiteur (pas d'intrusion), et la condition NoThink où le contrôle inhibiteur n'a pas joué son rôle (une intrusion est rapportée par le sujet). Ces trois conditions sont codées dans les données en qualifiant les portes correspondantes. Ces trois conditions viennent en plus de la condition hors stimulation (Base).

Nous faisons porter les deux étapes d'inversion en considérant des combinaisons de conditions :

- nous contrastons la condition NoThink où le contrôle inhibiteur doit être activé (avec ou sans intrusion) avec la condition de base ou Think (les portes Think ne sont pas activées dans le modèle direct),
- nous contrastons la condition NoThink, en activant les portes Think, ce qui fait une inversion sur trois paramètres,
- nous contrastons les quatre conditions, ce qui fait une inversion sur quatre paramètres.



## 11.1 Première étape de l'inversion : du BOLD à $sn^*$

Nous présentons les résultats d'inversion obtenus sans avoir inversé la condition initiale<sup>1</sup>. Les densités a priori utilisées pour les paramètres hémodynamiques et l'inverse de la variance du bruit sont celles qui ont été fixées à la section 10.4. Pour le bruit, nous fixons ainsi un écart-type du bruit a priori du même ordre de grandeur que l'amplitude maximale du signal non bruité prédit sous l'hypothèse du modèle<sup>2</sup>.

L'inversion des séries temporelles de 15 régions, sur 940 TR est très rapide. Elle dure respectivement 2, 3 ou 4 minutes pour 2, 3 ou 4 paramètres par région<sup>3</sup>.

La figure 11.1 présente l'ensemble des résultats de cette première étape d'inversion pour les trois combinaisons de conditions.

La figure 11.1(A) montre le contraste entre la condition NoThink (activation du contrôle inhibiteur) par rapport aux deux autres conditions (base et Think). Si on peut observer que les  $sn^*$  inversés pour les conditions base+Think sont à peu près homogènes (autour de 0.18), la réaction à la présentation du stimulus requérant le déclenchement du contrôle inhibiteur présente au contraire une tendance remarquable à se conformer à l'attendu lié à ce déclenchement : toutes les régions de contrôle (sauf RC 11) voient leur niveau d'activité augmenter, alors que les quatre régions-cibles voient leur niveau d'activité diminuer, ce qui correspond parfaitement au scénario exploré tout au long de notre travail, en tous cas chez ce participant.

En différenciant, figure 11.1(B), les conditions non déclenchantes du contrôle inhibiteur (Base et Think), l'image reste cohérente, avec des valeurs associées au Think qui sont toutes inférieures aux valeurs associées au NoThink dans les régions de contrôle, et même des valeurs qui peuvent devenir sensiblement inférieures au niveau de base pour certaines d'entre elles (RC 2-4, 7-9), ce qui pointerait vers l'idée d'une inhibition de certaines régions de contrôle en condition Think. Les valeurs inversées pour les régions-cibles en condition Think, quant à elles, diffèrent très peu de la condition Base (sauf peut-être C 13), là encore en conformité avec le scénario à l'étude.

Nous notons que la séparation des conditions non déclenchantes (l'activation des portes Think pour inverser SnThink) a un effet de bord sur l'estimation des valeurs inversées pour la condition déclenchante (les valeurs de  $sn^*$  obtenues pour la condition NoThink ne sont pas les mêmes, voir par exemple la RC 4), même si la tendance décrite ci-dessus pour la première combinaison reste globalement cohérente. Cette variation méritera d'être analysée plus en profondeur, mais, à ce stade, nous l'attribuons à l'effet d'auto-corrélation temporelle du modèle BOLD. En effet, entre les deux combinaisons (amalgame Base+Think, ou séparation), le pilotage de la prédiction du signal BOLD donné par l'expression 10.115 est donné, pour chaque exécution du modèle direct dans l'itération bayésienne, soit par une valeur unique contrainte par  $sn_{T_i} = sn_{B_i}$  pour la première combinaison, soit se différencie

---

1. nous avons vérifié que les résultats étaient identiques avec ou sans inversion de la condition initiale sur des simulations au chapitre précédent et sur les données du participant

2. Avec les hyperparamètres a priori qui se trouvent au paragraphe 10.4.2.1.2, l'espérance de la loi a priori que suit  $\sigma^{va}$  vaut  $10^4$ .  $\sigma^{va}$  est l'inverse de la variance du bruit. Donc l'espérance de la variance du bruit vaut  $10^{-4}$ , ce qui correspond à un écart-type de  $10^{-2}$ . Nous avons vu au chapitre 7, que sous l'hypothèse de notre modèle, l'amplitude maximale du signal BOLD d'une région vaut 0.04.

3. sans inversion de la condition initiale, une minute supplémentaire si nous inversons la condition initiale.

en deux prédictions selon que la porte activée soit la Base ou la porte Think, en fonction de l'évolution de leur distributions *a posteriori* respectives. La vraisemblance de la prédiction globale sur le signal BOLD peut en être affectée, ce qui, en retour, peut faire diverger la trajectoire de tuning sur les portes NoThink.

En différenciant, figure 11.1(C), les valeurs inversées en conditions NoThink avec ou sans intrusion signalée par le participant, le résultat est un peu flou pour les régions-cibles : pour deux d'entre elles (C 13 et 14), on constate un même niveau d'inhibition, même si pour C 12 et 15, on constate une possible association entre la présence d'une intrusion et un possible défaut d'inhibition par rapport à la condition sans intrusion (une moindre inhibition pourrait être associée à la présence d'une intrusion).

Ces interprétations d'une seule inversion chez un participant sont, bien entendu, à prendre avec la précaution requise...

Il reste que l'observation de la baisse d'activité dans les régions-cibles sous l'action de l'augmentation d'activité dans les régions de contrôle vient conforter la nécessité de choisir un schéma de connectivité qui la permette.

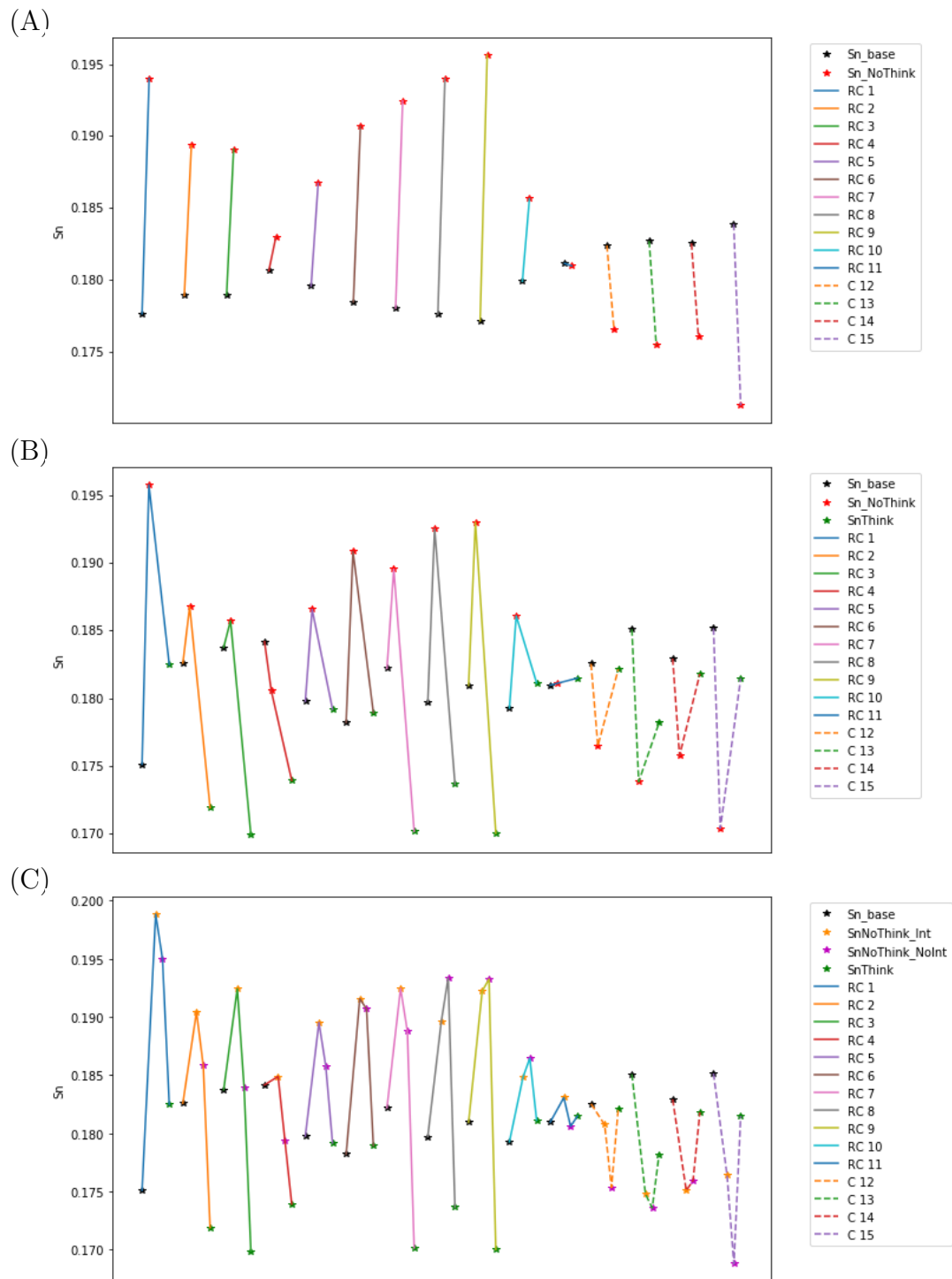


**Figure 11.1** – Résultat de la première étape d'inversion chez un participant, sur un sous-ensemble de 15 régions, et pour les trois combinaisons de conditions à contraster. Les 11 colonnes de gauche rapportent les valeurs de  $sn^*$  inversées pour les régions de contrôle (RC 1 à RC 11), et les quatre colonnes de droites celles pour les régions cibles (C 12 à C 15).

## 11.2 Deuxième étape

Pour la seconde étape de l'inversion, nous avons besoin de prendre en compte la connectivité entre régions. En cohérence avec notre question générale, nous avons considéré le schéma E-I tel qu'exposé dans l'article du chapitre 9, mais ici à plusieurs régions : nous considérons que les régions de contrôle sont connectées en excitations mutuelles, que les régions-cibles sont connectées également en excitations mutuelles, et que le schéma E-I s'applique entre ces deux groupes, chaque région de contrôle projetant ses connexions excitatrices de longue-portée vers les pools inhibiteurs des régions-cibles, alors que celles-ci projettent, en retour, leur connexions excitatrices de longue-portée vers les pools excitateurs des régions de contrôle. Cette architecture de connectivité a été présentée au chapitre 10, section 10.1.2.

### 11.2.1 $J_{gaba}$ par région

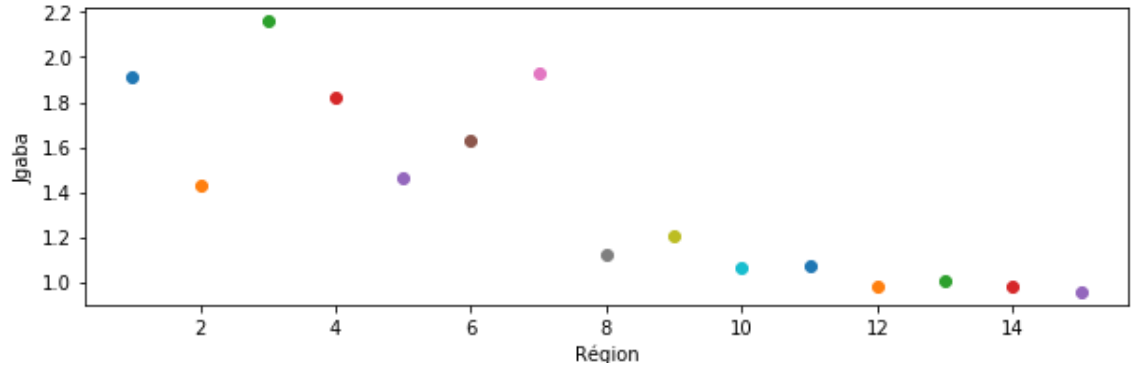


**Figure 11.2** – Résultat de la seconde étape d'inversion chez un participant pour la valeur de  $J_{gaba}$ . Les numéros de région correspondent à celle de la figure 11.1.

La figure 11.2 reporte les valeurs de  $J_{gaba}$  inversées pour chaque région à partir de la valeur de base de  $sn^*$  (nous reportons seulement l'inversion pour la première combinaison, les deux autres sont superposables).

Nous pouvons observer une tendance générale à des valeurs supérieures de  $J_{gaba}$  dans les régions de contrôle C 1-7 (autour de 1.8 nA), et une uniformité de celles-ci pour les régions-cibles (autour de 1 nA). Au regard de la cohérence d'ensemble du travail, il est intéressant de noter que cette distribution est assez conforme avec les valeurs de  $J_{gaba}$  inversées pour une situation de repos à 3 Hz, telles que présentées dans l'article pour la connectivité E-I, avec une valeur de référence  $J_{gaba,C} = 1.758$  nA dans les régions de contrôle, et  $J_{gaba,T} = 0.929$  nA dans les régions-cibles.

### 11.2.2 $I_{stim}$ par région

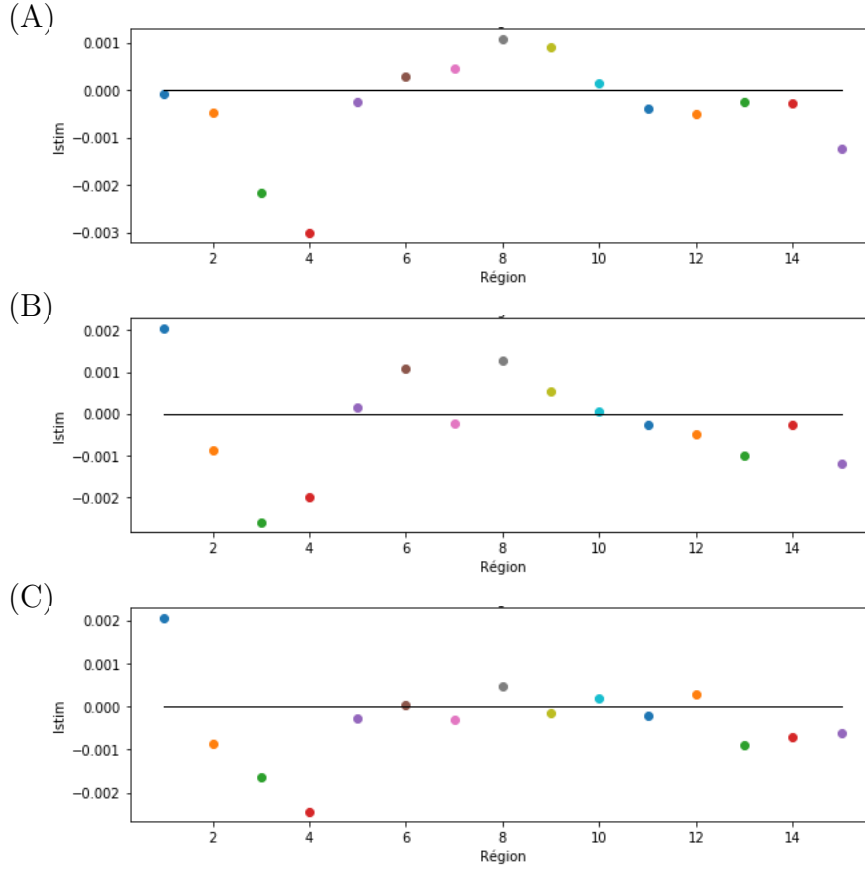


**Figure 11.3** – Résultat de la seconde étape d’inversion chez un participant pour la valeur de  $I_{stim}$  pour les trois combinaisons de conditions.

La figure 11.3 reporte les valeurs de  $I_{stim}$  inversées pour chaque région, en tenant compte des valeurs inversées en condition de stimulation, telles qu’elles doivent être spécifiées dans l’expression 10.132. Que ce soit en (A) où les valeurs stimulées amalgament la condition NoThink (avec ou sans intrusion) inversées en confondant Base et Think (première combinaison), en (B) où les valeurs stimulées amalgament la condition NoThink (avec ou sans intrusion) inversées en séparant Base et Think (seconde combinaison), ou en (C) où la valeur stimulée prise en compte est seulement la valeur inversée pour la condition NoThink avec intrusion, nous ne pouvons que constater que les valeurs inversées pour  $I_{stim}$  ne peuvent être interprétées autrement que comme du bruit autour de la valeur 0.

Ce constat est bien sûr en désaccord avec la prédiction qui aurait voulu que les régions de contrôle montrent un  $I_{stim}$  positif, avec des régions cibles montrant une valeur de stimulation externe nulle, sous l’hypothèse qu’on ait bien identifié que leurs entrées inhibitrices soient principalement celles venant des régions de contrôles présentes dans le choix du sous-système à 15 régions.

Nous notons, à ce stade de l’analyse de cette sortie de première inversion, que cet "échec" de l’inversion de  $I_{stim}$  nous amène à examiner en retour de plus près les valeurs obtenues à

l'inversion de première étape. Si, en effet, les variations de valeurs inversées alors pour  $sn^*$  vont bien dans le sens d'un schéma de contrôle inhibiteur de type E-I, les valeurs absolues des valeurs inversées se présentent dans un intervalle qui apparaît peu compatible avec les images d'activation / inhibition des niveaux d'activité, puisque la variation de  $sn^*$  dans un intervalle entre 0.17 et 0.2 comme en figure 11.1 est beaucoup plus faible que les variations considérées dans l'étude analytique.

Nous voyons donc à ce stade réapparaître la question de l'interprétation des valeurs absolues de  $sn^*$  obtenues par la première étape d'inversion. Sur une possible mise à l'échelle des résultats, pour correspondre aux gammes d'intervalles du modèle, il sera nécessaire d'étudier de manière plus approfondie l'effet du bruit (temporellement auto-corrélé) du signal BOLD, et la propagation, dans la chaîne d'inversion, des incertitudes attachées à chaque paramètre estimé.



# Cinquième partie

## Discussion



*Cette partie est constituée de trois chapitres. Le chapitre 12 est une synthèse des méthodes et principaux résultats. Le chapitre 13 revisite des concepts majeurs de ma thèse (la balance  $E/I$ , le contrôle inhibiteur, la sur-inhibition) en questionnant leur énoncé en langage naturel et leur traduction en langage mathématique. Il s'agit d'un compte-rendu de mon questionnement personnel, en tant que physicienne plongée dans une dynamique de recherche interdisciplinaire. Je le partage ici comme guide de lecture ou de relecture de ma thèse, parce que le travailler m'a énormément et précieusement éclairée, parce qu'il alimente ma pratique de la modélisation à l'interface de la physique du non linéaire et des neurosciences, et parce qu'il me permet d'élaborer sur de nouvelles questions, en particulier sur le contrôle exciteur et la moindre excitation au chapitre 14. Ce dernier chapitre regroupe plus largement des perspectives de mon travail, en modélisation et en inversion, pour alimenter de futures thèses.*

# Chapitre 12

## Synthèse

*Je vais rappeler les objectifs de la thèse et les principaux résultats.*

*Les sections 12.2.1, 12.2.2, 12.2.3 et 12.2.4 reprennent les points principaux du travail de modélisation et d'analyse présenté dans l'article du chapitre 9. La section 12.2.5 reprend les étapes de la procédure d'inversion présentée au chapitre 10, qui a été mise en oeuvre sur les données d'un participant au chapitre 11.*

*Je souligne en 12.3 les particularités des méthodes de travail que nous avons employées. La dynamique collective interdisciplinaire a notamment permis de mettre en oeuvre en neurosciences des méthodes d'analyses des systèmes non linéaires qui ont été éprouvées par les sciences du climat depuis les années 1980 (méthode dite de "découpage/raccordement").*

### 12.1 Rappel des objectifs de la thèse

Les objectifs de la thèse, présentés au chapitre 7, portaient sur :

- A) Comment connecter les régions au niveau neural pour obtenir une baisse de l'activité neurale excitatrice au régime stationnaire dans la région cible (i.e.  $\Delta sn_c^* = sn_{stim_c}^* - sn_{B_c}^* < 0$ ), quand l'activité neurale excitatrice au régime stationnaire dans la région de contrôle augmente (i.e.  $\Delta sn_{rc}^* = sn_{stim_{rc}}^* - sn_{B_{rc}}^* > 0$ ) ?
- B) Pour une architecture de connectivité fixée, comment  $I_{stim}$  dans les régions de contrôle et  $J_{gaba_c}$  dans les régions cibles influencent-ils les variations des activités neurales excitatrices au régime stationnaire, i.e.  $\Delta sn_i^*$ , des régions impliquées dans une tâche de contrôle inhibiteur ; en particulier la baisse d'activité des régions cibles  $\Delta sn_c^*$  due à l'augmentation d'activité des régions de contrôle  $\Delta sn_{rc}^*$  ?
- C) Comment l'analyse du modèle direct éclaire-t-elle le problème de l'inversion et nous permet de construire une procédure d'inversion (en 2 étapes) pour inférer les paramètres neuraux  $I_{stim}$  et  $J_{gaba}$  de chaque région à partir de signaux BOLD recueillis pendant une tâche de contrôle inhibiteur (quel choix de modèle direct pertinent, sur quelle observable porte l'inversion, comment construire une procédure d'inversion etc.) ?
- D) En prenant comme cas d'étude le contrôle inhibiteur des intrusions mnésiques, est-il possible de présenter une preuve de concept de l'inversion en estimant les paramètres  $I_{stim}$  et  $J_{gaba}$  à partir des mesures BOLD d'un sujet (un participant non exposé de l'étude REMEMBER) réalisant une tâche Think/No Think ?

## 12.2 Principaux résultats

### 12.2.1 Un nouveau modèle de connectivité

Nous proposons à la communauté un nouveau modèle de connectivité inter-régions qui généralise celui de [Naskar et al., 2021].

Les connexions inter-régions, de nature excitatrice, partent toujours des pools excitateurs. Une région peut émettre un faisceau de plusieurs fibres de son pool E vers le pool E et/ou I d'une autre région ; et recevoir de celle-ci de manière indépendante par son pool E et/ou I. Nous faisons l'hypothèse que chaque fibre de ce faisceau ne peut se projeter que sur l'un des deux pools de l'autre région, et pas les deux à la fois. Nous introduisons la proportion du nombre de fibres qui lie le pool E de la région  $j$  au pool E de la région  $i$ , par rapport au nombre total de fibres entre ces deux régions, noté  $k_{Eij}$ . La proportion du nombre de fibres qui lie le pool E de la région  $j$  au pool I de la région  $i$  vaut alors  $1 - k_{Eij}$ . Cette généralisation permet d'envisager différentes architectures de connectivité, dont celle proposée par [Naskar et al., 2021].

### 12.2.2 Structure de connectivité pertinente pour du contrôle inhibiteur

L'analyse du comportement de ce modèle avec la mise en oeuvre des outils d'analyses des systèmes non linéaires (régime stationnaire, temps caractéristiques, étude des sensibilités, gains de rétroaction, analyse par emboîtement) amène à proposer une structure de connectivité différente de celle du modèle de départ de [Naskar et al., 2021] pour pouvoir rendre compte des mécanismes d'inhibition.

Pour deux régions, une connexion du pool E de la région de contrôle vers le pool I de la région cible permet d'obtenir au niveau neural et BOLD une baisse de l'activité de la cible suite à une augmentation de l'activité de la région de contrôle (indépendamment du retour des cibles vers les régions de contrôle de E vers E ou de E vers I). Un retour de la cible vers la région de contrôle de E vers E permet d'obtenir des variations douces et non abruptes de l'activité de la cible quand on augmente la stimulation<sup>1</sup> de la région de contrôle. Cette structure de connectivité est notée E-I<sup>2</sup>. Au contraire, avec des connexions de type E-E envisagées dans le modèle d'origine ([Naskar et al., 2021]), surtout développé pour rendre compte de données en resting state au départ, l'activité des régions cibles ne peut qu'augmenter.

### 12.2.3 Effets de $I_{stim}$ sur la baisse d'activité $sn_c^*$

La stimulation de la région de contrôle se traduit par une augmentation du forçage sur son pool exciteur d'une valeur de  $I_{stim}$ . Plus  $I_{stim}$  augmente et plus la variation d'activité de la cible au régime stationnaire est grande en valeur absolue. Dans le cas d'une connectivité EI, plus  $I_{stim}$  augmente et plus l'activité de la cible  $sn_c^*$  baisse.

---

1. i.e. le forçage sur le pool exciteur

2. E : par quel pool reçoit la région de contrôle, I : par quel pool reçoit la cible

### 12.2.4 Effets de $J_{gaba_c}$ dans la cible sur la baisse d'activité $sn_c^*$

$J_{gaba_c}$  représente le poids synaptique des interneurons inhibiteurs vers les neurones excitateurs de la cible. Nous mettons en évidence une relation non linéaire (en forme de U pour une connectivité EI) entre la baisse d'activité de la région cible (pour  $I_{stim}$  fixé) et le paramètre inhibiteur  $J_{gaba_c}$  de celle-ci. Il existe une valeur de  $J_{gaba_c}$  au delà de laquelle plus on augmente l'inhibition dans la cible (i.e. le poids synaptique inhibiteur dans la cible  $J_{gaba_c}$ ), moins on abaisse l'activité de la cible. Ce résultat contre-intuitif nous semble important à prendre en considération dans la perspective de donner du sens au rôle des paramètres du modèle.

### 12.2.5 Procédure d'inversion des paramètres neuraux à partir de mesures BOLD

La comparaison des temps d'installation du modèle neural et du modèle BOLD permet de construire une procédure d'inversion en deux étapes :

1. Inversion des signaux BOLD  $y(t)$  vers les points fixes neuraux correspondant aux états non stimulés  $sn_B^*$  et stimulés  $sn_{stim}^*$ .  
Nous passons par une méthode d'inversion bayésienne variationnelle et l'implémentation numérique de cette étape passe par l'adaptation des codes de la toolbox VBA (matlab).
2. Inversion des paramètres neuraux  $I_{stim}$  et  $J_{gaba}$  à partir de ces jeux de points fixes neuraux, par résolution algébrique et numérique (python).

## 12.3 Méthode de travail

### 12.3.1 Séparation de la modélisation et de la technique/méthode d'inversion

Pour permettre l'analyse de données mesurées en IRMf pendant une tâche de contrôle inhibiteur (en particulier des intrusions mnésiques), nous avons séparé les questions relatives au modèle direct de celles qui concernent la technique d'inversion.

Une première partie de notre travail a consisté à éclairer les propriétés d'un modèle hiérarchique, couplé, non linéaire du signal BOLD de plusieurs régions cérébrales et des activités neurales sous-jacentes. Ceci a surtout été fait pour 2 régions et nous a permis de :

1. identifier un jeu de paramètres pertinents ( $\overline{k_E}$  qui fixe une architecture de connectivité) permettant d'observer en période de stimulation une augmentation de l'activité des régions de contrôle et une baisse de l'activité des régions cibles<sup>3</sup>

---

3. Etude en sensibilités

2. proposer deux étapes pour la méthode d'inversion (des signaux BOLD  $y(t)$  aux activités neurales  $sn^*$  de toutes les régions d'intérêt, puis des  $sn^*$  aux paramètres neuraux  $I_{stim}$  et  $J_{gaba}$  de toutes les régions d'intérêt) <sup>4</sup>
3. mieux comprendre les effets des paramètres qu'on inverse sur le signal BOLD, dans l'objectif de pouvoir en proposer une interprétation en sortie d'inversion.

La seconde partie de notre travail a eu pour but de proposer une procédure d'inversion robuste, avec des durées de calculs raisonnables pour pouvoir analyser les résultats des participants de la cohorte REMEMBER. Nous avons fixé l'architecture de connectivité pertinente grâce à la première partie. Le séquençage en deux étapes rend possible l'inversion avec le modèle de connectivité pertinent et les durées d'inversion sont courtes. La première étape ne prend que quelques minutes pour 15 régions et la seconde est quasi-instantanée du fait de la résolution algébrique.

### 12.3.2 Méthode d'analyse du modèle direct

Nous nous sommes intéressés au système composé de deux régions couplées : une région de contrôle et une région cible. Les variations respectives de leur activité neurale ( $sn_{rc}^*$  et  $sn_c^*$ ) et BOLD ( $y_{rc}(t)$  et  $y_c(t)$ ) sont décrites sous l'hypothèse du modèle présenté au chapitre 9.

Pour mieux comprendre comment l'architecture de connectivité impacte le signe des variations du signal BOLD de chaque région lorsque nous appliquons une stimulation de la région de contrôle (modélisée par une augmentation du forçage sur son pool exciteur), nous avons construit :

1. un passage d'échelle entre le fonctionnement d'un pool exciteur découplé <sup>5</sup> et celui d'un pool inhibiteur découplé dans une région cérébrale, jusqu'à la réponse neurale  $sn_i^*$  dans le système couplé.

Ce passage d'échelle repose sur la mise en oeuvre d'une méthode particulière d'analyse, dite par découpage/raccordement, utilisée notamment par des physiciens de la communauté climatique. Elle constitue l'un des apports majeurs issus de la collaboration scientifique qui s'est déroulée pendant les deux dernières années de cette thèse. La méthode par découpage/raccordement nous a permis de construire un emboîtement de sensibilités en exprimant pour chaque niveau d'emboîtement les sensibilités couplées en fonction de sensibilités découplées calculées au niveau d'emboîtement inférieur <sup>6</sup>.

Le système de deux régions couplées constitue le niveau d'emboîtement le plus haut. Tout autre niveau d'emboîtement est une sous-partie (ou un ensemble de sous-parties disjointes) du niveau d'emboîtement supérieur, dans laquelle nous connaissons les sensibilités des variables d'état ( $sn^*$  ou  $sg^*$ ) aux différents forçages, que nous pouvons évaluer pour n'importe quelles valeurs de ces forçages, et notamment pour des valeurs

---

4. Etude des temps caractéristiques

5. sans aucun autocouplage, seulement avec un forçage

6. i.e. sensibilités des variables d'état ( $sn^*$  ou  $sg^*$ ) aux différents forçages dans le sous-système considéré, évaluées à des valeurs particulières de ces forçages qui s'expriment en fonction des points fixes dans le système couplé

particulières et pertinentes qui s'expriment en fonction des points fixes obtenus dans le niveau d'emboîtement le plus haut (i.e. deux régions couplées).

Nous avons défini 4 niveaux d'emboîtement (dans un ordre descendant) :

- deux régions couplées
- deux régions découplées entre elles, constituées de leurs pools excitateur et inhibiteur couplés entre eux
- dans chaque région : deux pools excitateur et inhibiteur, découplés entre eux, chacun conservant leur autocouplage.
- chaque pool sans autocouplage, mais seulement avec un forçage

2. un passage des  $sn_i^*$  au niveau de description BOLD  $y_i(t)$



# Chapitre 13

## Préciser le langage pour partager le questionnement

*Dans une thèse disciplinaire, il serait attendu que la discussion porte sur des conclusions au regard de la question posée et la manière dont elles s'inscrivent dans les travaux de la communauté concernée. Cependant, la thèse présentée dans ce manuscrit se situe à l'interface de plusieurs disciplines : neurosciences expérimentales (neuropsychologie, neurobiologie, neuroimagerie), neurosciences computationnelles et physique du non linéaire.*

*Physicienne de formation, j'ai mené ce travail dans un laboratoire de neuropsychologie où j'ai été entourée de personnes expertes en neurosciences expérimentales (neuropsychologie et neuroimagerie) avec une culture scientifique, des lignes de raisonnements et un langage différents de ceux de ma communauté d'origine.*

*J'ai pris part à de nombreux dialogues interdisciplinaires et j'ai beaucoup travaillé pour réussir à entendre les collègues, que j'écoutais sans trop comprendre au début de ma thèse. La communication à l'interface est un exercice compliqué, exigeant et très formateur, et m'a permis de travailler tant en émission qu'en réception. En plus de la dimension pédagogique nécessaire pour communiquer mes idées et permettant de faciliter plus largement les échanges interdisciplinaires, celle du langage à l'interface est devenue très rapidement pour moi un des lieux du travail, un objet d'étude en soi.*

*Il m'a été nécessaire de déconstruire des implicites dans le langage naturel et les raisonnements (chez moi et mes interlocuteurs) et de définir un vocabulaire commun. Je vais profiter de la discussion pour rendre compte de cette dimension de mon travail personnel, partie importante de ma formation à la recherche dans un environnement interdisciplinaire.*

*Tout ceci a été alimenté par le travail formel en langage mathématique qui interroge le langage naturel à l'interface des disciplines, et permet en retour de le préciser. Pour alimenter cette dynamique entre langage mathématique et modèles exprimés en langage naturel, nous avons notamment accordé une importance particulière au choix des symboles mathématiques, en essayant d'adopter des notations porteuses de sens afin de faciliter la communication et la compréhension des objets manipulés<sup>1</sup>.*

---

1. e.g.  $\mathcal{P}_{sn^*, z_E}$  pour la sensibilité de  $sn^*$  à  $z_E$  dans un "Pool" isolé,  $\mathcal{A}_{sn^*, x_E}$  pour une région ("Area") isolée,  $B_E$  pour le forçage Basal sur le pool Excitateur d'une région etc.



*Je reprends et discute au sens littéral le contenu de ce qui a été présenté jusqu'ici. Je vais notamment interroger les concepts de balance E/I, paramètre E/I, contrôle inhibiteur, régulation top-down, augmentation de l'inhibition dans un contexte de contrôle inhibiteur, puis au chapitre suivant de contrôle excitateur, moindre excitation et augmentation de l'inhibition dans un contexte de contrôle excitateur.*

*Cela a plusieurs objectifs :*

- donner un autre guide de lecture de ma thèse en définissant le vocabulaire et en revisitant le langage et les raisonnements implicites qu'il induit menant à la formulation d'hypothèses intuitives (vérifiées ou non) sur le comportement du modèle,*
- pouvoir préciser les nouvelles questions qu'ouvre la thèse, c'est-à-dire élaborer de la manière la plus précise et explicite possible,*
- participer et aider au dialogue entre les communautés qui s'intéressent aux neurosciences théoriques (notamment celles des neurosciences computationnelles et de la modélisation physique) et expérimentales (notamment celles de la neuropsychologie et neuroimagerie).*

*Au regard de l'attendu d'une discussion dans une thèse disciplinaire, ma proposition peut paraître inattendue. C'est précisément parce que l'éclairage apporté permet de préciser les nouvelles questions et de nourrir le dialogue interdisciplinaire qu'un tel compte-rendu me semble important à partager.*

## 13.1 Balance E/I locale

Nous avons présenté aux chapitres 3 et 7 (section 7.3) la métaphore de la balance locale de Roberval, couramment utilisée dans la communauté des neurosciences. Elle permet notamment de penser l'effet compensatoire des propriétés excitatrices et inhibitrices dans une région isolée.

Par ailleurs, le modèle neural a pu être écrit au régime stationnaire (chapitre 7) et nous avons travaillé avec des points fixes stables. Les paramètres du modèle qui donnent l'activité  $sn^*$  d'une région isolée au régime stationnaire peuvent être classés en deux catégories, "excitateur" ou "inhibiteur", que nous commençons par définir formellement.

L'objectif de cette section est de préciser les liens entre les changements de  $sn^*$  d'une région isolée lorsque l'on modifie des paramètres de chacune de ces catégories et la métaphore de la balance de Roberval. Ceci permet de montrer les avantages de l'usage de cette métaphore pour nous aider à penser le comportement du modèle<sup>2</sup> et également les limites qui peuvent conduire à des raisonnements erronés.

---

2. en région isolée ou découplée comme nous verrons à la section 13.2.3.2

### 13.1.1 Comment définit-on l'effet excitateur ou inhibiteur : Etude de l'effet d'un paramètre

Nous souhaitons d'abord définir formellement l'effet *excitateur* ou *inhibiteur* d'un paramètre, de manière plus générale que dans le cadre de la balance E/I. Nous distinguons un effet infinitésimal et un effet fini. Ces définitions seront utilisées et serviront de fil conducteur à travers tout ce chapitre.

Nous voyons ensuite comment nous les utilisons pour préciser ce que traduit dans le modèle neural le concept de balance E/I et comment la variation d'un paramètre modifie l'état de cette balance.

Pour définir l'effet excitateur ou inhibiteur d'un paramètre, il est nécessaire d'avoir une pensée en perturbation ou variation. Dans tout ce chapitre, nous adoptons les notations et le vocabulaire suivant :  $\delta$  désigne une perturbation infinitésimale (ou perturbation), alors que  $\Delta$  désigne une variation finie (ou variation). Nous les définissons positives. Nous préciserons à chaque fois s'il s'agit de la perturbation ou de la variation :

- d'un paramètre  $\pi$  (notées  $\delta\pi$  ou  $\Delta\pi$ )
- d'une variable d'observation  $\eta^*$  (notées  $\delta\eta^*$  ou  $\Delta\eta^*$ ).

#### 13.1.1.1 Etude en perturbation infinitésimale : effet excitateur ou inhibiteur infinitésimal d'un paramètre

##### Définir un effet excitateur ou inhibiteur infinitésimal

Pour parler d'un effet excitateur ou inhibiteur infinitésimal, nous proposons de préciser à chaque fois le triplet suivant :

- le **paramètre perturbé** ( $\pi$ )
- la **variable d'observation** ( $\eta^*$ <sup>a</sup>)
- la **boucle de rétroaction** ( $\mathcal{B}$ )

Dans un système complexe, de plus non linéaire, composé d'un grand nombre d'équations algébriques, il se présente un très grand nombre de couplages entre variables d'état, connectées de façon très imbriquées : l'évolution finale au point fixe de ces variables peut passer par un grand nombre de rétroactions, qu'on visualise couramment à l'aide d'un graphe orienté. Un noeud contient une variable (d'état ou intermédiaire, comme une variable de transfert) et chaque arc orienté porte un couplage.

Le qualificatif *excitateur* ou *inhibiteur* est choisi en fonction du signe de la perturbation de la variable d'observation  $\eta^*$  (i.e.  $\delta\eta^*$ ), produite par la perturbation positive du paramètre  $\pi$  (i.e.  $\delta\pi > 0$ ), à travers la boucle de rétroaction  $\mathcal{B}$ .

Si  $\delta\eta^*$  est positive (resp. négative), alors le paramètre  $\pi$  a un effet excitateur (resp. inhibiteur) infinitésimal sur la variable d'observation quand on considère cette boucle de rétroaction  $\mathcal{B}$ .

---

a. elle est définie à partir d'une fonction qui dépend du paramètre que l'on perturbe :  $\eta^* = f(\pi)$

*Comment se servir des sensibilités pour préciser cet effet ?*

Le signe de  $\delta\eta^*$  peut être établi directement à partir d'un calcul de la sensibilité de  $\eta^*$  par rapport au paramètre  $\pi$ , notée  $\frac{\delta\eta^*}{\delta\pi}$  :

$$\delta\eta^* = \frac{\delta\eta^*}{\delta\pi} \delta\pi \quad (13.1)$$

Le signe de  $\delta\eta^*$  est identique à celui de la sensibilité  $\frac{\delta\eta^*}{\delta\pi}$  (puisque par définition  $\delta\pi > 0$ ). La définition de l'effet *excitateur* ou *inhibiteur* d'un paramètre pourrait être exprimée uniquement en fonction du signe de la sensibilité de la variable d'observation  $\eta^*$  par rapport au paramètre  $\pi$ . Si la sensibilité est positive (resp. négative), le paramètre  $\pi$  a un effet excitateur (resp. inhibiteur) sur la variable d'observation  $\eta^*$  quand on considère la boucle de rétroaction  $\mathcal{B}$ .

### 13.1.1.2 Cas de la balance E/I locale : caractérisation de l'effet excitateur ou inhibiteur infinitésimal d'un paramètre

La métaphore de la balance E/I est une image locale i.e. dans une région isolée.

La variable d'observation d'intérêt est en général implicite : on s'intéresse à l'activité neurale de la région, pour nous il s'agit de  $sn^*$ .

Les paramètres d'une région isolée peuvent alors être classés en deux catégories en fonction de l'effet excitateur ou inhibiteur de leur perturbation sur l'activité  $sn^*$  de la région d'intérêt.

En perturbation, nous proposons de caractériser l'étude de la balance E/I locale à partir de la définition du triplet suivant :

- un paramètre perturbé ( $\pi$ )
- la variable d'observation ( $\eta^a$ ) : activité excitatrice  $sn^*$  de la région
- la boucle de rétroaction ( $\mathcal{B}$ ) : une région isolée, i.e. 2 pools connectés. Nous la représentons sur la figure [13.1](#)

---

*a.* elle est définie à partir d'une fonction qui dépend du paramètre que l'on fait varier

La formulation mathématique du modèle au régime stationnaire pour une région isolée se trouve en annexe [18.1.1](#).

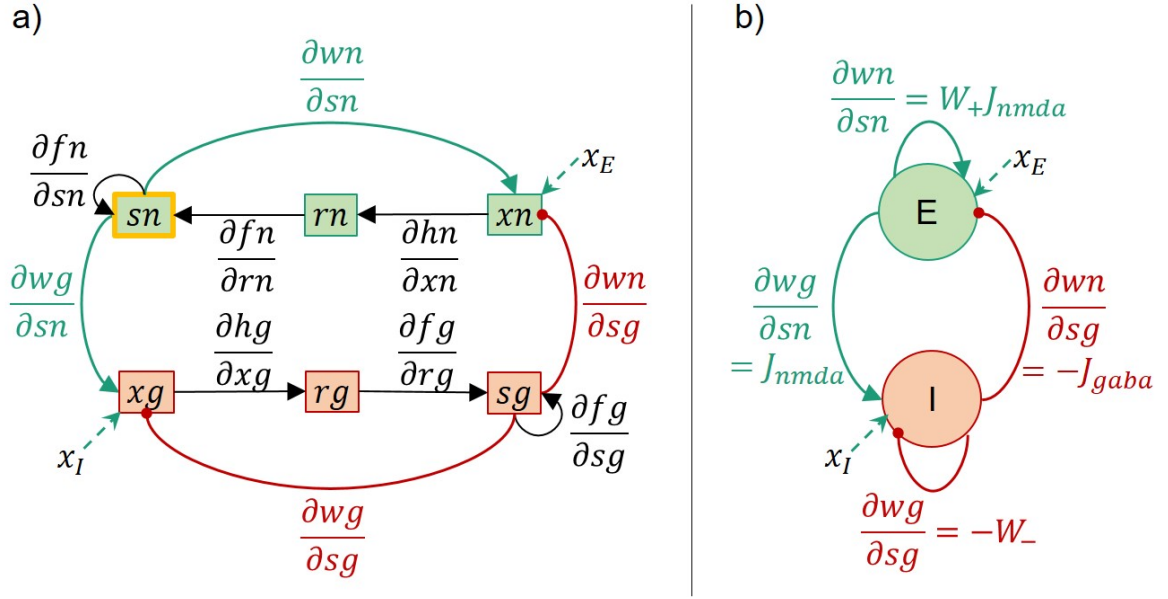


**Figure 13.1** – a) Boucle de rétroaction complète : graphe qui montre les variables du modèle (noeuds) et les dépendances entre ces variables (arêtes). Nous avons explicité la variable d'observation  $sn^*$  d'intérêt pour caractériser la balance E/I locale dans le cadre de notre modèle.  
 b) Boucle de rétroaction simplifiée pour ne garder uniquement que les couplages entre le pool E et le pool I, qui représentent des poids synaptiques. C'est le graphe plus couramment représenté (e.g. [Deco et al., 2014], [Deco and Jirsa, 2012]). La variable d'observation y est implicite.  
 Nous avons rajouté les forçages  $x_E$  et  $x_I$ , de manière à mettre à vue tous les paramètres qui figurent dans l'expression des courants  $xn^*$  et  $xg^*$ .

Dans le cadre de cette thèse, nous avons choisi d'étudier l'effet de la perturbation de deux paramètres en particulier (cf chapitre 7) : le forçage sur le pool exciteur ( $\pi \equiv x_E$ ) et le couplage du pool inhibiteur vers le pool exciteur ( $\pi \equiv J_{gaba}$ ). Nous avons vu que  $x_E$  et  $J_{gaba}$  ont respectivement un effet exciteur et inhibiteur infinitésimal sur  $sn^*$  d'une région isolée. L'effet de tous les autres paramètres qui figurent dans l'expression des courants  $xn^*$  et  $xg^*$  peut se retrouver dans l'annexe 18.1.1.

Pendant une tâche de contrôle inhibiteur, nous avons modélisé la stimulation d'une région de contrôle par une variation du forçage sur son pool exciteur (cf section 6.2.4). Nous étendons alors la définition de l'effet exciteur ou inhibiteur dû à une perturbation d'un paramètre (e.g. le forçage) dans le cas d'une variation finie.

### 13.1.1.3 Etude en variation finie : effet excitateur ou inhibiteur fini d'un paramètre

Définir un effet excitateur ou inhibiteur fini

Pour parler d'un effet excitateur ou inhibiteur fini, nous proposons de préciser à chaque fois le quadruplet suivant :

- le **paramètre** qu'on fait varier ( $\pi$ )
- l'**intervalle de variation du paramètre** ( $[\hat{\pi}; \hat{\pi} + \Delta\hat{\pi}]$ )
- la **variable d'observation** ( $\eta^*$ )
- la **boucle de rétroaction** ( $\mathcal{B}$ )

Si la variation de la variable d'observation  $\Delta\eta^*$  produite par la variation du paramètre  $\Delta\pi$  est positive (resp. négative), alors le paramètre a un effet excitateur (resp. inhibiteur) fini sur la variable d'observation quand on considère cette boucle de rétroaction  $\mathcal{B}$ .

Si le paramètre  $\pi$  a un effet excitateur (ou inhibiteur) infinitésimal sur tout l'intervalle de variation  $[\hat{\pi}; \hat{\pi} + \Delta\hat{\pi}]$ , alors le paramètre a un effet excitateur (ou inhibiteur) fini sur cet intervalle.

*Comment se servir des sensibilités pour préciser cet effet ?*

La valeur de la variable d'observation peut s'écrire comme une fonction du paramètre  $\pi$  :  $\eta^* = f(\pi)$ .

Si nous considérons une variation du paramètre  $\pi$  entre deux valeurs finies  $\hat{\pi}$  et  $\hat{\pi} + \Delta\hat{\pi}$ , la variation de la variable d'observation est définie par la différence entre deux valeurs de  $\eta^*$  évaluées pour  $\pi = \hat{\pi} + \Delta\hat{\pi}$  et pour  $\pi = \hat{\pi}$  :

$$\Delta\eta^* \equiv f(\hat{\pi} + \Delta\hat{\pi}) - f(\hat{\pi}) \quad (13.2)$$

Cette variation de la variable d'observation peut également s'écrire comme la somme des perturbations  $\delta\eta^*$  due à des perturbations  $\delta\pi$  entre  $\hat{\pi}$  et  $\hat{\pi} + \Delta\hat{\pi}$ . Il est alors possible de construire la variation  $\Delta\eta^*$  due à la variation  $\Delta\pi$  comme une intégrale de la sensibilité  $\frac{\delta\eta^*}{\delta\pi}$ <sup>3</sup> :

$$\Delta\eta^* = \int_{\hat{\pi}}^{\hat{\pi} + \Delta\hat{\pi}} \frac{\delta\eta^*}{\delta\pi} d\pi \quad (13.3)$$

Si la sensibilité  $\frac{\delta\eta^*}{\delta\pi}$  est de signe constant sur l'intervalle d'intégration de  $\pi$  alors l'intégrale est du même signe. Il y a alors conservation du signe entre la sensibilité sur l'intervalle d'intégration et la variation  $\Delta\eta^*$ .

Soit  $\pi \in [\hat{\pi}; \hat{\pi} + \Delta\hat{\pi}]$ . Supposons que  $\pi$  soit un paramètre avec un effet excitateur (resp. inhibiteur) infinitésimal sur  $\eta^*$  au sens de la définition précédente :  $\forall \pi \in [\hat{\pi}; \hat{\pi} + \Delta\hat{\pi}]$ , une perturbation  $\delta\pi > 0$  cause une perturbation  $\delta\eta^* > 0$  (resp.  $\delta\eta^* < 0$ ). Autrement dit,  $\forall \pi \in$

3. à condition que celle-ci soit continue et dérivable sur l'intervalle de variation du paramètre

$[\hat{\pi}; \hat{\pi} + \Delta\hat{\pi}]$ , la sensibilité  $\frac{\delta\eta^*}{\delta\pi}$  est positive (resp. négative). Nous pouvons en déduire que la variation totale  $\Delta\hat{\pi}$  cause une variation  $\Delta\eta$  positive (resp. négative). La variation de  $\pi$  a donc un effet excitateur (resp. inhibiteur) fini sur la variable d'observation, quand on considère cette boucle de rétroaction.

En conclusion, si un paramètre a un *effet excitateur* (resp. *inhibiteur*) *infinitésimal* sur un intervalle, alors la variation de ce paramètre entre deux valeurs finies de cet intervalle cause un *effet excitateur* (resp. *inhibiteur*) *fini* sur la variable d'observation, quand on considère cette boucle de rétroaction.

#### 13.1.1.4 Augmentation de la variation du paramètre : prédiction de l'effet ?

Supposons un paramètre  $\pi$  avec un effet excitateur (resp. inhibiteur) fini sur  $\eta^*$  entre  $\hat{\pi}$  et  $\hat{\pi} + \Delta\hat{\pi}$ .

Supposons qu'on augmente la variation du paramètre  $\pi$  d'un  $\Delta\hat{\pi}_s$  supplémentaire. L'intervalle de variation du paramètre devient :  $[\hat{\pi}; \hat{\pi} + \Delta\hat{\pi} + \Delta\hat{\pi}_s]$ .

Pour pouvoir prédire l'effet de cette augmentation supplémentaire du paramètre  $\pi$  sur la variation de la variable d'observation  $\eta^*$ , il est nécessaire que la sensibilité  $\frac{\delta\eta^*}{\delta\pi}$  soit de signe constant sur tout l'intervalle de variation du paramètre.

En effet, la variation de la variable d'observation  $\Delta\eta^*$  est égale à :

$$\Delta\eta^* = \int_{\hat{\pi}}^{\hat{\pi}+\Delta\hat{\pi}} \frac{\delta\eta^*}{\delta\pi} d\pi + \int_{\hat{\pi}+\Delta\hat{\pi}}^{\hat{\pi}+\Delta\hat{\pi}+\Delta\hat{\pi}_s} \frac{\delta\eta^*}{\delta\pi} d\pi \quad (13.4)$$

Si la sensibilité est toujours positive (resp. négative) sur  $[\hat{\pi} + \Delta\hat{\pi}; \hat{\pi} + \Delta\hat{\pi} + \Delta\hat{\pi}_s]$ , i.e. si  $\pi$  a toujours un effet excitateur (resp. inhibiteur) infinitésimal sur tout l'intervalle de sa variation, alors la variation de la variable d'observation  $\Delta\eta^*$  reste positive (resp. négative) et augmente en valeur absolue. La quantité supplémentaire est égale au deuxième terme de la somme (13.4).

#### Précision sur l'emploi du vocabulaire

Notons que nous prenons le soin d'attribuer l'adjectif "excitateur" (ou "inhibiteur") à l'effet que produit la perturbation ou variation du paramètre sur la variable d'observation. Nous évitons autant que possible la formulation "paramètre excitateur" (ou "paramètre inhibiteur"), et si nous venions à l'utiliser ce serait un raccourci de langage pour l'expression "paramètre dont la perturbation ou variation a un effet excitateur (ou inhibiteur) sur une variable d'observation, étant donné une boucle de rétroaction". La boucle de rétroaction et la variable d'observation doivent être précisées.

Cette nuance aura toute son importance dans la section 13.2. Nous verrons que dans un système composé de deux régions couplées entre elles (qui définit une boucle de rétroaction), l'effet excitateur ou inhibiteur de la perturbation ou variation du forçage sur le pool excitateur de la région de contrôle dépend du choix de la variable d'observation. Ce n'est donc pas le paramètre qui est en soi excitateur ou inhibiteur. Il est important de préciser la variable d'observation, c'est-à-dire "sur quoi il y a un effet excitateur (ou inhibiteur)", en plus du signe de sa variation.

### 13.1.1.5 Cas de la balance E/I : caractérisation de l'effet excitateur ou inhibiteur fini d'un paramètre

Faire varier l'état de la balance E/I locale implique de s'intéresser aux éléments suivants :

- le paramètre qu'on fait varier ( $\pi$ )
- l'intervalle de variation du paramètre ( $[\hat{\pi}; \hat{\pi} + \Delta\hat{\pi}]$ )
- la variable d'observation ( $\eta$ ) : activité excitatrice  $sn^*$  de la région
- la boucle de rétroaction ( $\mathcal{B}$ ) : une région isolée, i.e. 2 pools connectés (cf figure 13.1)

Pour évaluer la variation de l'activité  $sn^*$  due à une variation du paramètre d'intérêt  $\pi$  (par exemple  $B_E$  ou  $J_{gaba}$ ), nous pouvons directement soustraire les valeurs de points fixes :

$$\Delta sn^* = f(\hat{\pi} + \Delta\hat{\pi}) - f(\hat{\pi}) \quad (13.5)$$

Une deuxième méthode consiste à écrire la variation  $\Delta sn^*$  en fonction de la sensibilité de  $sn^*$  au paramètre d'intérêt  $\pi$  :

$$\Delta sn^* = \int_{\hat{\pi}}^{\hat{\pi} + \Delta\hat{\pi}} \frac{\delta sn^*}{\delta \pi} d\pi \quad (13.6)$$

Pour une valeur fixée de  $J_{gaba}$ <sup>4</sup>, nous faisons varier le forçage  $x_E$  entre 0.382 nA et 0.482 nA (ce qui correspond à une valeur de  $I_{stim}$  entre 0 et 0.1 nA). Sur les graphes de la figure 13.2, nous avons représenté  $sn^*$  en fonction de  $x_E$  (en haut), ainsi que la sensibilité de  $sn^*$  par rapport à  $x_E$  (en bas). L'expression analytique de la sensibilité (qui dépend de  $sn^*$ ) est donnée dans les annexes de l'article du chapitre 9 et est égale à la pente de la tangente à la courbe  $sn^* = f(x_E)$  au voisinage de  $x_E$ . Les pentes représentées sur le graphe du haut sont positives, comme le signe de la sensibilité  $\frac{\delta sn^*}{\delta x_E}$ . La perturbation de  $x_E$  a ainsi un effet excitateur infinitésimal sur  $sn^*$ , quand on considère la boucle de rétroaction correspondant à une région isolée (sur la figure 13.1).

La variation de  $sn^*$  entre  $x_E = \hat{x}_E$  et  $x_E = \hat{x}_E + \Delta\hat{x}_E$ , peut être construite comme une somme de  $\delta sn^*$ . Elle se traduit par l'intégrale de la sensibilité entre ces deux valeurs de  $x_E$ . Sur la figure 13.2, nous prenons l'exemple d'une variation entre  $x_E = 0.382$  et  $x_E = 0.460$ . Comme la sensibilité est positive sur l'intervalle d'intégration, l'intégrale est égale à l'aire hachurée sur le graphe du bas. La variation de  $sn^*$  résultante est donc positive (représentée sur le graphe du haut par la flèche verte pointant vers le haut). La variation de  $x_E$  entre  $x_E = 0.382$  et  $x_E = 0.460$  a un effet excitateur fini sur  $sn^*$ , quand on considère la boucle de rétroaction correspondant à une région isolée.

Plus  $\Delta\hat{x}_E$  augmente et plus l'aire sous la courbe de sensibilité est grande, donc plus la variation de  $sn^*$  est grande. Nous vérifions sur le graphe du haut de la figure 13.2 que  $sn^*$  augmente quand  $x_E$  augmente.

La métaphore de la balance correspondant à cette variation est représentée à la figure 13.3. Ici, augmenter  $x_E$  est représenté par l'ajout d'un poids sur le plateau excitateur (vert).

4. choisie ici de manière à avoir  $rn = 3$  Hz au repos, i.e. quand  $x_E = 0.382$  nA.

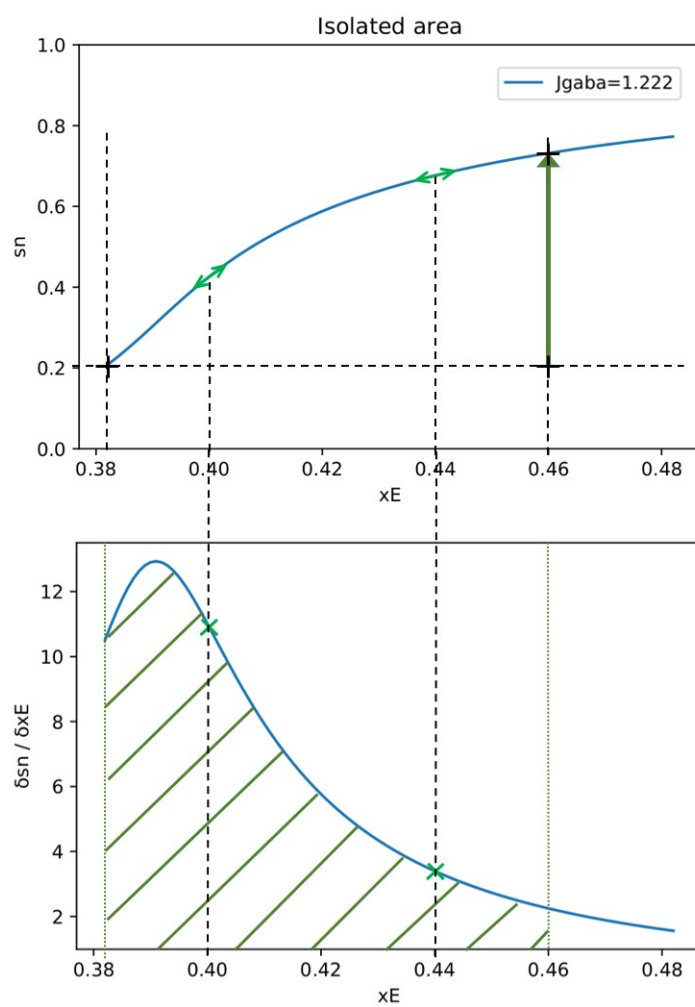


**Figure 13.2** – Activité  $sn^*$  en fonction de  $x_E$  (en haut), et sensibilité de  $sn^*$  par rapport à  $x_E$  (en bas) à  $J_{gaba}$  fixé.



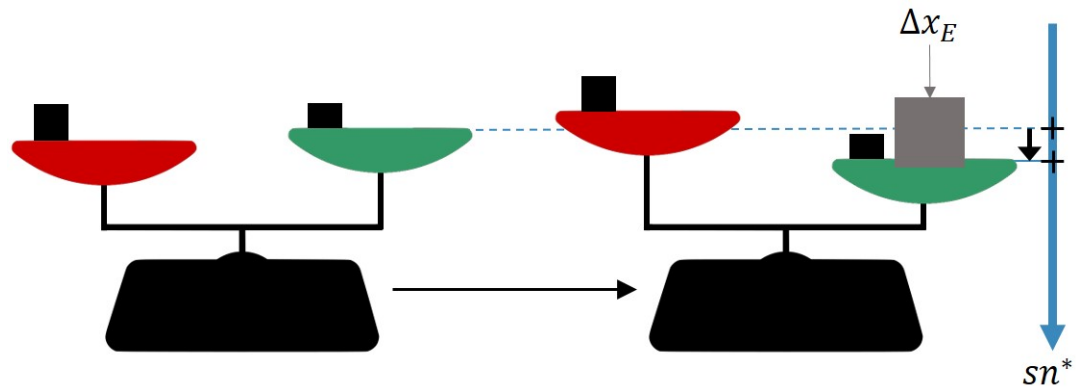


**Figure 13.3** – Métaphore de la balance E/I locale en variation correspondant à la figure 13.2.

Le niveau du plateau vert fixe l'activité  $sn^*$  de la région .

A gauche : Les paramètres du modèle sont représentés par des poids noirs qui fixent le niveau des plateaux, correspondant à une activité  $sn^*$  avant toute variation de  $x_E$ .

A droite : La variation du paramètre  $x_E$ , notée  $\Delta x_E$ , est représentée par l'ajout d'un nouveau poids sur la balance. Celui-ci n'est pas qualifié en soi d'excitateur ou d'inhibiteur (il n'est ni vert ni rouge), contrairement aux plateaux de la balance qui sont qualifiés respectivement d'excitateur et d'inhibiteur (ils sont vert et rouge). L'analyse du comportement du modèle a permis de savoir que le paramètre produit un effet excitateur fini sur la variable d'observation  $sn^*$ . Le poids  $\Delta x_E$  est alors placé sur le plateau excitateur. La balance penche en faveur du plateau E, ce qui entraîne une augmentation de  $sn^*$  :  $\Delta sn^* > 0$ .

Pour une valeur fixée de  $x_E$ <sup>5</sup>, nous faisons varier le coefficient de couplage  $J_{gaba}$  entre 0.2 nA et 2 nA. Sur les graphes de la figure 13.4, nous avons représenté  $sn^*$  en fonction de  $J_{gaba}$  (en haut), ainsi que la sensibilité de  $sn^*$  par rapport à  $J_{gaba}$  (en bas). L'expression analytique de la sensibilité (qui dépend de  $sn^*$ ) est donnée dans l'annexe 18.1.1 et est égale à la pente de la tangente à la courbe  $sn^* = f(J_{gaba})$  au voisinage de  $J_{gaba}$ . Les pentes représentées sur le graphe du haut sont négatives, comme le signe de la sensibilité  $\frac{\delta sn}{\delta J_{gaba}}$ . La perturbation de  $J_{gaba}$  a ainsi un effet inhibiteur infinitésimal sur  $sn^*$ , quand on considère la boucle de rétroaction correspondant à une région isolée (sur la figure 13.1).

La variation de  $sn^*$  entre  $J_{gaba} = \widehat{J_{gaba}}$  et  $J_{gaba} = \widehat{J_{gaba}} + \Delta \widehat{J_{gaba}}$ , peut être construite comme une somme de  $\delta sn^*$ . Elle se traduit par l'intégrale de la sensibilité entre ces deux valeurs de  $J_{gaba}$ . Sur la figure 13.4, nous prenons l'exemple d'une variation entre  $J_{gaba} = 0.2$  et  $J_{gaba} = 1.5$ . Comme la sensibilité est négative sur l'intervalle d'intégration, l'intégrale est négative et égale en valeur absolue à l'aire hachurée sur le graphe du bas. La variation de  $sn^*$  résultante est donc négative (représentée sur le graphe du haut par la flèche rouge pointant vers le bas). La variation de  $J_{gaba}$  entre  $J_{gaba} = 0.2$  et  $J_{gaba} = 1.5$  a un effet inhibiteur fini sur  $sn^*$ , quand on considère la boucle de rétroaction correspondant à une région isolée. Plus  $\Delta \widehat{J_{gaba}}$  augmente et plus l'aire hachurée est grande, donc plus la variation négative de  $sn^*$  est grande en valeur absolue. Nous vérifions sur le graphe du haut de la figure 13.4 que  $sn^*$  diminue quand  $J_{gaba}$  augmente.

La métaphore de la balance correspondant à cette variation est représentée à la figure 13.5. Ici, augmenter  $J_{gaba}$  est représenté par l'ajout d'un poids sur le plateau inhibiteur (rouge).

---

5. choisie ici au repos, i.e.  $x_E = 0.382$  nA.

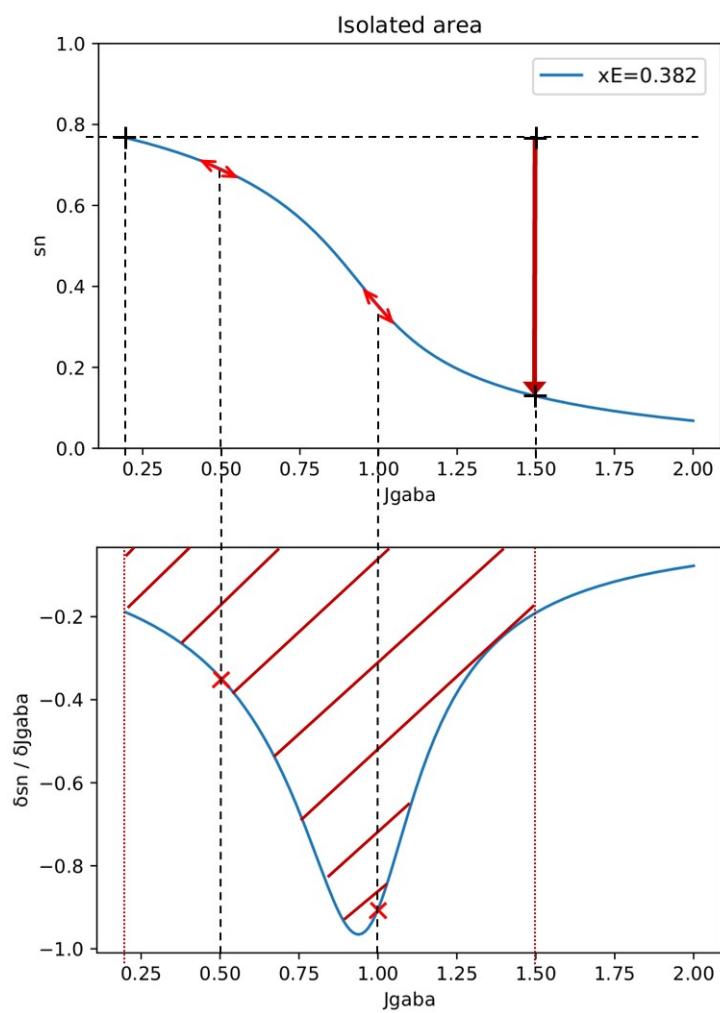


**Figure 13.4** – Activité  $sn^*$  en fonction de  $J_{gaba}$  (en haut), et sensibilité de  $sn^*$  par rapport à  $J_{gaba}$  (en bas), pour une valeur de  $x_E$  fixée.

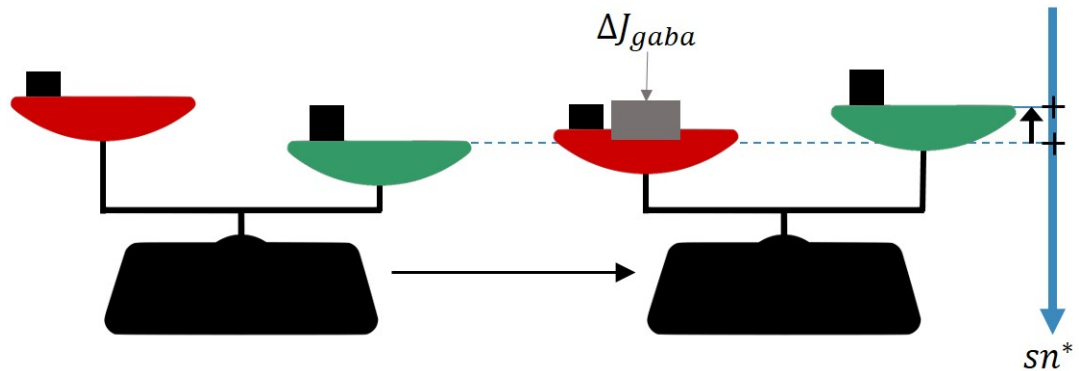


**Figure 13.5** – Métaphore de la balance E/I locale en variation correspondant à la figure 13.4.

Le niveau du plateau vert fixe l'activité  $sn^*$  de la région.

A gauche : Les paramètres du modèle sont représentés par des poids noirs qui fixent le niveau des plateaux, correspondant à une activité  $sn^*$  avant toute variation de  $J_{gaba}$ .

A droite : La variation du paramètre  $J_{gaba}$ , notée  $\Delta J_{gaba}$ , est représentée par l'ajout d'un nouveau poids sur la balance. Celui-ci n'est pas qualifié en soi d'excitateur ou d'inhibiteur (il n'est ni vert ni rouge), contrairement aux plateaux de la balance qui sont qualifiés respectivement d'excitateur et d'inhibiteur (ils sont vert et rouge). L'analyse du comportement du modèle a permis de savoir que le paramètre produit un effet inhibiteur fini sur la variable d'observation  $sn^*$ . Le poids  $\Delta J_{gaba}$  est alors placé sur le plateau inhibiteur. La balance penche en défaveur du plateau E, ce qui entraîne une diminution de  $sn^*$  :  $\Delta sn^* < 0$ .

### 13.1.2 Etude de l'effet de plusieurs paramètres

Dans la section précédente, nous avons perturbé ou fait varier un seul paramètre de la balance E/I. Dans cette section, nous étudions l'effet de la perturbation ou variation de plusieurs paramètres de cette balance.

#### 13.1.2.1 Perturbation simultanée de plusieurs paramètres

Nous perturbons simultanément les deux paramètres  $\pi_i$  et  $\pi_j$ . La linéarisation du modèle perturbé autour du point fixe permet de séparer et sommer les contributions dues à la perturbation de chaque paramètre (comme dans l'annexe 18.1.1). Nous pouvons alors écrire :

$$\delta sn = S_{\pi_i}(\pi_i, \pi_j)\delta\pi_i + S_{\pi_j}(\pi_i, \pi_j)\delta\pi_j \quad (13.7)$$

$S_{\pi_i}(\pi_i, \pi_j)$  désigne la sensibilité de  $sn^*$  à une perturbation  $\delta\pi_i$ , qui dépend des valeurs fixées de  $\pi_i$  et  $\pi_j$ .

$S_{\pi_j}(\pi_i, \pi_j)$  désigne la sensibilité de  $sn^*$  à une perturbation  $\delta\pi_j$ , qui dépend des valeurs fixées de  $\pi_i$  et  $\pi_j$ .

En reprenant les mêmes notations que dans l'article au chapitre 9 :  $S_{\pi_i}(\pi_i, \pi_j) \equiv \mathcal{A}_{sn, \pi_i}$  et  $S_{\pi_j}(\pi_i, \pi_j) \equiv \mathcal{A}_{sn, \pi_j}$ .

$$\delta sn = \mathcal{A}_{sn, \pi_i}\delta\pi_i + \mathcal{A}_{sn, \pi_j}\delta\pi_j \quad (13.8)$$

Si  $\mathcal{A}_{sn, \pi_i}$  et  $\mathcal{A}_{sn, \pi_j}$  sont de signes contraires, les contributions se compensent et les paramètres n'appartiennent pas au même plateau de la balance. C'est ce que nous avons vu à la section 7.3 avec les perturbations de  $x_E$  et de  $J_{gaba}$  (notion de compensation de poids placés de part et d'autre de la balance).

#### 13.1.2.2 Variation simultanée de plusieurs paramètres

L'objectif de ce paragraphe est de relier le comportement de  $sn^*$  dans une région cérébrale isolée quand on fait varier au moins deux paramètres  $\pi_i$  et  $\pi_j$ , aux images que l'on se fait du comportement d'une balance sur laquelle on ajouterait simultanément au moins deux poids sur les plateaux.

Nous passons d'abord par un travail formel pour construire les variations d'activité  $sn^*$  dues aux variations des paramètres, avant de pouvoir faire des liens avec la métaphore de la balance E/I.

Au point fixe, la valeur  $sn^*$  dépend des paramètres du modèle. On note  $sn^* = f(\pi_i, \pi_j)$ . Nous fixons les paramètres  $\pi_i$  et  $\pi_j$  et partons de la valeur de point fixe correspondante :  $sn^* = f(\widehat{\pi_i}, \widehat{\pi_j})$ .

Nous faisons ensuite varier simultanément  $\pi_i$  entre  $\widehat{\pi_i}$  et  $\widehat{\pi_i} + \Delta\widehat{\pi_i}$ ; et  $\pi_j$  entre  $\widehat{\pi_j}$  et  $\widehat{\pi_j} + \Delta\widehat{\pi_j}$ .  $\Delta sn^*$  est la variation totale de  $sn^*$  due à la variation simultanée de ces deux paramètres. Par définition, elle est égale à la différence entre les valeurs de  $sn^*$  évaluées entre les valeurs finales et initiales des intervalles de variations des paramètres :

$$\Delta sn^* \equiv f(\widehat{\pi_i} + \Delta\widehat{\pi_i}, \widehat{\pi_j} + \Delta\widehat{\pi_j}) - f(\widehat{\pi_i}, \widehat{\pi_j}) \quad (13.9)$$

Nous montrons en annexe 18.1.2.1 que  $\Delta sn^*$  se décompose en deux étapes successives correspondant à la variation successive de chaque paramètre. Elle peut s'écrire de deux manières :

$$\begin{cases} \Delta sn = \Delta sn_{\pi_i}(\widehat{\pi_j}) + \Delta sn_{\pi_j}(\widehat{\pi_i} + \Delta \widehat{\pi_i}) \\ \quad = \Delta sn_{\pi_j}(\widehat{\pi_i}) + \Delta sn_{\pi_i}(\widehat{\pi_j} + \Delta \widehat{\pi_j}) \end{cases} \quad (13.10)$$

avec la notation  $\Delta sn_{\pi_i}(\pi_j)$  égale à la variation de  $sn^*$  quand on fait varier  $\pi_i$  entre  $\widehat{\pi_i}$  et  $\widehat{\pi_i} + \Delta \widehat{\pi_i}$ , évaluée pour une valeur particulière de  $\pi_j$ . De la même manière,  $\Delta sn_{\pi_j}(\pi_i)$  est égale à la variation de  $sn^*$  quand on fait varier  $\pi_j$  entre  $\widehat{\pi_j}$  et  $\widehat{\pi_j} + \Delta \widehat{\pi_j}$ , évaluée pour une valeur particulière de  $\pi_i$ .

La première égalité contient des différences calculées entre deux points successifs le long du chemin suivant :

$$sn^* = f(\widehat{\pi_i}, \widehat{\pi_j}) \longrightarrow sn^* = f(\widehat{\pi_i} + \Delta \widehat{\pi_i}, \widehat{\pi_j}) \longrightarrow sn^* = f(\widehat{\pi_i} + \Delta \widehat{\pi_i}, \widehat{\pi_j} + \Delta \widehat{\pi_j}).$$

On a d'abord fait varier  $\pi_i$ , puis  $\pi_j$ .

La seconde égalité traduit des différences calculées le long du chemin :

$$sn^* = f(\widehat{\pi_i}, \widehat{\pi_j}) \longrightarrow sn^* = f(\widehat{\pi_i}, \widehat{\pi_j} + \Delta \widehat{\pi_j}) \longrightarrow sn^* = f(\widehat{\pi_i} + \Delta \widehat{\pi_i}, \widehat{\pi_j} + \Delta \widehat{\pi_j}).$$

On a cette fois-ci fait varier  $\pi_j$  en premier, puis  $\pi_i$ .

Pour ces deux chemins, le point fixe intermédiaire diffère puisqu'il est obtenu après une variation soit de  $\pi_i$  soit de  $\pi_j$ . Cependant, la variation totale est identique : elle est égale à la différence entre l'état final et l'état initial, qui sont identiques entre les deux chemins.

La variation totale  $\Delta sn^*$  est égale à la somme des effets produits par la variation de chaque paramètre un par un, en partant à chaque fois du point fixe final atteint après la variation du paramètre précédent.

Contrairement à l'étude en perturbation, il ne suffit pas de sommer les deux variations prises séparément pour obtenir la variation totale de  $sn^*$  :  $\Delta sn^* \neq \Delta sn^*_{/\pi_i}(\widehat{\pi_j}) + \Delta sn^*_{/\pi_j}(\widehat{\pi_i})$ <sup>6</sup>. En effet, la variation de  $sn^*$  due à celle d'un des paramètres dépend de la valeur du second. Il n'y a pas de raison de penser que la variation d'un paramètre produise toujours le même effet sur  $sn^*$ , i.e. a priori  $\Delta sn^*_{/\pi_i}(\widehat{\pi_j}) \neq \Delta sn^*_{/\pi_i}(\widehat{\pi_j} + \Delta \widehat{\pi_j})$  et  $\Delta sn^*_{/\pi_j}(\widehat{\pi_i}) \neq \Delta sn^*_{/\pi_j}(\widehat{\pi_i} + \Delta \widehat{\pi_i})$ .

Les variations  $\Delta sn^*_{/\pi_i}$  et  $\Delta sn^*_{/\pi_j}$  peuvent ne pas être de même signes et se compenser. C'est ce que nous illustrons à la figure 13.6.

Tous ces éléments sont en cohérence avec la métaphore de la balance E/I sur laquelle on poserait plusieurs poids (cf figure 13.6). La position finale de la balance est le résultat de la somme des déplacements successifs produits par les poids posés un par un sur les plateaux. En fonction de l'ordre dans lequel les poids sont posés, les déplacements intermédiaires des plateaux diffèrent. Les positions initiale et finale sont cependant identiques.

6. L'application qui lie les variations des deux paramètres à  $\Delta sn^*$  n'est donc pas une application linéaire (au sens algébrique) car elle ne vérifie pas la propriété d'additivité.

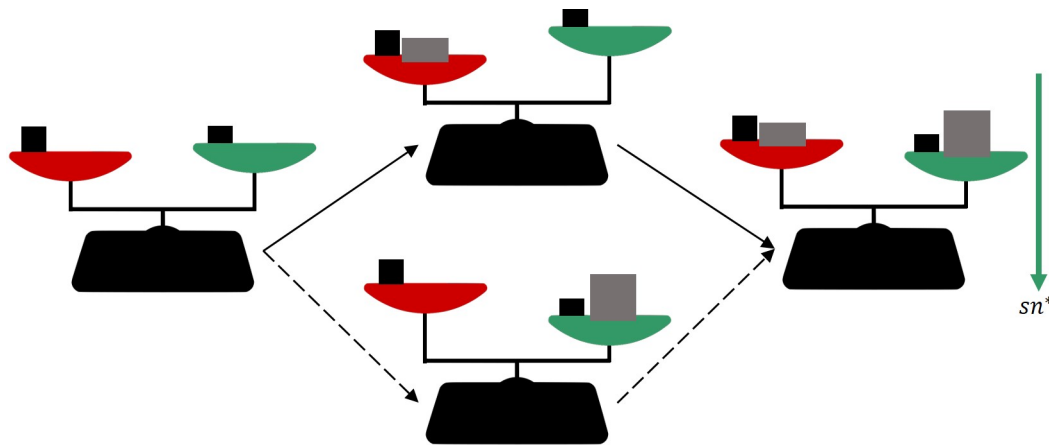


**Figure 13.6** – Ajout de deux poids successifs sur une balance, avec deux chemins possibles (en trait plein et en pointillés).

A gauche : situation initiale. Des poids noirs sont posés sur chacun des plateaux et correspondent aux paramètres du modèle qui fixent l'activité  $sn^*$  de la région. Celle-ci est donnée par la hauteur du plateau vert.

Nous ajoutons deux nouveaux poids (en gris), qui représentent les variations respectives de deux paramètres. Ces poids ne sont pas qualifiés en soi d'excitateurs ou d'inhibiteurs (ils ne sont ni verts ni rouges) et sont positionnés sur les plateaux excitateurs ou inhibiteurs (verts ou rouges) en fonction de l'effet qu'ils produisent sur  $sn^*$ .

Au milieu : deux situations intermédiaires possibles en fonction de l'ordre dans lequel les deux poids sont placés sur chacun des plateaux (plateau rouge en premier pour le chemin du haut et plateau vert en premier pour celui du bas).

A droite : situation finale. L'ajout du second poids amène à un même état final de la balance.

Nous pouvons observer l'effet compensatoire de poids placés de part et d'autre de la balance, représentant la compensation des effets excitateurs et inhibiteurs sur la variable d'observation  $sn^*$ .

### 13.1.3 Résumé de la métaphore de la balance E/I locale pour penser le comportement du modèle d'une région isolée

Pour faire les liens entre notre modèle au régime stationnaire et la balance E/I, il est nécessaire d'explicitier l'idée que le niveau du plateau E correspond à un niveau d'activité  $sn^*$  de la région, comme représenté sur la figure 13.3 par exemple. C'est la variable d'observation sur laquelle nous regardons l'effet des paramètres et qui est rendue implicite dans la métaphore de la balance E/I. A la section 3.3.1, nous avons déjà énoncé l'idée que le niveau du plateau E indiquait un niveau d'activité de la région hippocampique, associé à des performances mnésiques (cf figure 3.7).

La métaphore de la balance E/I locale est surtout utile pour se représenter les effets de la modification de poids sur le niveau des plateaux, et donc sur le changement de l'activité  $sn^*$ . Elle permet d'illustrer l'effet excitateur ou inhibiteur d'une modification des paramètres sur l'activité neurale du pool excitateur de la région au régime stationnaire.

Nous avons distingué les effets E ou I infinitésimaux (pensée en perturbation), des effets E ou I finis (pensée en variation). L'image de la balance peut être utilisée dans les deux cas. Dans le cas infinitésimal, l'expérience de pensée consiste à manipuler des "masses infinitésimales"  $\delta\pi$ , qui causent des mouvements de plateaux infinitésimaux.

Dans le cas fini, la variation d'un paramètre  $\Delta\pi$  correspond cette fois-ci à l'ajout d'une masse (un "poids") sur un des plateaux de la balance, tel que nous en avons déjà fait l'expérience perceptive.

#### 13.1.3.1 Au régime stationnaire

Les paramètres du modèle fixent un état de la balance dont le niveau du plateau E correspond à un niveau d'activité  $sn^*$ .

Nous pouvons ajouter des poids sur chacun des plateaux de la balance pour produire une variation de l'activité  $sn^*$ .

#### 13.1.3.2 Changement du poids sur le plateau excitateur

*Pensée en perturbation ou variation :*

Augmenter la valeur d'un paramètre à effet excitateur (respectivement de  $\delta\pi_E > 0$  ou  $\Delta\pi_E > 0$ ) se traduit par le fait de poser une masse supplémentaire sur le plateau E de la balance. En conséquence, la balance penche en faveur du plateau E qui s'abaisse, l'activité  $sn^*$  augmente (resp. de  $\delta sn^* > 0$  ou  $\Delta sn^* > 0$ ). C'est ce que nous avons représenté à la figure 13.3.

Inversement, diminuer la valeur d'un paramètre à effet excitateur ( $\delta\pi_E < 0$  ou  $\Delta\pi_E < 0$ ) se traduit par le fait d'enlever une masse du plateau E. En conséquence, la balance penche en défaveur du plateau E qui remonte et l'activité  $sn^*$  diminue ( $\delta sn^* < 0$  ou  $\Delta sn^* < 0$ ).



*Pensée en variation :*

Plus le poids  $\Delta\pi_E$  ajouté est grand, et plus  $sn^*$  augmente (i.e. plus la variation de la variable d'observation  $\Delta sn^* > 0$  est grande), et inversement.

### 13.1.3.3 Changement du poids sur le plateau inhibiteur

*Pensée en perturbation ou variation :*

Augmenter la valeur d'un paramètre à effet inhibiteur (respectivement de  $\delta\pi_I > 0$  ou  $\Delta\pi_I > 0$ ) se traduit par le fait de poser une masse supplémentaire sur le plateau I de la balance. Le plateau E remonte alors et l'activité  $sn^*$  diminue (resp. de  $\delta sn^* < 0$  ou  $\Delta sn^* < 0$ ). C'est ce que nous avons représenté à la figure 13.5

Inversement, diminuer la valeur d'un paramètre à effet inhibiteur ( $\delta\pi_I < 0$  ou  $\Delta\pi_I < 0$ ) se traduit par le fait d'enlever une masse du plateau I. En conséquence, la balance penche en faveur du plateau E qui baisse et l'activité  $sn^*$  augmente ( $\delta sn^* > 0$  ou  $\Delta sn^* > 0$ ).

*Pensée en variation :*

Plus le poids  $\Delta\pi_I$  ajouté est grand, et plus  $sn^*$  diminue (i.e. la variation négative de la variable d'observation  $\Delta sn^*$  devient plus grande en valeur absolue), et inversement.

Notons que c'est l'analyse du comportement du modèle qui permet de savoir si le paramètre a un effet excitateur ou inhibiteur (infinitésimal ou fini) sur  $sn^*$  et permet donc de savoir sur quel plateau poser le poids supplémentaire que représente la variation positive du paramètre.

### 13.1.3.4 Variation de plusieurs poids

La variation de plusieurs paramètres en même temps conduit à ajouter ou retirer plusieurs poids sur les plateaux E et/ou I. La position finale des plateaux résulte de la somme des déplacements produits par les ajouts successifs de poids sur chacun des plateaux. La balance E/I permet ainsi de se représenter les effets compensatoires sur le niveau des plateaux d'une modification des poids de part et d'autre, et donc sur le changement d'activité  $sn^*$ . C'est ce que nous avons représenté à la figure 13.6.

### 13.1.4 Limites de la métaphore de la balance

Dans cette section, nous discutons les limites de l'image de la balance E/I pour penser l'effet d'une variation des paramètres sur l'activité  $sn^*$  dans le cadre de notre modèle. Cette représentation intuitive amène en effet à penser l'effet de la variation de paramètres sur  $sn^*$  de manière linéaire<sup>7</sup>.

Nous traitons d'abord le cas de la variation d'un paramètre, puis nous travaillerons ensuite sur les limites de l'image de la balance pour penser le comportement du modèle dans le cadre de la variation de deux paramètres.

#### 13.1.4.1 Effet de la variation d'un paramètre sur $sn^*$

##### 13.1.4.1.1 Propriété de linéarité ?

Considérons une variation du paramètre  $\pi$  entre  $\hat{\pi}$  et  $\hat{\pi} + \Delta\hat{\pi}$ , qui produit une variation de l'activité  $sn^*$  de  $\Delta sn^*$ . Un raisonnement intuitif, basée sur l'image de la balance E/I (cf figure 13.7), consisterait à dire que multiplier la variation du paramètre par  $\alpha$  conduirait à multiplier aussi la variation de l'activité  $sn^*$  par  $\alpha$  (i.e. produirait une variation l'activité de  $\alpha\Delta sn^*$ ). Pour que ce soit le cas, nous montrons ci-dessous qu'il faudrait que la sensibilité de  $sn^*$  par rapport à  $\pi$  soit constante sur l'intervalle d'intégration, ce qui n'est pas nécessairement le cas. Nous montrons alors qu'il n'existe pas d'application linéaire entre la variation du paramètre ( $\Delta\pi$ ) et la variation de la variable d'observation  $sn^*$  ( $\Delta sn^*$ ).

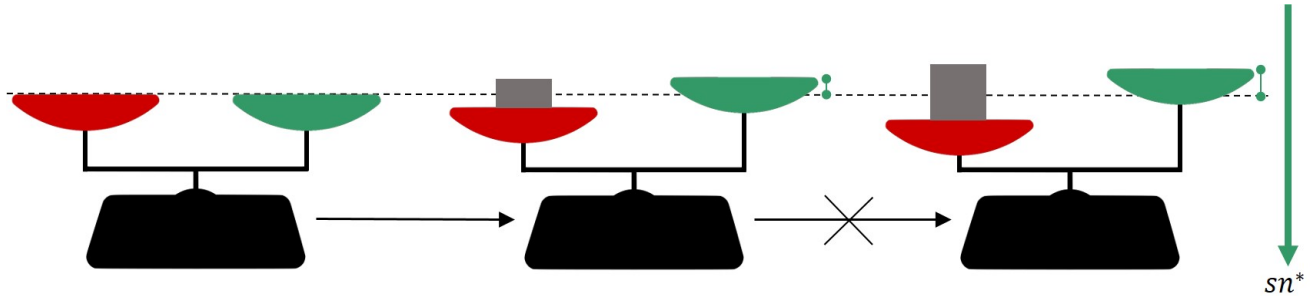


**Figure 13.7** – Multiplication de la masse par 2, représentant une multiplication par 2 de la variation d'un paramètre (ici à effet inhibiteur fini). La variation de  $sn^*$  n'est pas multipliée par deux dans le modèle : le déplacement des plateaux ne peut pas être multiplié par 2 si on se sert de la balance pour illustrer le comportement du modèle, contrairement à l'attendu. Pour visualiser plus simplement le déplacement des plateaux, nous avons représenté comme situation initiale une balance vide à l'équilibre. D'après le modèle de la balance expliqué à la section 13.1.3, elle ne l'est pas : des poids fixent l'activité  $sn^*$  avant toute variation de paramètre.

Prenons par exemple le cas du forçage  $x_E$  par rapport auquel nous avons calculé les sensibilités au chapitre 9. Supposons d'abord une variation du forçage  $x_E$  d'une valeur de

7. au sens algébrique d'une application linéaire

$I_{stim}$  ( $\Delta\widehat{x}_E = I_{stim}$ ). La variation d'activité peut s'écrire :

$$\Delta sn \equiv \int_{\widehat{x}_E}^{\widehat{x}_E + I_{stim}} \mathcal{A}_{sn, x_E} dx_E \quad (13.11)$$

Si la stimulation est multipliée par  $\alpha$ , la variation résultante s'écrit :

$$\int_{\widehat{x}_E}^{\widehat{x}_E + \alpha I_{stim}} \mathcal{A}_{sn, x_E} dx_E = \int_{\widehat{x}_E}^{\widehat{x}_E + I_{stim}} \mathcal{A}_{sn, x_E} dx_E + \int_{\widehat{x}_E + I_{stim}}^{\widehat{x}_E + \alpha I_{stim}} \mathcal{A}_{sn, x_E} dx_E \quad (13.12)$$

Pour avoir :

$$\int_{\widehat{x}_E}^{\widehat{x}_E + \alpha I_{stim}} \mathcal{A}_{sn, x_E} dx_E = \alpha \Delta sn \quad (13.13)$$

il faudrait que :

$$\int_{\widehat{x}_E + I_{stim}}^{\widehat{x}_E + \alpha I_{stim}} \mathcal{A}_{sn, x_E} dx_E = (\alpha - 1) \Delta sn \quad (13.14)$$

Ceci est toujours vérifié dans le cas où l'intégrande est constant. En effet, supposons que  $\forall x_E \in [\widehat{x}_E; \max(I_{stim}, \alpha I_{stim})]$ ,  $\mathcal{A}_{sn, x_E} \equiv \mathcal{A}$ .

Nous pouvons alors écrire :

$$\int_{\widehat{x}_E + I_{stim}}^{\widehat{x}_E + \alpha I_{stim}} \mathcal{A}_{sn, x_E} dx_E = \mathcal{A} \int_{\widehat{x}_E + I_{stim}}^{\widehat{x}_E + \alpha I_{stim}} dx_E = \mathcal{A}(\alpha - 1)I_{stim} = (\alpha - 1) \underbrace{\int_{\widehat{x}_E}^{\widehat{x}_E + I_{stim}} \mathcal{A}_{sn, x_E} dx_E}_{\Delta sn} \quad (13.15)$$

Nous avons vu, cependant, au chapitre 9 que la sensibilité de  $sn^*$  au forçage  $x_E$ , notée  $\mathcal{A}_{sn, x_E}$ , dépend de la valeur de  $x_E$  (cf figure 7.A. de l'article) et l'intégrande n'est pas constant. Si une variation du forçage de  $I_{stim}$  produit une variation de l'activité  $sn^*$  qui vaut  $\Delta sn^*$ , une variation du forçage de  $\alpha I_{stim}$  ne conduit pas nécessairement à une variation d'activité  $\alpha \Delta sn^*$ .

Dans la métaphore de la balance, le déplacement du plateau n'est pas proportionnel au poids ajouté.

D'autre part,  $\mathcal{A}_{sn, x_E}$  dépend de la valeur de tous les paramètres de la région. Pour une même variation du forçage  $\Delta\widehat{x}_E = I_{stim}$ , la variation  $\Delta sn^*$  résultante n'est pas toujours identique. Autrement dit, le déplacement des plateaux de la balance dû à l'ajout du poids excitateur  $I_{stim}$  ne produit pas toujours le même déplacement des plateaux de la balance : du fait des non linéarités, la réactivité de la balance à l'ajout d'un même poids  $I_{stim}$  dépend de la valeur des autres paramètres.

### 13.1.4.1.2 Déplacements identiques du plateau E et du plateau I ?

Comme pour le plateau E, nous ajoutons l'idée qu'à un niveau du plateau I correspond un niveau d'activité du pool inhibiteur de la région, i.e.  $sg^*$  (cf figure 13.8). Quand la balance bascule, les deux plateaux étant solidaires, le plateau I se déplace autant que le plateau E, mais dans le sens opposé.

Dans le modèle neural, si nous perturbons un paramètre  $\pi$ , nous pouvons regarder l'effet produit sur  $sn^*$  et sur  $sg^*$ . Les sensibilités de ces variables d'état par rapport à  $\pi$  ne sont pas nécessairement égales et de signes opposés. C'est ce que nous montrons pour  $\pi = x_E$  dans les annexes de l'article au chapitre 9 (une région isolée, section 2.4) :

$$\mathcal{A}_{sn, x_E} = \mathcal{A}_{sg, x_E} \times J_{nmda} \mathcal{A}_{sg, x_I}^O \quad (13.16)$$

En variation, puisqu'en plus les sensibilités varient sur l'intervalle d'intégration, il n'y a pas de raison pour que les sensibilités soient égales :  $\Delta sn^* \neq -\Delta sg^*$ . Si ces variations représentent les déplacements respectifs des plateaux E et I, ce résultat est contraire à la prédiction faite avec la métaphore de la balance E/I. Les plateaux ne se déplacent pas d'une même quantité.

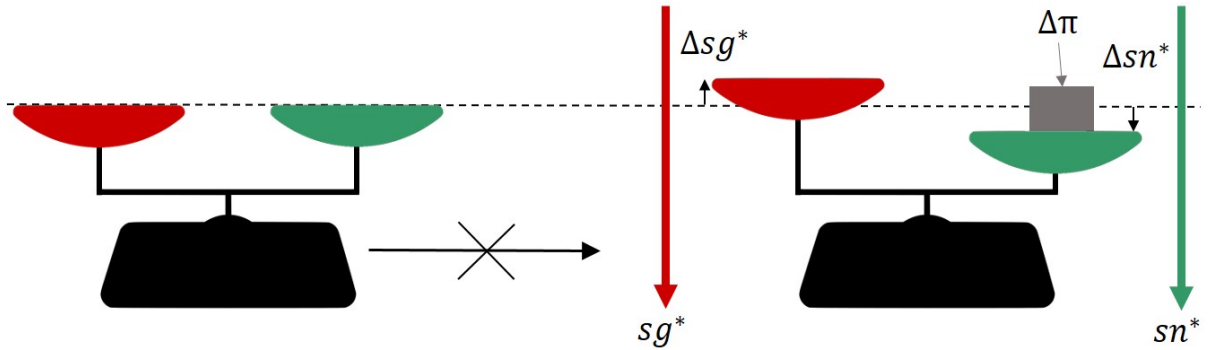


**Figure 13.8** – Le déplacement des plateaux dû à l'ajout d'un poids n'est pas de même longueur : la variation  $\Delta\pi$  ne produit pas la même variation d'activité sur les pools excitateur et inhibiteur d'une région isolée ( $\Delta sn^* \neq -\Delta sg^*$ ). Même remarque qu'à la figure 13.7 pour l'état initial de la balance.

### 13.1.4.2 Effet de la variation de plusieurs paramètres : penser la compensation

Dans le cas d'une balance de Roberval sur laquelle nous ajouterions deux masses identiques, une sur chaque plateau (e.g. une masse de 1 g), le déplacement total des plateaux serait nul : celui dû à l'ajout de la seconde masse compenserait exactement celui dû à la première. Nous détaillons dans ce paragraphe les conditions pour que la variation d'un paramètre à effet inhibiteur (ou exciteur) sur  $sn^*$  compense exactement l'effet d'un paramètre à effet exciteur (ou inhibiteur). Nous voyons qu'il ne suffit pas d'ajouter une variation identique des paramètres, comme dans le cas des masses de la balance. Il est nécessaire de prendre en

compte les sensibilités sans lesquelles nous ne pouvons prédire le déplacement des plateaux.

Nous partons d'une situation d'équilibre connue avec  $sn^* = f(\widehat{\pi}_E, \widehat{\pi}_I)$ . Les paramètres  $\widehat{\pi}_E$  et  $\widehat{\pi}_I$  sont connus et fixent le niveau des plateaux de la balance.

#### 13.1.4.2.1 En perturbation

La linéarisation autour du point fixe permet de séparer les contributions de l'effet de chaque perturbation prise séparément :

$$\delta sn = \mathcal{A}_{sn, \pi_E} \delta \pi_E + \mathcal{A}_{sn, \pi_I} \delta \pi_I \quad (13.17)$$

$$= S_{\pi_E}(\pi_E, \pi_I) \delta \pi_E + S_{\pi_I}(\pi_E, \pi_I) \delta \pi_I \quad (13.18)$$

Nous souhaitons annuler la perturbation de la variable d'observation :

$$\delta sn^* = 0 \iff \mathcal{A}_{sn, \pi_E} \delta \pi_E = -\mathcal{A}_{sn, \pi_I} \delta \pi_I \quad (13.19)$$

Si  $\delta \pi_I \equiv \delta \pi_E$ , alors  $\mathcal{A}_{sn, \pi_E} = -\mathcal{A}_{sn, \pi_I}$ . En perturbation, il est nécessaire que les sensibilités soit égales en valeur absolue et de signes opposés pour annuler  $\delta sn^*$ . Ceci n'est pas toujours vérifié puisque leur valeur dépend des paramètres  $\pi_E$  et  $\pi_I$  que l'on perturbe et des valeurs des paramètres auxquelles elles sont évaluées.

Il n'est donc pas suffisant de compenser exactement les perturbations pour annuler  $\delta sn^*$ . La sensibilité pondère l'effet de la perturbation du paramètre. Il est nécessaire de prendre en compte les sensibilités pour connaître  $\delta sn^*$  qui donne le déplacement du plateau E. Dans la métaphore de la balance, il faudrait penser le poids ajouté comme le produit de la perturbation par la sensibilité. Si on ajoute l'idée que cette dernière dépend des autres paramètres, alors l'aspect intuitif attaché à la balance de Roberval devient source de confusion.

#### 13.1.4.2.2 En variation

Considérons une variation du paramètre exciteur et du paramètre inhibiteur<sup>8</sup> respectivement entre  $\widehat{\pi}_E$  et  $\widehat{\pi}_E + \Delta \widehat{\pi}_E$  et entre  $\widehat{\pi}_I$  et  $\widehat{\pi}_I + \Delta \widehat{\pi}_I$ . Cela correspond à un ajout de poids sur le plateau exciteur (de  $\Delta \widehat{\pi}_E$ ) et sur le plateau inhibiteur (de  $\Delta \widehat{\pi}_I$ ).

Nous supposons de plus que ces poids sont identiques :  $\Delta \widehat{\pi}_I \equiv \Delta \widehat{\pi}_E$ .

Peu importe l'ordre dans lequel les variations de paramètres sont appliquées, nous allons voir que la variation d'activité  $sn^*$  produite par le second paramètre ne compense pas nécessairement celle due au premier, i.e. ne conduit pas nécessairement à annuler la variation totale d'activité  $sn^*$  ( $\Delta sn^* \neq 0$ ). Autrement dit, peu importe l'ordre dans lequel les poids sont positionnés sur les plateaux, l'ajout de ce deuxième poids ne permet pas de revenir à la position initiale des plateaux, comme l'illustre la figure 13.9.

---

8. abus de langage : à effet exciteur/inhibiteur sur  $sn^*$  quand on considère une région isolée

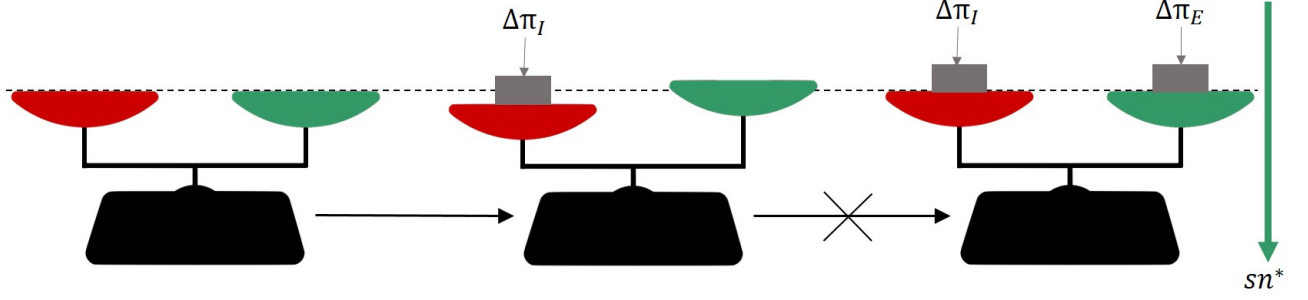


**Figure 13.9** – L'ajout de deux masses identiques de part et d'autre de la balance n'annule pas le déplacement des plateaux : la variation totale de l'activité  $sn^*$  n'est pas nulle ( $\Delta sn^* \neq 0$ ).  
Même remarque qu'à la figure 13.7 pour l'état initial de la balance.

Nous avons vu à la section 13.1.2.2 que la variation totale de l'activité excitatrice  $sn^*$  de la région s'écrit :

$$\Delta sn = \Delta sn_{/\pi_E}(\widehat{\pi_I}) + \Delta sn_{/\pi_I}(\widehat{\pi_E} + \Delta \widehat{\pi_E}) \quad (13.20)$$

$$= \Delta sn_{/\pi_I}(\widehat{\pi_E}) + \Delta sn_{/\pi_E}(\widehat{\pi_I} + \Delta \widehat{\pi_I}) \quad (13.21)$$

Pour que  $\Delta sn = 0$ , nous montrons en annexe 18.1.2.2 que les sensibilités doivent être constantes sur tout l'intervalle d'intégration ; et de plus identiques en valeurs absolues mais de signe opposé.

Cependant, il n'y a pas de raison que ces conditions soient respectées. En effet, leur valeur dépend des paramètres que l'on perturbe et des valeurs des paramètres auxquelles on les évalue.

Nous avons particulièrement travaillé avec le forçage sur le pool excitateur  $x_E$  et le couplage du pool inhibiteur vers le pool excitateur  $J_{gaba}$ . Nous avons écrit (cf section 7.3.2 ou annexe 18.1.1) :

$$\mathcal{A}_{sn, J_{gaba}} = \mathcal{A}_{sn, x_E} \times (-sg^*) \quad (13.22)$$

Les deux sensibilités  $\mathcal{A}_{sn, J_{gaba}}$  et  $\mathcal{A}_{sn, x_E}$  sont bien de signe opposé mais ne sont pas égales en valeur absolue. De plus, elles dépendent toutes les deux des valeurs de  $x_E$  et  $J_{gaba}$  : elles ne sont pas constantes.

La métaphore de la balance favorise donc un raisonnement intuitif de linéarité<sup>9</sup> pour prévoir le comportement du modèle, qui n'est pas toujours valide. Il n'est pas suffisant de compenser exactement les variations de paramètres pour annuler le déplacement des plateaux, en particulier du plateau E. Il est nécessaire de prendre en compte les sensibilités pour connaître la variation  $\Delta sn^*$  qui donne le déplacement du plateau E.

9. Linéaire au sens algébrique : application linéaire de 2 variables

### 13.1.5 Poids synaptiques GABAergique et glutamatergique : quels effets sur la balance E/I ?

La métaphore de la balance E/I est couramment employée dans la communauté des neurosciences pour aider à penser l'effet d'une modification des poids synaptiques excitateurs ou inhibiteurs sur l'activité cérébrale (d'une région). Nous l'avons par exemple employée au chapitre 3 pour comprendre chez des animaux les effets d'une levée de l'inhibition tonique dans l'hippocampe sur les apprentissages hippocampo-dépendants.

Lors de séances de travail avec un médecin psychiatre addictologue de mon entourage scientifique, nous avons commencé par employer de manière équivalente les expressions "augmentation du poids GABAergique" ou "augmentation de l'inhibition dans une région" (inhibition au sens de la capacité d'un neurone pré-synaptique à hyperpolariser un neurone post-synaptique dans le corpus de la neurophysiologie, cf chapitre 3). Ces expressions se traduisent mathématiquement par une augmentation d'un paramètre représentant un poids synaptique GABAergique dans une région. L'augmentation de  $J_{gaba_c}$  a d'ailleurs été présentée comme une augmentation de l'inhibition dans la cible au 12.2.4.

Il en est de même pour les expressions "augmentation du poids glutamatergique" et "augmentation de l'excitation dans une région" (excitation au sens de la capacité d'un neurone pré-synaptique à dépolariser un neurone post-synaptique dans le corpus de la neurophysiologie, cf chapitre 3), se traduisant par l'augmentation d'un paramètre représentant un poids synaptique glutamatergique dans une région.

Pour pouvoir complètement traduire ces expressions dans le cadre du modèle neural de la thèse, il a été nécessaire de préciser entre nous de quels poids synaptiques nous parlions (cf figure 13.10).

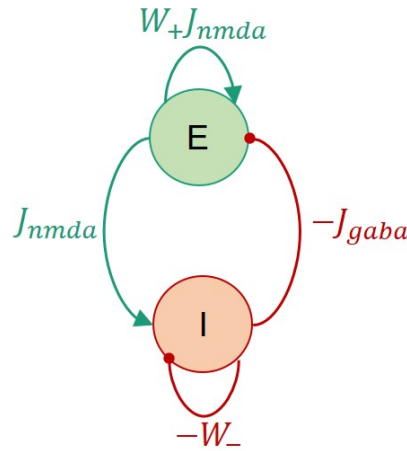


**Figure 13.10** – Représentation des poids synaptiques qui figurent dans l'expression des courants du modèle neural proposé par [Naskar et al., 2021].

### 13.1.5.1 Poids GABAergiques

Dans le modèle, il existe deux poids synaptiques GABAergiques de natures différentes :

- le poids synaptique entre inter-neurones inhibiteurs
- le poids synaptique des inter-neurones inhibiteurs vers les neurones excitateurs

Ils sont respectivement traduits mathématiquement par :

- l'auto-couplage du pool inhibiteur sur lui-même ( $W_-$ ).
- le couplage du pool inhibiteur vers le pool exciteur ( $J_{gaba}$ )

L'effet d'une perturbation de chacun de ces paramètres sur l'activité  $sn^*$  d'une région isolée n'est pas le même (cf annexes 18.1.1) :

- pour  $W_-$ , nous montrons un effet exciteur sur l'activité  $sn^*$  de la région : augmenter  $W_-$  conduit à augmenter l'activité de la région.
- pour  $J_{gaba}$ , nous montrons un effet inhibiteur sur l'activité  $sn^*$  de la région : augmenter  $J_{gaba}$  conduit à diminuer l'activité de la région.

Nous voyons alors qu'augmenter un poids synaptique GABAergique<sup>10</sup> ne mène pas nécessairement à une inhibition de la région (i.e. à une baisse de son activité).

Par ailleurs, ces deux poids synaptiques sont modélisés de manière indépendante :  $W_-$  et  $J_{gaba}$  ne dépendent pas l'un de l'autre. Nous pouvons nous interroger sur la pertinence de cette hypothèse de modélisation quand on travaille en perturbation. Est-ce qu'il existe des mécanismes biologiques qui permettent de perturber/modifier sélectivement les poids synaptiques entre interneurones inhibiteurs, ou entre neurones inhibiteurs et excitateurs, mais pas les deux en même temps ? Par exemple, est-ce qu'une augmentation de la concentration en GABA dans une région peut n'atteindre que certains types de synapses ?

### 13.1.5.2 Poids glutamatergiques

Dans le modèle, il existe deux poids synaptiques glutamatergiques de nature différente :

- le poids synaptique entre neurones excitateurs
- le poids synaptique des neurones excitateurs vers les inter-neurones inhibiteurs

Ils sont respectivement traduits mathématiquement par :

- l'auto-couplage du pool exciteur sur lui-même ( $W_+ J_{nmda}$ ).
- le couplage du pool exciteur vers le pool inhibiteur ( $J_{nmda}$ )

Dans le modèle, ces deux couplages sont proportionnels. Nous voyons qu'au contraire des poids synaptiques GABAergiques, ils ne sont pas modélisés de manière indépendante. La perturbation de  $W_+$  conduit à une perturbation uniquement de l'autocouplage exciteur alors que la perturbation de  $J_{nmda}$  conduit à perturber à la fois l'autocouplage exciteur et le couplage de E vers I.

---

10. i.e. inhibiteur dans le corpus de la neurophysiologie au sens "qui hyperpolarise un neurone post-synaptique"



L'effet d'une perturbation de chacun de ces paramètres sur l'activité  $sn^*$  d'une région isolée a été étudié en annexe 18.1.1 :

- La perturbation de  $W_+$  conduit à une augmentation de  $sn^*$  : effet excitateur sur  $sn^*$  dans une région
- La perturbation de  $J_{nmda}$  peut conduire à une augmentation (effet excitateur) ou une diminution (effet inhibiteur) de  $sn^*$  en fonction du signe de la somme  $\mathcal{A}_{sn,x_E}W_+ + \mathcal{A}_{sn,x_I}$  :

$$\delta sn^* = \underbrace{(\mathcal{A}_{sn,x_E}W_+ + \mathcal{A}_{sn,x_I})}_{>0} \underbrace{sn^*}_{>0} \delta J_{nmda} \quad (13.23)$$

La variation  $\Delta sn^*_{J_{nmda}}$  dépendra du signe et des valeurs de l'intégrande  $\delta sn^*_{J_{nmda}}$  sur l'intervalle d'intégration de  $J_{nmda}$ .

Supposons que l'autocouplage E et le couplage de E vers I ne dépendent pas l'un de l'autre. Nous les notons uniquement  $W_+$  pour le premier, et toujours  $J_{nmda}$  pour le second. D'après le travail analytique mené en annexe 18.1.1 sur les sensibilités dans une région isolée :

- une perturbation uniquement de l'auto-couplage E (via  $W_+$ ) conduit à une augmentation de  $sn^*$  : effet excitateur sur  $sn^*$  dans une région
- une perturbation uniquement du couplage de E vers I (cf dernière remarque après le tableau dans l'annexe 18.1.1) conduit à une diminution de  $sn^*$  : effet inhibiteur sur  $sn^*$  dans une région.

Si l'autocouplage E et le couplage de E vers I ne dépendent pas l'un de l'autre, nous voyons qu'augmenter un poids synaptique glutamatergique<sup>11</sup> ne mène pas nécessairement à une excitation de la région (i.e. à une augmentation de son activité).

### 13.1.6 Bilan

Dans cette section, j'ai donc tenté d'éclairer, avec la plus grande précision de vocabulaire possible, les points concernant :

1. l'effet excitateur ou inhibiteur d'un paramètre,
2. les liens entre le comportement du modèle et la métaphore de la balance E/I,
3. les limites de cette dernière pour prédire la réponse du modèle à une perturbation / variation d'un paramètre.

J'espère que cet éclairage pourra nourrir le dialogue entre les communautés.

---

11. i.e. excitateur dans le corpus de la neurophysiologie au sens "qui dépolarise un neurone post-synaptique"

## 13.2 Comment définir le contrôle inhibiteur d'une région cible ? Que signifie augmenter l'inhibition d'une région cible ?

Nous avons détaillé à la section 13.1 la notion d'inhibition de l'activité d'une région. Celle-ci est une pensée en région isolée dans laquelle la variation d'un paramètre de la région conduit à abaisser son activité  $sn^*$  (e.g. figure 13.5). Elle diffère de la notion d'inhibition de l'activité d'une région cible dans un contexte de contrôle inhibiteur. Je poursuis donc ici l'exposition de ma mise au clair autour du concept central du contrôle inhibiteur.

Je reviens dans un premier temps sur le concept de *contrôle inhibiteur d'une région cible* en neurosciences et précise comment il a été traduit dans le modèle neural de la thèse. Nous voyons à la section 13.2.1 qu'il désigne le comportement d'un système composé de plusieurs régions, au moins une région de contrôle et au moins une région cible. La stimulation de la région de contrôle conduit à l'abaissement d'activité de la région cible. Nous nous remettons en tête l'importance de considérer la rétroaction de la cible vers la région de contrôle, notamment présente dans le modèle neural de DCM et dans notre modèle neural, et qui peut être oubliée et faciliter des raisonnements causaux pour prédire le comportement du modèle. Les régions de contrôle et les régions cibles ont des influences réciproques : la stimulation de la région de contrôle conduit donc à l'abaissement d'activité de la région cible toutes rétroactions prises en compte. Nous verrons à la section 13.2.2 comment dans notre modèle ont été traduits les mots "contrôle" et "inhibiteur", en reprenant les définitions d'*effet exciteur* et *effet inhibiteur* présentées à la section 13.1.

Dans un second temps, nous définissons ce que signifie *augmenter l'inhibition d'une région cible* et voyons comment cela peut être traduit dans le cadre de notre modèle neural (13.2.3), en partant des prédictions sur les effets des paramètres neuraux étudiés dans la thèse ( $I_{stim}$  dans la région de contrôle et  $J_{gaba_c}$  dans la cible).

Dans l'ensemble de cette section, l'architecture de connectivité E-I est privilégiée pour modéliser le contrôle inhibiteur avec le modèle neural couplé présenté dans la thèse (cf chapitre 8). C'est le système couplé. Le système découplé désigne le système sans retour de la cible vers la région de contrôle (i.e. coefficient de couplage nul de la cible vers la région de contrôle).

### 13.2.1 Contrôle inhibiteur d'une région cible : une pensée en variations qui implique des régions de contrôle et des régions cibles couplées entre elles

#### 13.2.1.1 Implication de régions différentes des cibles : les régions de contrôle

Brunton formule en 1883 dans l'Oxford English Dictionary une définition de l'inhibition comme concept physiologique, qui peut être appliquée à une variété de phénomènes en dehors du champ de la physiologie ([Bari and Robbins, 2013]). L'inhibition désigne *l'arrêt des fonctions d'une structure ou d'un organe par l'action d'une autre structure ou organe*. Quand cette action cesse, la fonction peut se manifester à nouveau, elle n'a pas été abolie (comme dans le cas d'une paralysie par exemple). Cette définition donne une interprétation différente de la baisse ou l'arrêt du niveau d'une fonction qui, avant le 19e siècle, était uniquement compris comme une baisse ou un arrêt de l'excitation, sans aucune intervention d'un processus antagoniste supplémentaire extérieur.

Les mécanismes cérébraux du contrôle inhibiteur décrits notamment dans les sections 2.3 et 5.5.2, et que nous avons modélisés dans la thèse concordent avec cette définition. Ils conduisent à une variation de l'activité des régions cibles (mnésiques par exemple), par l'intervention d'autres régions différentes des cibles : les régions de contrôle. La variation de l'activité des régions cibles n'est pas produite par les cibles elles-mêmes via des mécanismes neuraux internes à chaque région. La communication neurale entre régions de contrôle et régions cibles est centrale.

Plus précisément, nous avons étudié la baisse de l'activité d'une région cible due à l'augmentation de l'activité d'une région de contrôle. Dans le cadre de notre modèle, ces variations se traduisent formellement dans chaque région par une variation de la fraction moyenne de canaux excitateurs ouverts au régime stationnaire :  $\Delta sn_{rc}^* > 0$  pour la région de contrôle et  $\Delta sn_c^* < 0$  pour la région cible. Le contrôle inhibiteur est une pensée en variations par rapport à une situation de référence de l'activité de régions cérébrales qui s'influencent. Nous pouvons alors nous intéresser à la direction des couplages.

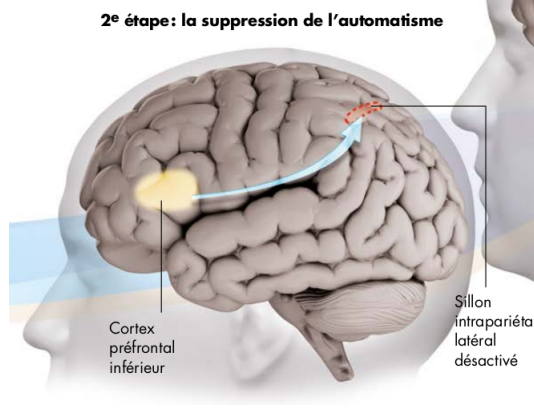
#### 13.2.1.2 Une pensée causale découplée ?

##### 13.2.1.2.1 Régulation top-down : absence de rétroaction

Le contrôle inhibiteur d'une région cible est souvent présenté comme un mécanisme de régulation top-down (cf introduction de la thèse, sections 2.3 et 5.5.2) : la région de contrôle envoie un signal qui se propage jusqu'à la région cible<sup>12</sup>. L'activité de la cible baisse en conséquence par rapport à son état de base : on parle ainsi de *downregulation* ([Levy and Anderson, 2012], [Anderson et al., 2016], [Apšvalka et al., 2022]). Dans cette grille de lecture du phénomène, la rétroaction de la cible vers la région de contrôle n'est pas pensée, elle n'existe pas dans le raisonnement. Le modèle est causal découplé ; une représentation typique est donnée par la figure suivante (tirée de "Les neurones de la pensée libre",

---

12. [Schmitz et al., 2017] (page 6) : "suppressing retrieval involves a goal-related signal that originates in the right DLPFC and spreads downstream, via polysynaptic pathways, to the hippocampus"



Dans le cadre de notre modèle, sans la rétroaction,  $I_{stim}$  produit d'abord une variation d'activité dans la région de contrôle  $\Delta sn_{rc}^* > 0$ . Puis l'activité de la cible baisse en conséquence, ce qui se traduit par  $\Delta sn_c^* < 0$ . On décrit une chaîne de causalité :

$$I_{stim} \Rightarrow \Delta sn_{rc}^* > 0 \Rightarrow \Delta sn_c^* < 0 \quad (13.24)$$

Cette manière dé耦lée de présenter le contrôle inhibiteur au niveau cérébral permet de proposer des hypothèses causales intuitives sur l'effet que produit la variation d'activité de régions de contrôle sur celle de régions cibles (comme nous le verrons en 13.2.2).

Cependant elle ne correspond pas au modèle neural présenté dans la thèse. En effet, celui-ci est plus complexe car il prend en compte la rétroaction de la cible vers la région de contrôle. Avec les rétroactions,  $I_{stim}$  met le système dans un état caractérisé à la fois par une augmentation de l'activité de la région de contrôle  $\Delta sn_{rc}^* > 0$  et une baisse de l'activité de la région cible  $\Delta sn_c^* < 0$ . Nous pourrions écrire :

$$I_{stim} \Rightarrow (\Delta sn_{rc}^* > 0, \Delta sn_c^* < 0) \quad (13.25)$$

### 13.2.1.2.2 Mesures de connectivité effective : rétroactions négligées

Nous pourrions alors nous poser la question de la pertinence d'un modèle avec des rétroactions quand la pensée top-down est dé耦lée.

DCM est une méthode couramment utilisée pour effectuer des mesures de connectivité effective, par exemple sur des données obtenues en IRMf pendant une tâche de contrôle inhibiteur (cf 5.5.2). Le modèle neural dynamique couramment utilisé dans DCM (cf 5.3.1.2) est un modèle avec des rétroactions.

En effet, les couplages intrinsèques<sup>13</sup> entre régions sont portés par la matrice  $\bar{\bar{A}}$  qui est non nulle et symétrique (comme la matrice de connectivité structurale utilisée dans le modèle étudié dans la thèse). Il existe donc des éléments  $A_{ij} = A_{ji} \neq 0$ , c'est-à-dire des coefficients de couplage non nuls (et égaux) d'une région de contrôle  $i$  vers une cible  $j$ , et également de la cible  $j$  vers la région de contrôle  $i$ . Certains de ces couplages peuvent varier en période de stimulation, sans que cette variation ne conduise nécessairement à une annulation de la

13. i.e. en l'absence de toute stimulation

valeur du couplage en question <sup>14</sup>.

Cependant, bien que les rétroactions existent, elles tendent à être négligées. L’emphase est portée sur les effets des régions de contrôle vers les cibles, en questionnant rarement l’importance ou non des rétroactions (i.e. des effets en retour des cibles vers les régions de contrôle). Nous développons ce point dans les deux prochains paragraphes.

Certaines études utilisent des modèles avec une modulation unidirectionnelle des coefficients de couplages (e.g. de la région de contrôle vers la cible uniquement) et semblent ne pas questionner l’importance de la rétroaction portée par les coefficients de couplages intrinsèques.

Au niveau représentationnel d’abord, nous notons que des connexions causales entre une région émettrice et une région réceptrice peuvent schématiser les couplages modulés, par exemple d’une région de contrôle vers une région cible pour une modulation top-down ([Apšvalka et al., 2022], [Leone et al., 2022]). Dans ces schémas, les rétroactions portées par les couplages intrinsèques ne sont pas représentées, elles sont implicites. Ce biais représentationnel peut encourager à les rendre implicites dans la pensée et donc à minimiser leur importance pour comprendre les mécanismes de contrôle inhibiteur.

Dans un modèle avec une modulation unidirectionnelle, l’effet de la variation du couplage sur la baisse d’activité de la cible peut dépendre de la valeur du couplage intrinsèque en retour de la cible vers la région de contrôle. Si ce couplage retour est identique chez tous les participants de l’étude, la comparaison seule des valeurs estimées pour le couplage top-down suffit pour éclairer la question des mécanismes responsables d’une plus ou moins grande baisse d’activité dans la cible, sous l’hypothèse du modèle. S’il ne le sont pas, il semble intéressant de vérifier si la variation d’activité de la cible en fonction du couplage top-down reste bien de plus en plus négative quand le couplage top-down devient de plus en plus négatif, indépendamment de la valeur du couplage retour de la cible vers la région de contrôle.

D’autres études ne semblent pas non plus questionner l’importance de la modulation de la rétroaction, portée par la variation du couplage de la cible vers la région de contrôle, bien qu’elle soit présente dans le modèle. En effet, dans plusieurs études (cf 5.5.2.2), l’hypothèse de modélisation la plus probable, étant donné les mesures BOLD, est celle d’une modulation bidirectionnelle des coefficients de couplage entre une région de contrôle et une région cible, c’est-à-dire une variation du coefficient de couplage de la région de contrôle vers la cible (top-down) et une autre variation du coefficient de couplage de la cible vers la région de contrôle (bottom-up).

Certaines de ces études ne comparent entre participants que les estimations de couplages top-down (e.g. [Benoit and Anderson, 2012], [Benoit et al., 2015], [Gagnepain et al., 2017]), sans rapporter les résultats obtenus pour les estimations de couplages bottom-up. Seuls ces couplages top-down semblent être utilisés pour comparer l’inhibition des cibles entre participants.

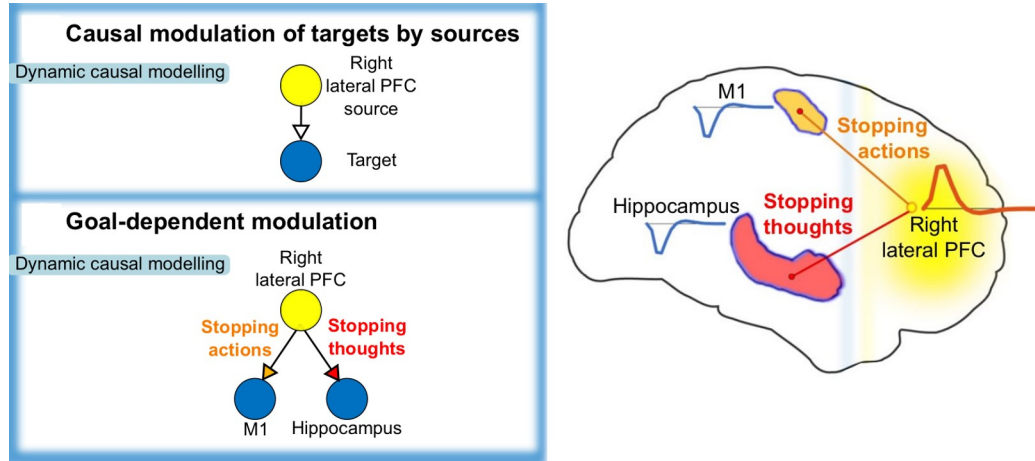
Au contraire, d’autres de ces études comme [Mary et al., 2020] comparent bien les couplages top-down entre participants et les couplages bottom-up entre participants. Néanmoins, ces

---

14. l’amplitude de la variation est estimée à partir de données, cf 5.5.2

analyses statistiques sont menées séparément, de sorte que l'effet du couplage top-down sur la baisse d'activité de la cible n'est pas étudié en prenant en compte la rétroaction. Pourtant, sous l'hypothèse du modèle direct employé, pour un participant, la variation d'activité de la région cible pourrait ne pas devenir plus négative quand le couplage top-down négatif croît en valeur absolue (i.e. devient plus négatif). La baisse d'activité de la cible en fonction du coefficient de couplage top-down pourrait dépendre du coefficient bottom-up, et donc de sa modulation.

En conclusion, bien que les rétroactions existent dans le modèle neural de DCM et qu'elles puissent même varier au cours de la tâche pour modéliser une condition expérimentale, elles semblent négligées lors de l'étude des mécanismes cérébraux de contrôle inhibiteur et elles tendent à devenir un implicite de la pensée. Pour présenter l'étude du contrôle inhibiteur en IRMf, [Apšvalka et al., 2022] résume d'ailleurs le travail avec DCM par le schéma causal suivant :



### 13.2.1.2.3 Un signal "inhibiteur" ?

J'ai de plus pu entendre que "la région de contrôle envoie un signal inhibiteur vers la cible pour abaisser son activité". Dans cette description du phénomène, le signal qui est envoyé contient en soi le résultat.

Dans notre modèle, le "signal envoyé" consiste en une augmentation du forçage sur le pool excitateur de la région de contrôle, d'une valeur de  $I_{stim}$ . Le résultat sur la cible diffère en fonction de la structure de connectivité (cf chapitre 9). L'activité de la cible est abaissée uniquement lorsque la région de contrôle connecte le pool I de la cible, peu importe la nature de la rétroaction (i.e. du pool E de la cible vers le pool E ou I de la région de contrôle). Ceci correspond aux deux architectures de connectivité EI et II. Ce n'est donc pas le signal qui est inhibiteur en soi, mais l'effet qu'il produit du fait de la structure de connectivité, i.e. quand on prend en compte tous les couplages entre régions. De plus, nous avons vu que l'abaissement de l'activité de la cible diffère en amplitude pour un même  $I_{stim}$  entre la structure de connectivité EI ou II. Cela montre l'importance de la rétroaction sur l'effet inhibiteur que produit  $I_{stim}$ .

[Schmitz et al., 2017] parle d'un signal de contrôle inhibiteur, émis depuis le cortex pré-frontal, dans une pensée top-down causale, mais explicite l'importance des interneurons inhibiteurs de la région cible dans l'émergence de sa baisse d'activité : "suppressing unwanted thoughts engages GABAergic interneurons local to the hippocampus in a functionally distinct manner to implement an inhibitory control signal driven by DLPFC" (fin du premier paragraphe de la page 8 de cet article).

### 13.2.2 Modélisation du contrôle inhibiteur dans la thèse

Nous avons présenté dans la thèse un modèle neural du contrôle inhibiteur. Dans le cadre d'une connectivité EI, la variation du forçage  $B_{E_{rc}}$  sur le pool excitateur de la région de contrôle conduit à une augmentation de l'activité  $sn_{rc}^*$  de la région de contrôle et une baisse de l'activité  $sn_c^*$  de la région cible, toutes rétroactions prises en compte.

Dans cette partie, nous reprenons les clés de lecture présentées à la section 13.1 (cf encadrés 13.1.1.1 et 13.1.1.3) et nous nous redonnons tous les objets de l'analyse nécessaires pour préciser le qualificatif "excitateur" ou "inhibiteur" de l'effet d'une augmentation de ce forçage. Nous étudierons d'abord l'effet d'une augmentation du forçage, en perturbation et en variation, sur l'activité de la région cible (13.2.2.1), puis sur celle de la région de contrôle (13.2.2.2). Ceci nous permet de mettre en avant l'importance de toujours préciser la variable d'observation (13.2.2.3). Nous nous remettons alors en tête que dans l'expression "contrôle inhibiteur", la notion de contrôle fait référence au signe positif de la variation d'activité de la région de contrôle alors que le qualificatif "inhibiteur" désigne de manière implicite le signe négatif de la variation d'activité de la région cible (13.2.2.4).

#### 13.2.2.1 Etude de l'effet sur la cible d'une augmentation du forçage sur le pool excitateur de la région de contrôle

##### 13.2.2.1.1 En perturbation

Pour préciser l'effet excitateur ou inhibiteur de la perturbation d'un paramètre, nous proposons de toujours expliciter le triplet de l'encadré 13.1.1.1. Dans le cas qui nous intéresse ici, nous précisons donc :

- le paramètre perturbé ( $\pi$ ) :  $B_{E_{rc}}$  (i.e. le forçage sur le pool excitateur)
- la variable d'observation ( $\eta$ ) : activité excitatrice de la région cible  $sn_c^*$
- la boucle de rétroaction ( $\mathcal{B}$ ) : 2 régions : une région de contrôle et une région cible, couplées entre elles en EI. La boucle de rétroaction est représentée à la figure 13.11.



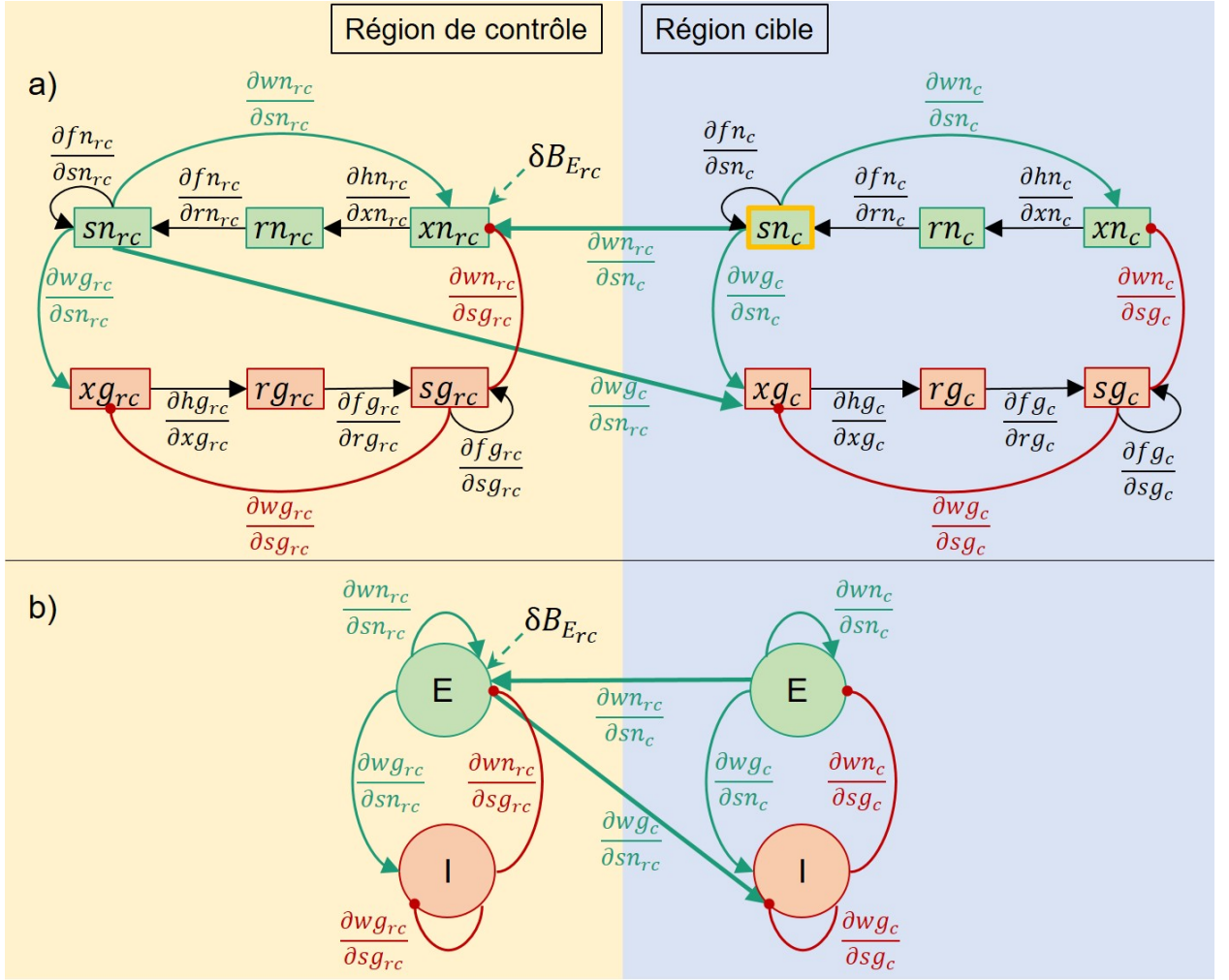


**Figure 13.11** – a) Boucle de rétroaction complète : graphe qui montre les variables du modèle (noeuds) et les dépendances entre ces variables (arêtes). Cette représentation a le mérite de rendre visible la variable d'observation  $sn_c^*$  (en jaune). b) Nous la simplifions pour ne garder uniquement que les couplages entre les pools  $E$  et  $I$ , qui représentent des poids synaptiques. C'est le schéma plus couramment représenté (e.g. [Deco et al., 2014], [Deco and Jirsa, 2012]). Il rend implicite toutes les variables du modèle et en particulier la variable d'observation.  $E$  et  $I$  représentent des sous-graphes.

Nous avons en plus représenté la perturbation du forçage  $B_{Erc}$ .

Au point fixe, la valeur de  $sn_c^*$  dépend de  $B_{Erc}$  et  $J_{gaba_c}$ . Nous noterons donc :

$$sn_c^* = f_c(B_{Erc}, J_{gaba_c}) \quad (13.26)$$

Nous perturbons  $B_{Erc}$ , pour une valeur fixée de  $J_{gaba_c}$ .

Dans un système de 2 régions couplées nous avons vu que l'effet de  $B_{Erc}$  sur l'activité  $sn_c^*$  de la cible dépend du type de connectivité. Pour les différents type de connectivité, nous



avons calculé analytiquement la sensibilité de  $sn_c^*$  par rapport à  $B_{Erc}$  (cf chapitre 9). Elle dépend de  $B_{Erc}$  et  $J_{gaba_c}$  : nous la notons  $S_{sn_c/B_{Erc}}(B_{Erc}, J_{gaba_c})$ . La perturbation de  $sn_c^*$  due à  $\delta B_{Erc}$  s'écrit :

$$\delta sn_{c/B_{Erc}} = S_{sn_c/B_{Erc}}(B_{Erc}, J_{gaba_c})\delta B_{Erc} \quad (13.27)$$

Pour une connexion asymétrique EI en particulier, nous avons vu que  $\delta sn_{c/B_{Erc}} < 0$ , ce qui nous permet de déduire que  $B_{Erc}$  a un effet inhibiteur infinitésimal sur l'activité  $sn_c^*$  de la cible au régime stationnaire.

### 13.2.2.1.2 En variation

Nous partons de  $sn_c^*$  évalué en  $B_{Erc} = \widehat{B_{Erc}}$  et  $J_{gaba_c} = \widehat{J_{gaba_c}}$ .

Pour préciser l'effet excitateur ou inhibiteur fini de la variation de  $B_{Erc}$ , nous nous donnons les éléments du quadruplet de l'encadré 13.1.1.3 :

- le paramètre perturbé ( $\pi$ ) :  $B_{Erc}$  (i.e. le forçage sur le pool excitateur)
- intervalle de variation : nous faisons varier le forçage  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$ . Nous notons la variation du forçage  $\Delta\widehat{B_{Erc}} \equiv I_{stim}$ .
- la variable d'observation ( $\eta$ ) : activité excitatrice de la région cible  $sn_c^*$
- la boucle de rétroaction ( $\mathcal{B}$ ) : 2 régions : une région de contrôle et une région cible, couplées entre elles en EI. La boucle de rétroaction est représentée à la figure 13.11.

Pour connaître le signe de la variation de la variable d'observation  $sn_c^*$ , nous l'écrivons comme une somme de perturbations :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) \equiv \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}} \delta sn_{c/B_{Erc}} \quad (13.28)$$

$$= \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}} S_{sn_c/B_{Erc}}(B_{Erc}, \widehat{J_{gaba_c}}) dB_{Erc} \quad (13.29)$$

$$\equiv f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (13.30)$$

$\Delta sn_{c/B_{Erc}}$  est la fonction qui donne la variation de l'activité  $sn_c^*$  de la cible due à cette variation fixée du forçage  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$  dans la région de contrôle. Elle dépend de la valeur fixée de  $J_{gaba_c}$  dans la cible, c'est-à-dire d'un poids de la balance E/I dans la cible.

Comme nous l'avons montré dans l'article du chapitre 9, le signe de  $\delta sn_{c/B_{Erc}}$  dépend de l'architecture de connectivité, et donc également celui de  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ .

Dans le cas d'une connectivité EI,  $\delta sn_{c/B_{Erc}} < 0$  sur tout l'intervalle d'intégration, donc  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) < 0$ . L'application de  $I_{stim}$  dans la région de contrôle a un effet inhibiteur fini sur l'activité de la cible  $sn_c$ .

### 13.2.2.2 Etude de l'effet sur la région de contrôle d'une augmentation du forçage sur le pool excitateur de la région de contrôle

Nous reprenons exactement les mêmes étapes de raisonnement que dans la partie précédente, à la seule différence que nous changeons de variable d'observation( $\eta$ ) : activité excitatrice de la région de contrôle  $sn_{rc}^*$ .

#### 13.2.2.2.1 En perturbation

Au point fixe, la valeur de  $sn_{rc}^*$  dépend de  $B_{E_{rc}}$  et  $J_{gaba_c}$ . Nous noterons donc :

$$sn_{rc}^* = f_{rc}(B_{E_{rc}}, J_{gaba_c}) \quad (13.31)$$

Nous perturbons  $B_{E_{rc}}$ , pour une valeur fixée de  $J_{gaba_c}$ .

Nous avons calculé analytiquement la sensibilité de  $sn_{rc}^*$  par rapport à  $B_{E_{rc}}$ , qui dépend de  $B_{E_{rc}}$  et  $J_{gaba_c}$ . Nous la notons  $S_{sn_{rc}/B_{E_{rc}}}(B_{E_{rc}}, J_{gaba_c})$ . La perturbation de  $sn_{rc}^*$  due à  $\delta B_{E_{rc}}$  est donnée par :

$$\delta sn_{rc}/B_{E_{rc}} = S_{sn_{rc}/B_{E_{rc}}}(B_{E_{rc}}, J_{gaba_c})\delta B_{E_{rc}} \quad (13.32)$$

Pour une connexion EI, nous avons vu que  $\delta sn_{rc}/B_{E_{rc}} > 0$ , ce qui nous permet de déduire que  $B_{E_{rc}}$  a un effet excitateur infinitésimal sur l'activité  $sn_{rc}^*$  de la région de contrôle au régime stationnaire.

#### 13.2.2.2.2 En variation

Nous écrivons la variation de la variable d'observation comme une somme de perturbations :

$$\Delta sn_{rc}/B_{E_{rc}}(\widehat{J_{gaba_c}}) \equiv \int_{\widehat{B_{E_{rc}}}}^{\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}}} \delta sn_{rc}/B_{E_{rc}} \quad (13.33)$$

$$= \int_{\widehat{B_{E_{rc}}}}^{\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}}} S_{sn_{rc}/B_{E_{rc}}}(B_{E_{rc}}, \widehat{J_{gaba_c}}) dB_{E_{rc}} \quad (13.34)$$

$$\equiv f_{rc}(\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}}, \widehat{J_{gaba_c}}) - f_{rc}(\widehat{B_{E_{rc}}}, \widehat{J_{gaba_c}}) \quad (13.35)$$

$\Delta sn_{rc}/B_{E_{rc}}$  est la fonction qui donne la variation de l'activité  $sn_{rc}^*$  de la région de contrôle due à cette variation fixée du forçage  $B_{E_{rc}}$  entre  $\widehat{B_{E_{rc}}}$  et  $\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}}$  dans la région de contrôle. Elle dépend de la valeur fixée de  $J_{gaba_c}$ , c'est-à-dire d'un poids de la balance E/I dans la cible. Nous notons que cette propriété est contre-intuitive dans une pensée top-down, c'est-à-dire si on ne pense plus à l'existence du retour de la cible vers la région de contrôle. En effet, l'activité de la région de contrôle ne dépend pas de  $J_{gaba_c}$  dans la cible s'il n'y a pas de retour de la cible vers la région de contrôle.

Dans le cas d'une connectivité EI, nous avons montré dans l'article du chapitre 9 que  $\delta sn_{rc}/B_{E_{rc}} > 0$  sur tout l'intervalle d'intégration, donc  $\Delta sn_{rc}/B_{E_{rc}}(\widehat{J_{gaba_c}}) > 0$ . L'application de  $I_{stim}$  dans la région de contrôle a un effet excitateur fini sur l'activité de la cible  $sn_{rc}$ .

### 13.2.2.3 Conclusion : importance de préciser la variable d'observation pour caractériser l'effet excitateur ou inhibiteur d'un paramètre

Dans le cas d'une architecture de connectivité de type EI, nous avons montré :

- un effet inhibiteur (infinitésimal et fini) de  $B_{E_{rc}}$  si la variable d'observation est l'activité de la cible au régime stationnaire ( $sn_c^*$ )
- un effet excitateur (infinitésimal et fini) de  $B_{E_{rc}}$  si la variable d'observation est l'activité région de contrôle au régime stationnaire ( $sn_{rc}^*$ ).

#### Note sur l'importance de préciser la variable d'observation

Pour une même boucle de rétroaction, le qualificatif "excitateur" ou "inhibiteur" dépend donc de la variable d'observation. Nous ne pouvons l'utiliser comme qualificatif du paramètre en soi. J'ai souvent entendu et moi-même utilisé l'expression "forçage excitateur" pour désigner le paramètre  $B_{E_{rc}}$ . Cette expression ne fait pas de sens, même comme raccourci de langage pour "effet excitateur du paramètre", si on ne précise pas la variable d'observation. Nous prendrons donc soin de ne pas l'utiliser dans la suite de l'exposé et de toujours rendre explicite la variable d'observation, c'est-à-dire "sur quoi l'augmentation du paramètre produit un effet excitateur ou inhibiteur".

### 13.2.2.4 Notions de *contrôle* et *contrôle inhibiteur*

La région dite de contrôle subit une variation du forçage sur son pool excitateur, d'une valeur de  $I_{stim}$ . Le contrôle peut être défini par l'augmentation résultante de son activité  $sn_{rc}^*$ , i.e.  $\Delta sn_{rc}^* > 0$ . Il peut être décrit par la chaîne de causalité suivante :

$$\boxed{\text{Contrôle : } I_{stim} \Rightarrow \Delta sn_{rc}^* > 0} \quad (13.36)$$

Cette région de contrôle est couplée à une région cible. Le qualificatif 'inhibiteur' fait référence au signe de la variation d'activité de la cible, due à la variation du forçage de la région de contrôle :  $\Delta sn_c^* < 0$ .

Le contrôle inhibiteur est modélisé par une variation positive du forçage  $B_{E_{rc}}$  dans la région de contrôle qui produit une augmentation d'activité de la région de contrôle et une baisse d'activité de la région cible. En prenant en compte la rétroaction de la région cible vers la région de contrôle, nous pouvons définir le contrôle inhibiteur par la chaîne de causalité suivante :

$$\boxed{\text{Contrôle Inhibiteur : } I_{stim} \Rightarrow (\Delta sn_{rc}^* > 0, \Delta sn_c^* < 0)} \quad (13.37)$$

Dans ce contexte, augmenter le contrôle inhibiteur ou l'inhibition de la cible signifie augmenter la baisse de l'activité de la région cible<sup>15</sup>. Cela se traduit formellement par une variation  $\Delta sn_c^* < 0$  plus grande en valeur absolue. Nous pouvons alors nous demander *comment augmenter l'inhibition de la cible dans le contexte du contrôle inhibiteur, c'est-à-dire comment augmenter la baisse de l'activité de la région cible quand on stimule la région de contrôle ?*

### 13.2.3 Comment augmenter l'inhibition de l'activité d'une région cible ?

Nous avons étudié l'effet de deux paramètres sur la baisse d'activité de la région cible dans un contexte de contrôle inhibiteur (i.e. sous l'effet de la stimulation d'une région de contrôle) :

- l'amplitude  $I_{stim}$  de la variation du forçage sur le pool exciteur de la région de contrôle
- le coefficient de couplage  $J_{gaba_c}$  du pool inhibiteur vers le pool exciteur de la région cible.

Dans la section 13.2.3.1, nous réexposons la prédiction sur l'effet d'une augmentation de  $I_{stim}$ , qui avait été donnée dans l'introduction de la thèse 7.3 : augmenter  $I_{stim}$  permet d'abaisser davantage l'activité de la cible. Nous la rationalisons avec la formulation intégrale de la variation d'activité en sensibilité. Puis nous soulignons la limite d'un raisonnement linéaire qui tenterait de prédire quantitativement la baisse d'activité de la cible en fonction de l'amplitude de  $I_{stim}$ .

Dans la section 13.2.3.2, je présente à nouveau la prédiction sur l'effet d'une augmentation de  $J_{gaba_c}$ , qui se trouve en 7.3 : augmenter  $J_{gaba_c}$  permet d'abaisser davantage l'activité de la cible. Nous revenons sur les étapes de raisonnement permettant de l'énoncer, dans une pensée sans rétroaction de la cible vers la régions de contrôle (c'est-à-dire un modèle top-down du contrôle inhibiteur, cf 13.2.1.2). Sans rétroaction, nous validons cette prédiction. Nous nous redonnons ensuite la rétroaction, et la prédiction reste vraie, avec des différences d'amplitudes. Ces deux études, sans et avec rétroaction, nous permettent de fixer le vocabulaire : nous définissons comment dans le modèle peuvent être traduites l'*inhibition* et la *sur-inhibition* de l'activité de la cible.

Cependant, dans le chapitre 9, nous avons mis en évidence une relation non linéaire entre la baisse d'activité de la cible, et l'augmentation de  $J_{gaba_c}$ . Nous explicitons complètement la variable d'observation qui a servi alors à étudier cet effet et nous montrons qu'elle diffère de la variable d'observation considérée pour formuler la prédiction ci-dessus. Ceci me permet de partager un retour sur l'importance du choix de la variable d'observation en fonction de la question de recherche.

---

15. L'augmentation de l'inhibition de l'activité d'une région cible dans un contexte de contrôle inhibiteur n'est pas à confondre avec l'augmentation de l'inhibition dans une région cible qui est une pensée en région isolée (cf 13.1.5).

### 13.2.3.1 Prediction sur l'effet d'une augmentation de $I_{stim}$ dans la région de contrôle

Une prédiction immédiate de l'effet d'une augmentation de  $I_{stim}$  sur la baisse d'activité de la cible peut être formulée comme suit :

*"Plus  $I_{stim}$  augmente et plus la baisse d'activité dans la cible augmente."*

#### 13.2.3.1.1 Confirmation donnée par l'étude de sensibilité

Supposons que  $J_{gaba_c}$  soit fixé et égal à  $\widehat{J_{gaba_c}}$ . Nous avons vu que  $I_{stim}$  cause une variation d'activité dans la cible égale à  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ .

La variation d'activité de la cible  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$  peut s'écrire comme l'intégrale de la sensibilité de  $sn_c^*$  par rapport à  $B_{Erc}$  :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + I_{stim}} S_{sn_c/B_{Erc}}(B_{Erc}, \widehat{J_{gaba_c}}) dB_{Erc} \quad (13.38)$$

Nous avons vu que  $S_{sn_c/B_{Erc}}$  est négative dans le cas d'une connectivité EI pour toute valeur de  $B_{Erc}$ . Une augmentation de  $I_{stim}$  se traduit mathématiquement par une augmentation de la borne supérieure de l'intégrale. En conséquence,  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$  reste négative et augmente en valeur absolue quand  $I_{stim}$  augmente. Cette augmentation tend à s'amortir quand  $S_{sn_c/B_{Erc}}(B_{Erc}, \widehat{J_{gaba_c}})$  tend vers 0.

Sur la figure 13.12, nous avons représenté sur le graphe A) l'activité  $sn_c^*$  de la cible, en fonction de  $I_{stim}$  (c'est-à-dire pour une valeur de forçage sur le pool excitateur de la région de contrôle égal à  $B_{Erc} = 0.382 + I_{stim}$ ), pour une valeur de  $J_{gaba_c}$  fixée dans la région cible. Le couple  $(J_{gaba_{rc}}, J_{gaba_c})$  assure que  $rn_{rc} = rn_c = 3Hz$ . Il est approximativement égal à (1.758, 0.929). Pour calculer la variation, nous pouvons faire une différence entre  $sn_c^*$  stimulé et non stimulé (cf flèche rouge sur le graphe A, reportée sur le graphe B). La variation d'activité de la cible décroît quand  $I_{stim}$  augmente.

Sur le graphe C, nous avons tracé la sensibilité de  $sn_c^*$  à  $B_{Erc}$  à partir de son expression analytique établie au chapitre 9.  $S_{sn_c/B_{Erc}}(B_{Erc}, \widehat{J_{gaba_c}})$  est évaluée pour une valeur fixée de  $J_{gaba_c} = \widehat{J_{gaba_c}} \approx 0.929$  et pour différentes valeurs du forçage  $B_{Erc} = \widehat{B_{Erc}} + I_{stim} = 0.382 + I_{stim}$ . Par ailleurs, elle est aussi égale à la pente de la tangente à la courbe  $sn_c^* = f_c(I_{stim}, J_{gaba_c})$  (graphe A). Nous vérifions bien que les pentes sont négatives et varient en fonction de  $I_{stim}$ . L'intégration de toutes les sensibilités entre 0 et par exemple  $I_{stim} = 0.06$ , égale à l'opposé de l'aire hachurée, permet de calculer par une deuxième méthode la variation  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ . Elle peut être traduite par une somme de  $\delta sn_c^*$  (cf petites doubles flèches rouges sur le graphe A).

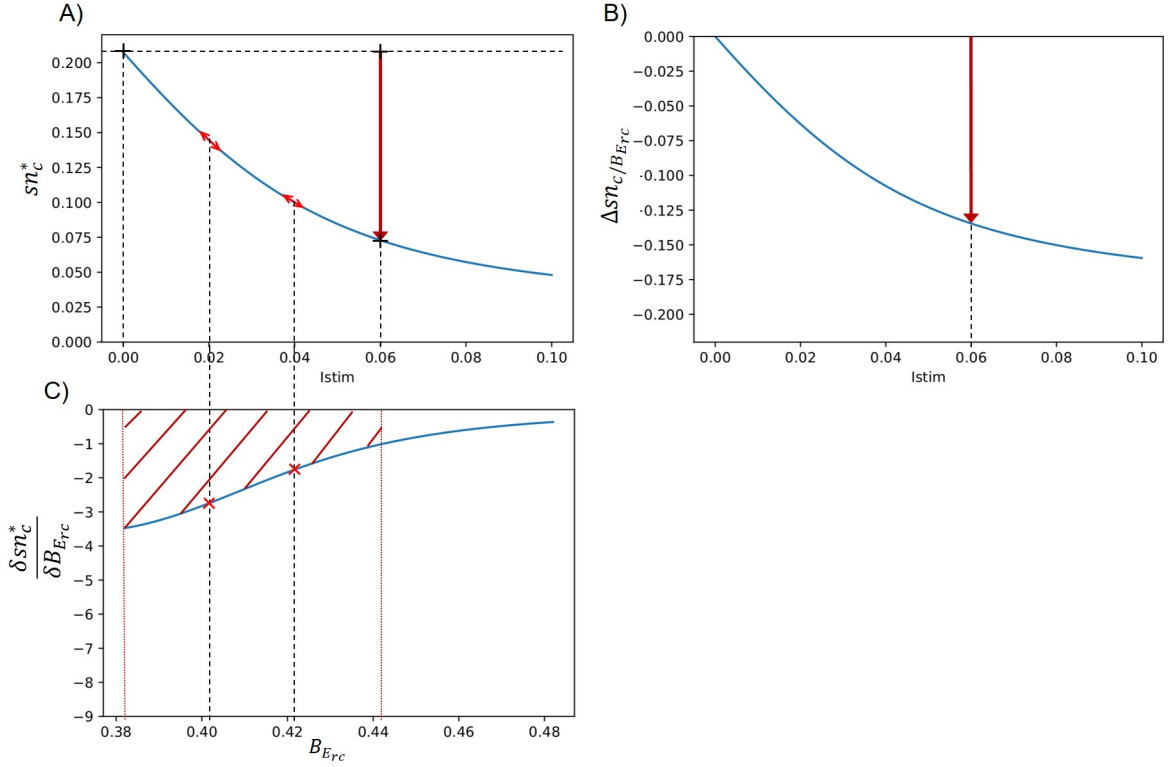


**Figure 13.12** – A : Activité  $sn_c^*$  de la cible en fonction de  $I_{stim}$  sur le pool excitateur de la région de contrôle, pour  $J_{gaba_c}$  fixé. Le forçage sur le pool excitateur de la région de contrôle vaut  $B_{Erc} = 0.382 + I_{stim}$ .  
 B : Variation d'activité de la cible calculée entre la stimulation ( $I_{stim}$ ) et le repos ( $I_{stim} = 0$  i.e.  $B_{Erc} = 0.382$ ), pour  $J_{gaba_c}$  fixé.  
 C : Sensibilité de  $sn_c^*$  par rapport au forçage  $B_{Erc}$ , en fonction de  $B_{Erc}$ , pour  $J_{gaba_c}$  fixé.

Plus l'amplitude  $I_{stim}$  de la variation du forçage sur le pool excitateur de la région de contrôle augmente et plus l'activité de la cible diminue.  
 Autrement dit, pour augmenter l'inhibition de l'activité de la cible, il est possible d'augmenter la stimulation de la région de contrôle.

Nous avons vérifié dans l'annexe 18.2 que sans rétroaction, plus  $I_{stim}$  augmente et plus la baisse d'activité dans la cible augmente. L'ajout de la rétroaction de la cible vers la région de contrôle conserve donc le sens des variations d'activité dans la région de contrôle et dans la cible, mais en modifie les amplitudes.

### 13.2.3.1.2 Hypothèse de linéarité

D'autre part, une autre hypothèse immédiate de l'effet d'une augmentation de la stimulation de la région de contrôle sur la baisse d'activité de la cible pourrait être formulée comme suit :

*"Pour multiplier par  $\alpha$  la baisse d'activité de la cible, nous avons besoin de multiplier par  $\alpha$  la stimulation de la région de contrôle."*

Sur la figure 13.12 B, pour  $I_{stim} = 0.05$  nous avons  $\Delta sn_{c/B_{Erc}} \approx -0.125$ . Si la stimulation est multipliée par 2, i.e.  $I_{stim} = 0.1$ , la variation d'activité n'est pas multipliée par 2 ( $\Delta sn_{c/B_{Erc}} \approx -0.150$ ). Plus généralement, nous constatons que  $\Delta sn_{c/B_{Erc}}$  ne varie pas de manière linéaire avec  $I_{stim}$ , i.e. il n'existe pas de coefficient de proportionnalité entre ces deux grandeurs.

En annexe 18.2, nous expliquons que l'hypothèse de linéarité aurait été vraie si la sensibilité était constante sur tout l'intervalle d'intégration du forçage, ce qui n'est le cas ni ici (cf graphe en bas à gauche), ni dans le cas sans rétroaction (cf annexe 18.2).

Pour abaisser deux fois plus l'activité de la cible, il ne suffit pas de stimuler deux fois plus la région de contrôle.

### 13.2.3.2 Prédiction sur l'effet d'une augmentation de $J_{gaba_c}$ dans la cible

Dans le contexte du contrôle inhibiteur, la stimulation de la région de contrôle produit un abaissement de l'activité de la cible.

Nous nous sommes intéressés à l'effet de la variation d'un paramètre de la balance locale E/I de la région cible sur cette baisse d'activité. Plus précisément, nous avons exploré l'effet d'une variation de  $J_{gaba_c}$  dans la région cible, qui peut s'interpréter comme un poids synaptique moyen des neurones inhibiteurs vers les neurones excitateurs de la région cible.

Pour une région isolée, nous avons vu que  $J_{gaba}$  a un effet inhibiteur sur l'abaissement de l'activité  $sn^*$  : plus  $J_{gaba}$  augmente et plus l'activité de la région baisse (propriété vérifiée, cf section 7.3.2 et 13.1).

#### 13.2.3.2.1 Sans rétroaction de la cible vers la région de contrôle

Dans un système sans rétroaction de la cible vers la région de contrôle, nous présentons le raisonnement causal qui nous a permis de prédire, à partir de cette propriété vérifiée, l'effet sur-inhibiteur d'une augmentation de  $J_{gaba_c}$  sur la baisse d'activité de la cible due à la stimulation de la région de contrôle. Cette prédiction avait été énoncée à la section 7.3.

Nous adoptons la notation  $\eta^{*dec}$  pour désigner une variable d'état  $\eta^*$  dans le système découplé, ce qui signifie ici sans rétroaction de la cible vers la région de contrôle.

##### 13.2.3.2.1.1 Raisonnement causal et usage de la métaphore de la balance E/I

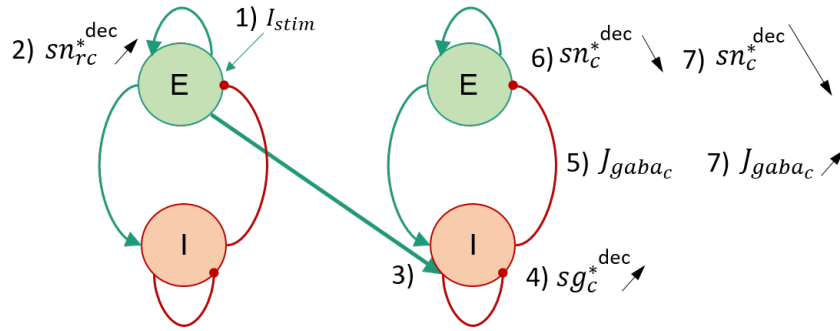


**Figure 13.13** – Schéma représentant les différentes étapes du raisonnement causal qui permet de prédire l'effet sur-inhibiteur de  $J_{gaba_c}$ .

Une manière immédiate de se représenter l'effet d'une augmentation de  $J_{gaba_c}$  sur la baisse d'activité de la cible, peut être décrite à travers le raisonnement causal suivant. A partir du système représenté à la figure 13.13, nous pouvons penser dans un ordre successif :

1. stimulation de la région de contrôle par augmentation du forçage sur son pool excitateur (de  $I_{stim}$ )
2. augmentation de l'activité  $sn_{rc}^{*dec}$  de la région de contrôle (de  $\Delta sn_{rc/B_{Er_c}}^{*dec} > 0$ )



3. augmentation du forçage appliqué sur le pool inhibiteur dans la région cible (de  $\kappa_{c-rc}\Delta sn_{rc/B_{Er c}}^{*\text{dec}}$ )
4. augmentation de l'activité du pool inhibiteur de la région cible  $sg_c^{*\text{dec}}$
5. transmission de l'activité du pool inhibiteur au pool excitateur de la région cible par l'intermédiaire de  $J_{gaba_c}$
6. baisse de l'activité  $sn_c^{*\text{dec}}$  de la cible, avec toutes les rétroactions dans la région cible prises en compte<sup>16</sup> : on parle d'inhibition de la cible.
7. augmentation de  $J_{gaba_c}$  qui produit une diminution<sup>17</sup> supplémentaire de l'activité  $sn_c^{*\text{dec}}$  de la cible : on parle d'augmentation de l'inhibition ou de sur-inhibition.
8. plus  $J_{gaba_c}$  augmente et plus la baisse d'activité de la cible est grande, i.e. plus on augmente l'inhibition de la cible.

### Prédiction de l'effet d'une augmentation de $J_{gaba_c}$

A partir de ce raisonnement, il est alors possible d'émettre une prédiction sur l'effet d'une augmentation de  $J_{gaba_c}$  sur la baisse d'activité de la cible :

*"Plus  $J_{gaba_c}$  dans la cible augmente et plus la baisse d'activité de la cible est grande.*

*Augmenter  $J_{gaba_c}$  produit une sur-inhibition ou augmentation de l'inhibition de la cible (inhibition qui est due à la stimulation de la région de contrôle)."*

**Propriétés du raisonnement :** Ce raisonnement fait appel à une pensée causale dé-couplée dans laquelle la rétroaction de la région de contrôle vers la cible n'est pas présente. Les étapes 1 à 6 décrivent la pensée causale du contrôle inhibiteur présentée à la section 13.2.1.2. Le point 7 considère l'effet d'une augmentation de  $J_{gaba_c}$  sur l'activité de la cible seule, i.e. dé-couplée de la région de contrôle. D'autre part, la pensée est linéaire : la variation d'activité de la cible due à  $I_{stim}$  dans la région de contrôle (6) est sommée à celle produite par une variation de  $J_{gaba_c}$  dans la cible (7).

Nous pouvons représenter ce raisonnement avec la métaphore de la balance E/I de la région cible, comme sur la figure 13.14.

---

16. En effet nous avons vu au chapitre 7 qu'une augmentation du forçage  $x_I$  sur le pool inhibiteur d'une région conduit à une diminution de son activité  $sn^*$ , toutes les rétroactions ayant été prises en compte.

17. En effet, nous avons vu que dans une région,  $J_{gaba}$  a un effet inhibiteur sur l'activité  $sn^*$ .

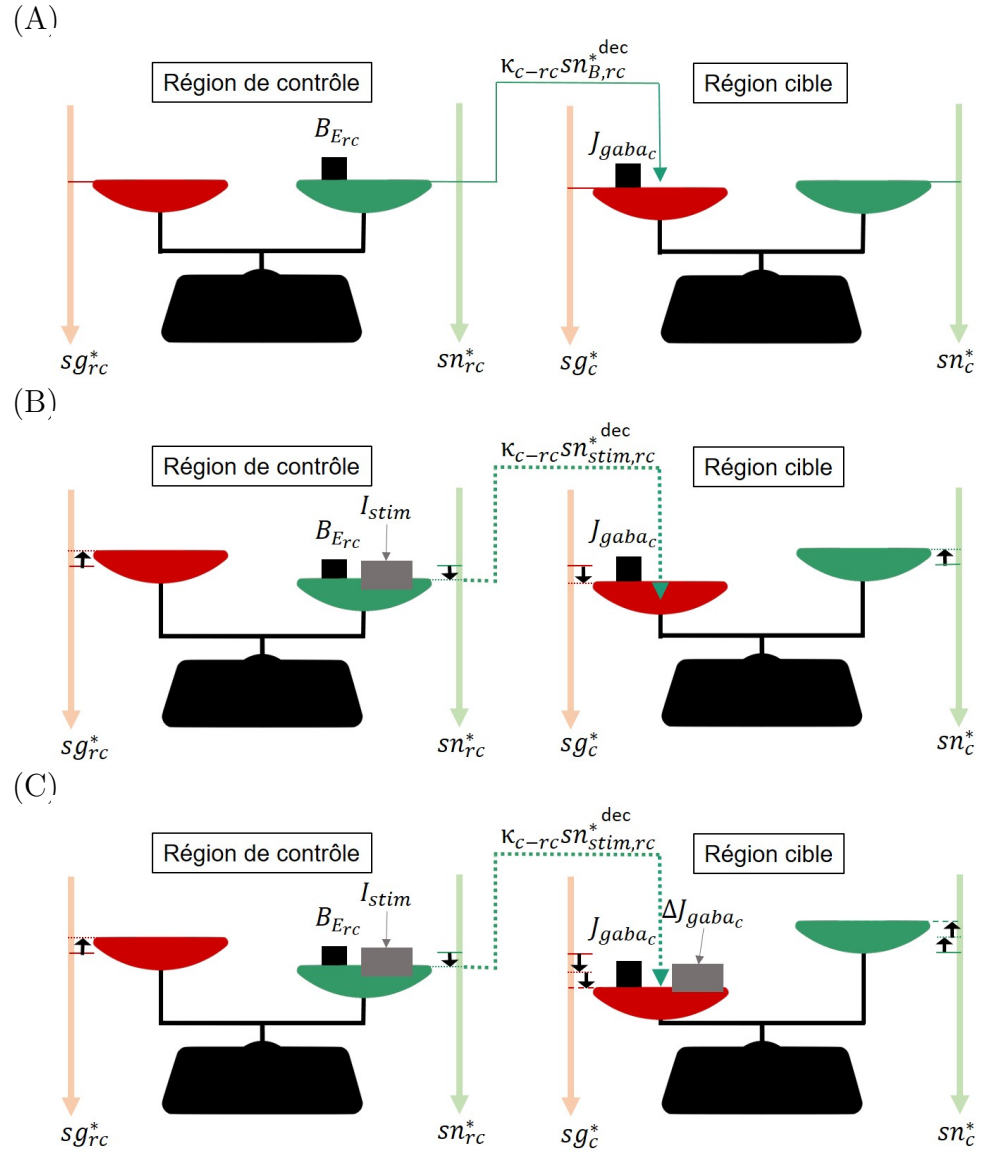


**Figure 13.14** – Représentation du raisonnement sans rétroaction avec la métaphore de la balance. Les poids  $I_{stim}$  et  $\Delta J_{gaba_c}$  sont placés sur les plateaux E ou I en fonction de l'effet excitateur ou inhibiteur qu'ils produisent dans une région (cf figures 13.3 et 13.5).

### 13.2.3.2.1.2 Variations des points fixes : définitions formelles de l'*inhibition* et de la *sur-inhibition*

Nous pouvons vérifier la validité de ce raisonnement causal, découplé et linéaire, avec le graphe 13.15, qui représente l'activité de la région cible  $sn_c^{*dec}$  dans un système sans rétro-action, pour différentes valeurs de  $J_{gaba_c}$  dans la cible et pour différentes valeurs du forçage  $B_{Erc}$  dans la région de contrôle.

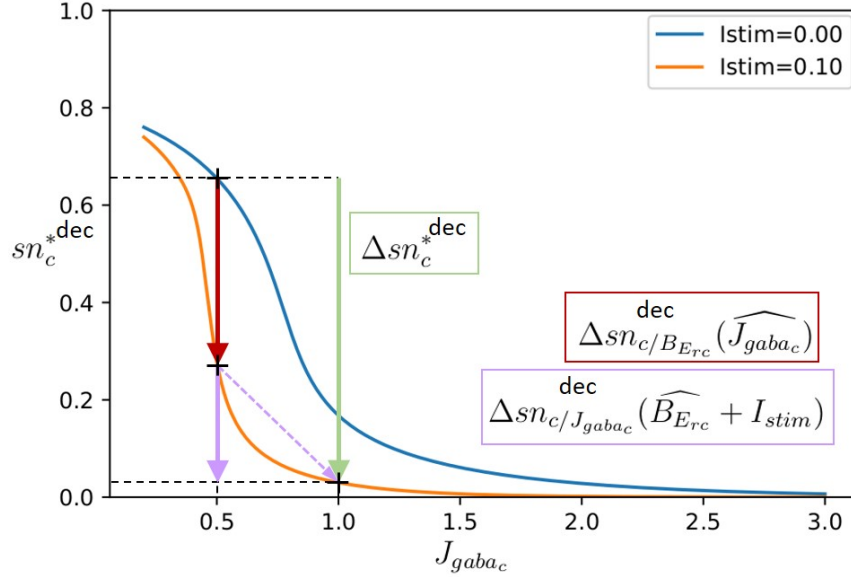


**Figure 13.15** – Sans rétroaction de la cible vers la région de contrôle. Activité  $sn_c^{*dec}$  de la cible en fonction de  $J_{gaba_c}$ , pour deux valeurs fixées du forçage dans la région de contrôle (pour rappel le forçage vaut  $B_{Erc} = 0.382 + I_{stim}$ ). La stimulation de la région de contrôle est modélisée par une variation du forçage, i.e. un passage de  $I_{stim} = 0$  (courbe bleue) à  $I_{stim} = 0.1$  (courbe orange).

Flèche rouge : inhibition (obtenue pour une stimulation de la région de contrôle, à  $J_{gaba_c}$  fixé).

Flèche violette : sur-inhibition (obtenue pour une augmentation supplémentaire de  $J_{gaba_c}$ ). La flèche verte représente la somme des deux autres.

La courbe bleue représente les activités  $sn_c^{*dec}$  au repos ( $I_{stim} = 0$ ) et la courbe orange les activités  $sn_c^{*dec}$  stimulées ( $I_{stim} > 0$ ). Pour une valeur fixée du forçage (au repos ou en stimulation),  $sn_c^{*dec}$  diminue quand  $J_{gaba_c}$  augmente, ce qui était attendu (effet inhibiteur de  $J_{gaba_c}$  sur  $sn_c^{*dec}$ ).

Sur le graphe 13.15, la flèche rouge représente la variation d'activité de la cible due à la stimulation de la région de contrôle : c'est la différence entre  $sn_c^{*\text{dec}}$  stimulé (i.e. pour  $I_{stim} = 0.1$ ) et  $sn_c^{*\text{dec}}$  non stimulé (i.e. pour  $I_{stim} = 0$ ), à  $J_{gaba_c} = \widehat{J_{gaba_c}}$  fixé.

Pour rappel, une valeur fixée de  $I_{stim}$  donne une valeur fixée du forçage  $\widehat{B_{Erc}} = 0.382 + I_{stim}$  et  $I_{stim} \neq 0$  permet de faire varier le forçage. Dans tout ce qui suit nous avons explicité ce point en notant :  $\Delta\widehat{B_{Erc}} = I_{stim}$ .

Nous utilisons la notation  $\Delta sn_{c/B_{Erc}}^{*\text{dec}}(J_{gaba_c})$  pour désigner la variation d'activité  $sn_c^{*\text{dec}}$  entre  $B_{Erc} = \widehat{B_{Erc}}$  et  $B_{Erc} = \widehat{B_{Erc}} + \Delta\widehat{B_{Erc}} = \widehat{B_{Erc}} + I_{stim}$ , évaluée à une valeur fixée de  $J_{gaba_c}$ .

Nous notons  $sn_c^{*\text{dec}} = f_c^{\text{dec}}(B_{Erc}, J_{gaba_c})$  qui permet d'écrire explicitement que  $sn_c^{*\text{dec}}$  est une fonction de  $B_{Erc}$  et  $J_{gaba_c}$ .

La différence entre  $sn_c^{*\text{dec}}$  stimulé (i.e. pour  $I_{stim} = 0.1$ ) et  $sn_c^{*\text{dec}}$  non stimulé (i.e. pour  $I_{stim} = 0$ ), pour une valeur de fixée de  $J_{gaba_c} = \widehat{J_{gaba_c}}$  ( $= 0.5$  sur le graphe) est notée  $\Delta sn_{c/B_{Erc}}^{*\text{dec}}(\widehat{J_{gaba_c}})$  et est définie par :

$$\Delta sn_{c/B_{Erc}}^{*\text{dec}}(\widehat{J_{gaba_c}}) \equiv f_c^{\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) - f_c^{\text{dec}}(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (13.39)$$

Sur le graphe, nous voyons que cette différence est négative, comme attendu. La flèche rouge représente ainsi l'inhibition de l'activité de la cible par la région de contrôle (étape 6 du raisonnement causal).

La flèche violette représente la variation d'activité supplémentaire due à l'augmentation de  $J_{gaba_c}$  : c'est la différence entre  $sn_c^{*\text{dec}}$  stimulé pour  $J_{gaba_c} = \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$  ( $= 1$  sur le graphe) et  $sn_c^{*\text{dec}}$  stimulé pour  $J_{gaba_c} = \widehat{J_{gaba_c}}$  ( $= 0.5$ ).

Nous adoptons la notation  $\Delta sn_{c/J_{gaba_c}}^{*\text{dec}}(B_{Erc})$  pour désigner la variation d'activité  $sn_c^{*\text{dec}}$  entre  $J_{gaba_c} = \widehat{J_{gaba_c}}$  et  $J_{gaba_c} = \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$ , évaluée à une valeur fixée du forçage  $B_{Erc}$ .

La différence entre les deux valeurs de  $sn_c^{*\text{dec}}$  stimulées précédentes s'écrit donc  $\Delta sn_{c/J_{gaba_c}}^{*\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}) = \Delta sn_{c/J_{gaba_c}}^{*\text{dec}}(\widehat{B_{Erc}} + I_{stim})$  et est définie par :

$$\Delta sn_{c/J_{gaba_c}}^{*\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}) \equiv f_c^{\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c^{\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (13.40)$$

Sur le graphe, cette différence est négative, ce qui est cohérent avec l'effet inhibiteur de  $J_{gaba_c}$  sur la région cible découplée de la région de contrôle (cf propriété vérifiée de début de section). La flèche violette représente l'augmentation de l'inhibition de la cible, i.e. la sur-inhibition (étape 7 du raisonnement causal).

Au final, l'activité de la cible vaut  $sn_c^{*\text{dec}} = f_c^{\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}})$ . La variation totale d'activité entre  $sn_c^{*\text{dec}}$  stimulée évaluée à  $J_{gaba_c} = \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$  et  $sn_c^{*\text{dec}}$  non stimulée évaluée à  $J_{gaba_c} = \widehat{J_{gaba_c}}$  est notée  $\Delta sn_c^{*\text{dec}}$  (flèche verte sur le graphe). Elle est égale à la somme des deux variations précédentes (flèches rouge et violette) :

$$\Delta sn_c^{*\text{dec}} \equiv f_c^{\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c^{\text{dec}}(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (13.41)$$

$$= \underbrace{\Delta sn_{c/B_{Erc}}^{*\text{dec}}(\widehat{J_{gaba_c}})}_{\text{Inhibition}} + \underbrace{\Delta sn_{c/J_{gaba_c}}^{*\text{dec}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}})}_{\text{Sur-inhibition}} \quad (13.42)$$

Puisque  $\Delta sn_{c/J_{gaba_c}}^{*dec}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}) < 0$  (flèche violette), nous pouvons comparer  $\Delta sn_{c/B_{Erc}}^{*dec}(\widehat{J_{gaba_c}})$  (flèche rouge), par rapport à  $\Delta sn_c^{*dec}$  (flèche verte) :

$$\boxed{\Delta sn_c^{*dec} < \Delta sn_{c/B_{Erc}}^{*dec}(\widehat{J_{gaba_c}}) < 0} \quad (13.43)$$

Par rapport à l'effet produit par la stimulation de la région de contrôle (flèche rouge), l'augmentation de  $J_{gaba_c}$  conduit alors à diminuer encore plus l'activité de la cible.

Par ailleurs, nous constatons graphiquement que plus nous augmentons  $J_{gaba_c}$  par rapport à la référence  $\widehat{J_{gaba_c}}$ , et plus l'activité de la cible  $sn_c^{*dec}$  stimulé diminue (courbe orange décroissante). Donc plus  $J_{gaba_c}$  augmente et plus on augmente l'inhibition de l'activité de la cible.

Remarque : Le graphe équivalent pour la région de contrôle se trouve en annexe [18.10](#) et permet de vérifier que dans le système découplé, l'activité de la région de contrôle n'est pas affectée par la variation de  $J_{gaba_c}$  dans la cible.

### 13.2.3.2.2 Avec rétroaction : cas de deux régions couplées en EI

Nous nous redonnons la rétroaction de la région cible vers la région de contrôle (cf figure 13.16). Nous allons voir si l'augmentation de  $J_{gaba_c}$  produit toujours une augmentation de l'inhibition de l'activité de la cible. Comme dans le cas sans rétroaction, nous allons essayer de raisonner d'abord avec la métaphore des balances E/I.

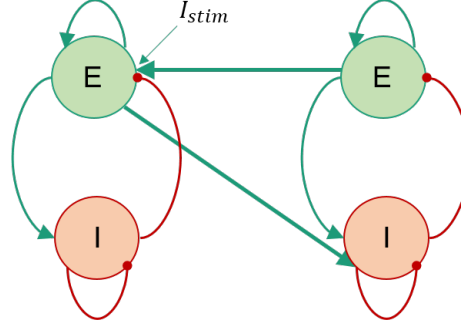


**Figure 13.16** – Deux régions connectées en E-I. La région de contrôle est à gauche et la région cible à droite.

#### 13.2.3.2.2.1 Peut-on prévoir le comportement du modèle couplé avec la métaphore de la balance E/I ?

Nous représentons à la figure 13.17 la métaphore de la balance dans le cas de deux régions couplées selon l'architecture de connectivité E-I.

Au repos (13.17 A), les balances sont couplées : elles agissent l'une sur l'autre sur les niveaux de leurs plateaux. Les valeurs de points fixes dans le système couplé sont différentes de celles du système découpé, i.e. sans rétroaction de la cible vers la région de contrôle.

Nous ajoutons successivement deux poids,  $I_{stim}$  (13.17 B) et  $\Delta J_{gaba_c}$  (13.17 C), qui sont posés sur les plateaux E ou I en fonction de l'effet excitateur ou inhibiteur qu'ils produisent dans une région isolée :  $I_{stim}$  est placé sur le plateau E de la région de contrôle (comme sur la figure 13.3) et  $\Delta J_{gaba_c}$  sur le plateau I de la région cible (comme sur la figure 13.5). Nous allons voir qu'il est compliqué de prédire le comportement des balances couplées à partir de celui des balances isolées ou découpées, d'abord dans le cas de l'application de  $I_{stim}$ , puis dans le cas de la variation  $\Delta J_{gaba_c}$ .

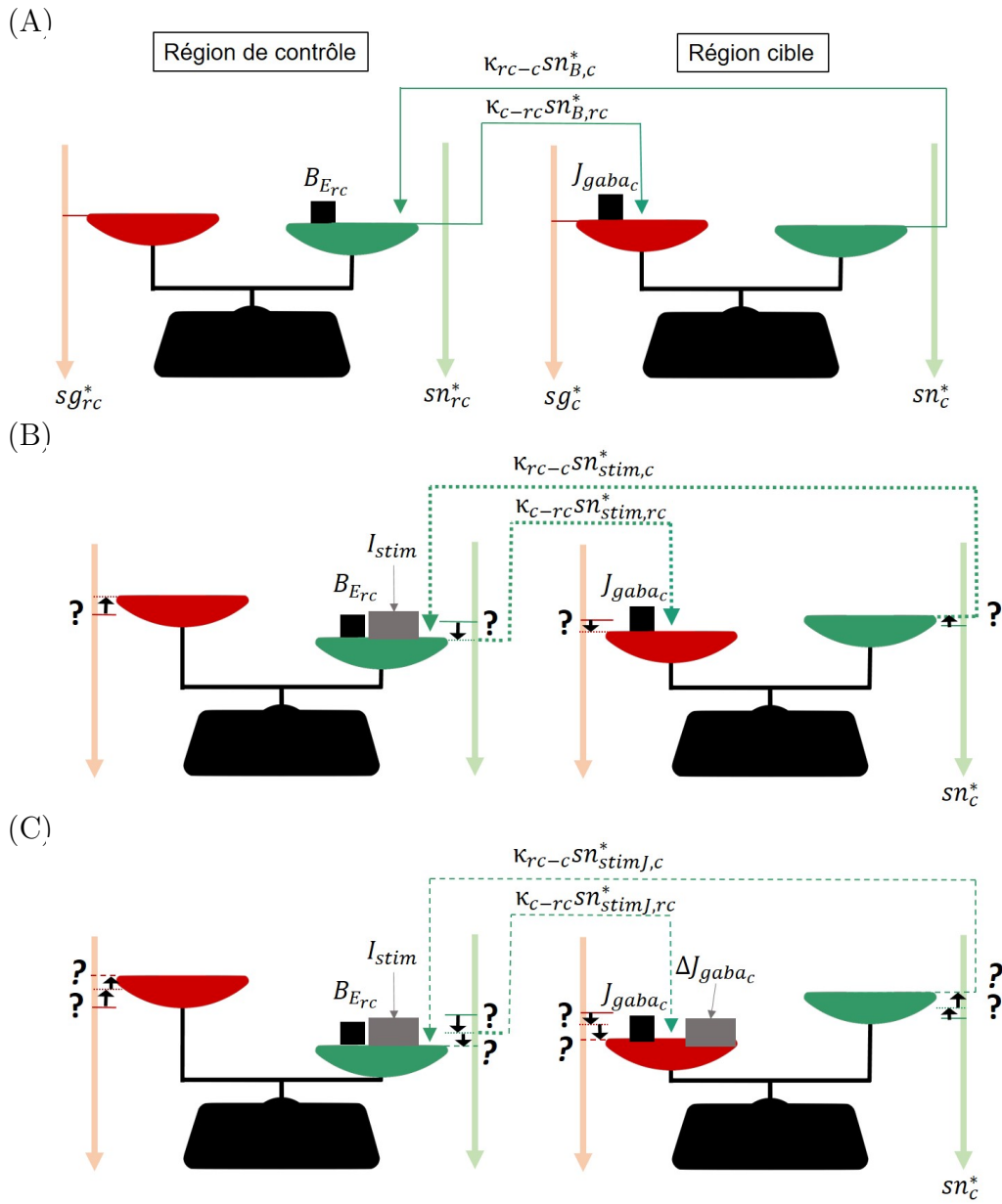


**Figure 13.17** – Balances de la région de contrôle et de la région cible couplées

A) au repos ;

B) avec stimulation de la région de contrôle (représentée par l'ajout du poids  $I_{stim}$  ;

C) avec variation supplémentaire de  $J_{gaba_c}$  dans la cible (représentée par l'ajout du poids  $\Delta J_{gaba_c}$ ). Les points fixes ne sont pas les mêmes que dans B), ils sont notés  $sn_{stimJ,rc}^*$  et  $sn_{stimJ,c}^*$  (obtenus avec la stimulation et en plus avec la variation de  $J_{gaba_c}$ ).

Lorsque la région de contrôle est isolée, nous connaissons le comportement de sa balance à l'application d'un poids  $\Delta x_E$  (i.e. sa variation d'activité  $sn^*$  en fonction de la variation du forçage  $x_E$  sur son pool excitateur, cf figure 13.2).

Dans le système découplé, la région de contrôle ne reçoit pas d'influence de la région cible, sa variation d'activité due à  $I_{stim}$  ne dépend pas de celle de la région cible (cf figure 13.14 B.). L'application d'un poids  $I_{stim}$  sur la balance de la région de contrôle découplée produit le même effet que  $\Delta x_E = I_{stim}$  pour la région de contrôle isolée. Plus précisément, le comportement de la région isolée avec  $x_E = B_{E_{rc}}$  et  $x_I = B_{I_{rc}}$  (valeurs de forçages au repos), sur laquelle nous appliquons  $\Delta x_E = I_{stim}$ , donne la variation d'activité de la région de contrôle dans le système découplé :  $\Delta sn^*_{/x_E} = \Delta sn^*_{rc/B_{E_{rc}}}^{dec}$  (nous pouvons comparer les figures 13.2 et 18.6 qui donnent respectivement des variations de  $sn^*$  en fonction de  $x_E$  et de  $sn^*_{rc/B_{E_{rc}}}^{dec}$  en fonction de  $I_{stim}$  et qui sont identiques). La région cible au contraire reçoit une influence de la région de contrôle. Sa variation d'activité dépend de celle de la région de contrôle et du couplage de la région de contrôle vers la cible<sup>18</sup>.

Dans le système couplé, la réponse de la balance de la région de contrôle due à l'ajout de  $I_{stim}$  ne peut pas être calculée uniquement à partir du comportement de la région isolée due à l'ajout d'un poids  $\Delta x_E = I_{stim}$ . En effet,  $I_{stim}$  a un effet sur la balance de la région de contrôle, qui a un effet sur la balance de la cible, qui a un effet retour sur la balance de la région de contrôle (cf figure 13.17 B). La variation d'activité de la cible diffère également de celle du système découplé, à cause de cet effet retour sur la région de contrôle. Nous montrons que dans le système couplé,  $I_{stim}$  conduit à une augmentation d'activité de la région de contrôle et une baisse d'activité de la cible (ce qui est représenté par les déplacements des plateaux de la figure 13.17 B), avec des amplitudes différentes de celles du cas découplé (cf annexes 18.2).

Lorsque la région cible est isolée, nous connaissons la réponse de sa balance à l'application d'un poids  $\Delta J_{gaba}$  (i.e. sa variation d'activité  $sn^*$  en fonction de la variation de  $J_{gaba}$ , cf figure 13.4).

Dans le cas découplé, nous avons vu que l'ajout de  $\Delta J_{gaba_c}$  sur le plateau inhibiteur de la région cible n'a pas d'effet sur la balance de la région de contrôle mais uniquement sur la cible (cf figure 13.14 C). La variation d'activité de la cible découplée n'est cependant pas égale à celle de la cible isolée. En effet, pour calculer cette variation d'activité, il faut pouvoir prendre en compte l'état de la balance dans le système découplé, qui est soumise à un forçage supplémentaire venant de la région de contrôle. La variation d'activité de la cible découplée est égale à la variation d'activité d'une cible isolée qui aurait une valeur particulière du forçage sur son pool inhibiteur :  $x_I = B_{I_c} + \kappa_{c-rc} sn^*_{stim,rc}^{dec}$  (i.e. une balance isolée avec un poids sur le plateau I qui appuie comme appuyait la région de contrôle dans le système découplé, cf 13.14 B.). Autrement dit, le comportement de la région cible isolée avec  $x_E = B_{E_c}$  et  $x_I = B_{I_c} + \kappa_{c-rc} sn^*_{stim,rc}^{dec}$ , sur laquelle nous appliquons  $\Delta J_{gaba} = \Delta J_{gaba_c}$

---

18. cf annexe 18.2.1.2 : dans le système sans rétroaction le calcul analytique de la sensibilité de  $sn^*_c$  par rapport à  $B_{E_{rc}}$  met bien en évidence un terme d'émission (sensibilité de la région de contrôle découplée), de transmission (couplage) et de réception (sensibilité de la cible découplée). La perturbation de  $B_{E_{rc}}$  a un effet sur la région de contrôle puis cet effet est transmis à la cible. Il s'agit d'une traduction en perturbation de la chaîne de causalité (13.24)



donne la variation d'activité de la cible dans le système découplé :  $\Delta sn_{/J_{gaba}}^* = \Delta sn_{c/J_{gaba_c}}^{*\text{dec}}$ . Dans le système couplé, la réponse de la balance de la cible due à l'ajout de  $\Delta J_{gaba_c}$  ne peut pas être calculée uniquement à partir du comportement de la région cible isolée après l'ajout d'un poids  $\Delta J_{gaba} = \Delta J_{gaba_c}$ . En effet, la variation de  $J_{gaba_c}$  agit sur la cible, qui agit sur la région de contrôle, qui a un effet retour sur la cible.

Le schéma C) de la figure 13.17 semble nous permettre de prédire qu'une augmentation de  $J_{gaba_c}$  conduise à un abaissement supplémentaire du plateau I de la cible, i.e. de  $sn_c^*$  stimulé : on appuie sur la plateau I de la cible, ce qui élève son plateau E, qui appuie en retour sur le plateau E de la région de contrôle, qui appuie alors davantage sur le plateau I de la cible. Ceci conduirait alors à une sur-inhibition de l'activité de la région cible. Nous ne pouvons néanmoins pas prédire l'amplitude à partir de celle obtenue dans le cas d'une région isolée. Seul le travail analytique permet de conforter cette prédiction et d'en connaître l'amplitude. Il est important de noter qu'à ce stade du travail, nous n'avons pas étudié l'effet d'une augmentation de  $J_{gaba_c}$  sur l'activité  $sn_c^*$  de la cible stimulée dans le système composé de deux régions couplées, mais seulement d'une région isolée (cf chapitre 7 ou annexe 18.1.1). Ce sera donc l'objet du paragraphe 13.2.3.2.2.2.

Bilan de l'usage de la métaphore de la balance E/I pour penser le comportement d'une région isolée, de régions dans le système sans et avec rétroaction

**La métaphore de la balance E/I locale permet de se représenter le comportement d'une région isolée ou découplée, mais présente des limites**

La balance E/I locale permet de se représenter l'effet de la variation d'un paramètre sur l'activité d'une région isolée, avec les limites que nous avons soulignées à la section 13.1.4, notamment le fait que sa réactivité (i.e. sensibilité) dépend de son état (i.e. des valeurs de tous les autres paramètres). Pour une même perturbation, le déplacement du plateau E donnant  $\delta sn^*$  n'est pas toujours le même parce qu'il dépend de tous les paramètres de la région isolée. En variation,  $\Delta sn^*$  peut s'écrire comme une somme de  $\delta sn^*$  et la précaution d'interprétation est la même : il faut lire le déplacement du plateau E donnant  $\Delta sn^*$  comme une fonction de l'état de la balance avant cette variation.

Quand on étudie le système découplé (sans rétroaction de la cible vers la région de contrôle), connaître le comportement des balances E/I des régions isolées pour différentes valeurs de forçages  $x_E$  ou  $x_I$  permet de connaître le comportement de la balance de chaque région découplée.

**La métaphore de la balance E/I est compliquée à utiliser quand le système comprend des rétroactions**

Pour prévoir et comprendre le comportement du modèle couplé, nous venons de voir que la présence du couplage en retour de la cible vers la région de contrôle rend l'utilisation de la métaphore de la balance E/I locale plus compliquée qu'utile. Il est nécessaire alors de l'étudier avec les outils d'analyses des systèmes non linéaires, qui permettent de prendre en compte les non linéarités (i.e. la réactivité d'une balance dépend de son état) et l'effet de tous les couplages.

Le découpage-raccordement est un exemple d'outil analytique utilisé dans le chapitre 9 qui nous a permis de construire, de manière intelligible dans le système couplé, les sensibilités de  $sn_{rc}^*$  et de  $sn_c^*$  à une perturbation du forçage  $B_{E_{rc}}$  sur le pool E de la région de contrôle. Nous avons écrit, dans le système couplé, la réactivité des balances E/I de la région de contrôle et de la cible dues à l'ajout d'un poids  $\delta B_{E_{rc}}$  sur le plateau E de la balance de la région de contrôle, en fonction des réactivités des balances prises de manière isolée, i.e. en coupant tous les couplages. Evidemment, comme la réactivité d'une balance isolée dépend de son état, il est nécessaire de l'évaluer en prenant en compte son état dans le système couplé (au repos, i.e. avant la perturbation). Quand nous coupons tous les couplages, il est nécessaire d'ajuster les poids  $x_E$  et  $x_I$  sur les plateaux des balances pour retrouver au repos les valeurs de points fixes du système couplé : pour la région de contrôle  $x_E = B_{E_{rc}} + \kappa_{rc-c} sn_c^*$  et  $x_I = B_{I_c}$  ; et pour la région cible  $x_E = B_{E_{rc}}$  et  $x_I = B_{I_c} + \kappa_{c-rc} sn_{rc}^*$ .

L'intégration des sensibilités permet ensuite de passer d'une pensée en perturbation du forçage  $\delta B_{E_{rc}}$  à une pensée en variation du forçage  $\Delta B_{E_{rc}} = I_{stim}$ .

### 13.2.3.2.2.2 Effet d'une augmentation de $J_{gaba_c}$ dans la cible sur l'activité $sn_c^*$ de la cible stimulée

#### En perturbation

Pour étudier l'effet excitateur ou inhibiteur infinitésimal de  $J_{gaba_c}$ , nous précisons les éléments suivant :

- paramètre perturbé ( $\pi$ ) :  $J_{gaba_c}$
- la variable d'observation ( $\eta$ ) : activité de la cible pour une valeur fixée du forçage dans la région de contrôle, en période de stimulation :  $B_{Erc} = \widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$ .  
On note cette activité :  $sn_c^* = f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, J_{gaba_c})$
- boucle de rétroaction ( $\mathcal{B}$ ) : 2 régions : une région de contrôle et une cible couplées entre elles avec une architecture de connectivité EI.

Dans un système de deux régions couplées, nous avons calculé en annexes 18.3 la sensibilité de  $sn_c^*$  par rapport à  $J_{gaba_c}$ , qui dépend de  $B_{Erc}$  et  $J_{gaba_c}$ . Nous la notons  $S_{sn_c/J_{gaba_c}}(B_{Erc}, J_{gaba_c})$ . C'est la réactivité de la balance de la cible à une perturbation  $\delta J_{gaba_c}$  dans le système couplé<sup>19</sup>.

La perturbation de  $sn_c^*$  due à  $\delta J_{gaba_c}$  est donnée par :

$$\delta sn_{c/J_{gaba_c}} = S_{sn_c/J_{gaba_c}}(B_{Erc}, J_{gaba_c})\delta J_{gaba_c} \quad (13.44)$$

Nous montrons que  $\delta sn_{c/J_{gaba_c}} < 0$  pour n'importe quel type de connectivité, y compris une connectivité EI, et n'importe quelle valeur du forçage, en particulier pour  $B_{Erc} = \widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$ .

Dans le système couplé, la perturbation de  $J_{gaba_c}$  dans la cible conduit donc à un abaissement de l'activité excitatrice de la cible  $sn_c^*$  stimulée.

$J_{gaba_c}$  a un effet inhibiteur infinitésimal sur l'activité  $sn_c^*$  stimulée de la cible.

#### En variation

Nous partons de  $sn_c^*$  évalué en  $B_{Erc} = \widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$  et  $J_{gaba_c} = \widehat{J_{gaba_c}}$ . Pour préciser l'effet excitateur ou inhibiteur fini de la variation de  $J_{gaba_c}$ , nous reprenons les éléments du triplet précédent, et nous ajoutons l'intervalle de variation du paramètre : nous faisons varier  $J_{gaba_c}$  entre  $\widehat{J_{gaba_c}}$  et  $\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$ .  $\Delta sn_{c/J_{gaba_c}}$  désigne la variation de  $sn_c^*$  dans la cible due à cette variation de  $J_{gaba_c}$  dans la cible. Nous pouvons l'écrire en fonction de la sensibilité de  $sn_c^*$  par rapport à  $J_{gaba_c}$  :

$$\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}) = \int_{\widehat{J_{gaba_c}}}^{\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}} S_{sn_c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, J_{gaba_c}) dJ_{gaba_c} \quad (13.45)$$

$$\equiv f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (13.46)$$

19. Nous l'exprimons dans l'annexe 18.3 en fonction de la réactivité des balances de la région de contrôle et de la cible à une perturbation de leur forçage si nous avons coupé tous les couplages

D'après le travail fait en annexes 18.3, nous savons que :

$$\forall (B_{E_{rc}}, J_{gaba_c}), S_{sn_c/J_{gaba_c}}(B_{E_{rc}}, J_{gaba_c}) < 0 \quad (13.47)$$

Donc  $\Delta sn_c/J_{gaba_c}(\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}}) < 0$ . Une variation de  $J_{gaba_c}$  produit une diminution de l'activité  $sn_c$  stimulée de la région cible.

#### Augmentation de la variation $\Delta \widehat{J_{gaba_c}}$

Quand  $\Delta \widehat{J_{gaba_c}}$  augmente, la borne supérieure de l'intégrale augmente. Comme l'intégrande est négative,  $\Delta sn_c/J_{gaba_c}(\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}})$  reste négative et augmente en valeur absolue<sup>20</sup>.

En conclusion, nous montrons que  $J_{gaba_c}$  a un effet inhibiteur fini sur l'activité de la région cible  $sn_c$  stimulée. Plus  $J_{gaba_c}$  augmente et plus l'activité de la cible baisse.

Nous savons que la stimulation de la région de contrôle conduit à abaisser l'activité de la cible. Nous savons maintenant qu'une augmentation de  $J_{gaba_c}$  dans le système couplé conduit à abaisser l'activité de la cible, en particulier de  $sn_c$  stimulée. Ceci nous permet de prévoir un effet sur-inhibiteur de l'augmentation de  $J_{gaba_c}$  dans le système couplé en EI.

#### 13.2.3.2.2.3 Variations des points fixes : définitions formelles de l'*inhibition* et de la *sur-inhibition*

Comme dans le cas sans rétroaction (paragraphe 13.2.3.2.1), nous pouvons étudier les variations de points fixes dans le système couplé en EI. Nous avons tracé à la figure 13.18 l'activité  $sn_c^*$ , pour différentes valeurs de  $J_{gaba_c}$  dans la cible et pour différentes valeurs du forçage  $B_{E_{rc}}$  dans la région de contrôle.

Nous représentons par des flèches sur le graphe 13.18 les différences introduites précédemment.

La flèche rouge représente la variation d'activité de la cible due à la stimulation de la région de contrôle. Elle est bien négative, comme attendu dans le cas de la connectivité EI (cf article du chapitre 9).

La flèche violette représente la variation supplémentaire d'activité due à l'augmentation de  $J_{gaba_c}$ . Elle est négative, ce qui est cohérent avec l'effet inhibiteur de  $J_{gaba_c}$  sur l'activité de la cible dans le système couplé (pour une valeur fixée du forçage dans la région de contrôle, ici en période de stimulation).

Les conclusions sont les mêmes qu'au paragraphe 13.2.3.2.1 : seules les valeurs des différences changent mais leur signe ne change pas.

---

20. Nous pourrions tracer  $\Delta sn_c/J_{gaba_c}$  en fonction de  $J_{gaba_c}$ , pour une valeur fixée du forçage  $B_{E_{rc}}$ . Pour n'importe quel type de connectivité, on s'attend à voir une courbe  $y$  négative et décroissante.

$$\forall J_{gaba_c} > \widehat{J_{gaba_c}}, y = f_c(\widehat{B_{E_{rc}}}, J_{gaba_c}) - f_c(\widehat{B_{E_{rc}}}, \widehat{J_{gaba_c}}) \quad (13.48)$$

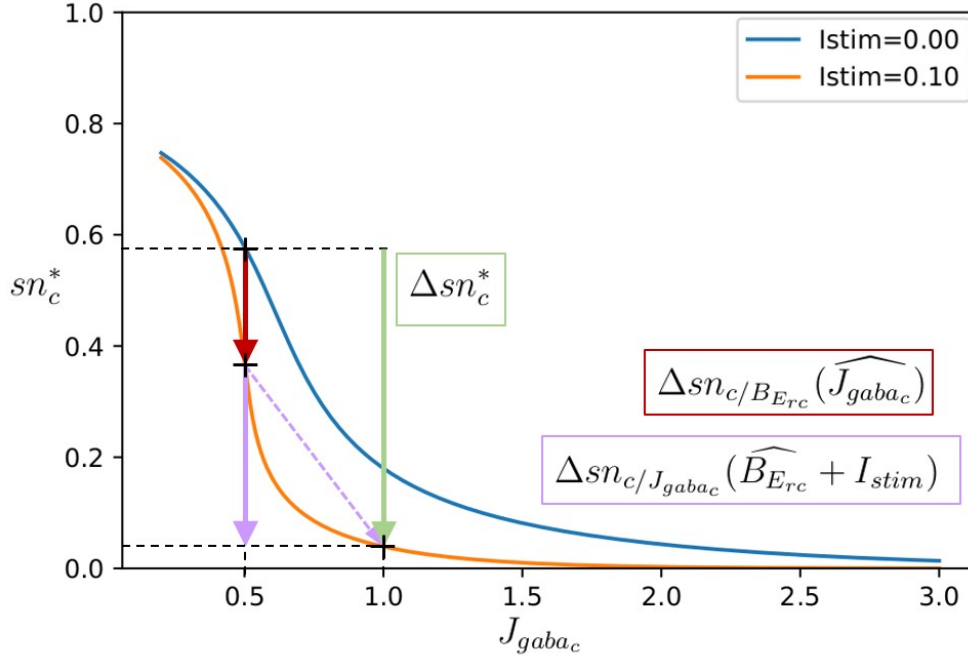


**Figure 13.18** – Avec rétroaction de la cible vers la région de contrôle. Même légende que la figure 13.15, en enlevant la notation <sup>dec</sup>.

Puisque  $\Delta sn_c/J_{gaba_c}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}) < 0$  (flèche violette), la variation totale d'activité de la cible  $\Delta sn_c^*$  (flèche verte) est négative et plus grande en valeur absolue que  $\Delta sn_c/B_{Erc}(\widehat{J_{gaba_c}})$  (flèche rouge). L'augmentation de  $J_{gaba_c}$  permet d'augmenter l'inhibition de l'activité de la cible (i.e. de sur-inhiber). Nous avons :

$$\Delta sn_c^* < \Delta sn_c/B_{Erc}(\widehat{J_{gaba_c}}) < 0 \quad (13.49)$$

Par ailleurs, plus nous augmentons  $\Delta \widehat{J_{gaba_c}}$  et plus la surinhibition est importante (i.e.  $\Delta sn_c/J_{gaba_c}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}) < 0$  devient grande en valeur absolue). En conséquence, l'activité de la cible  $sn_c^*$  diminue de plus en plus par rapport à la situation de départ  $sn_c^* = f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}})$  (i.e.  $\Delta sn_c^* < 0$  devient plus grande en valeur absolue).

Remarque :  $J_{gaba_c}$  a aussi un effet sur la région de contrôle, contrairement au cas sans rétroaction. Le graphe équivalent pour la région de contrôle se trouve en annexe 18.9 et permet de vérifier qu'avec l'effet retour de la cible, l'activité de la région de contrôle est affectée par la variation de  $J_{gaba_c}$  dans la cible. C'est ce que représentent les déplacements de plateaux de la région de contrôle à la figure 13.17 C. Une étude en perturbation, montre que la sensibilité de  $sn_{rc}^*$  à  $J_{gaba_c}$  n'est pas nulle. Nous avons  $\delta sn_{rc}/J_{gaba_c} < 0$  uniquement dans le cas des connexions E-E et E-I (cf annexe 18.3). En EI, pour toute valeur fixée du forçage  $B_{Erc}$ , la perturbation de  $J_{gaba_c}$  dans la cible produit donc une baisse de l'activité  $sn_{rc}^*$  dans la région de contrôle.  $J_{gaba_c}$  a un effet inhibiteur (infinitésimal et fini) sur l'activité  $sn_{rc}^*$  de la région de contrôle dans un système de deux régions couplées, pour toute valeur fixée du forçage  $B_{Erc}$ .

#### 13.2.3.2.2.4 Note sur la variation totale d'activité $\Delta sn_c^*$ : deux chemins possibles pour la construire

Nous avons construit analytiquement la variation totale d'activité  $\Delta sn_c^*$ , due à la variation de  $I_{stim}$  et de  $J_{gaba_c}$  dans un système couplé (cf annexes 18.4). La linéarisation autour du point fixe permet de séparer  $\Delta sn_c^*$  en deux contributions, chacune portant l'effet de la variation d'un seul des deux paramètres. Nous vérifions que le raisonnement linéaire qui consiste à ajouter les deux variations  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$  et  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}})$  est donc valide.

Si nous partons de l'activité de la cible au régime stationnaire évaluée en  $J_{gaba_c} = \widehat{J_{gaba_c}}$  et  $B_{Erc} = \widehat{B_{Erc}}$ , i.e.  $sn_c^* = f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}})$ , il existe alors deux chemins possibles pour arriver à l'état  $sn_c^* = f_c(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}})$ , représentés sur la figure 13.19 :

- premier chemin : variation de  $B_{Erc}$  puis de  $J_{gaba_c}$  (ce que nous venons de voir),
  - second chemin : variation de  $J_{gaba_c}$  puis de  $B_{Erc}$ . Nous pouvons en effet noter que pour arriver à la même variation totale d'activité  $\Delta sn_c^*$ , il aurait été possible de sommer les deux différences représentées sur le graphe de droite de la figure 13.19.
- 1)  $J_{gaba_c}$  a un effet inhibiteur sur l'activité de la cible au repos : quand on augmente  $J_{gaba_c}$  de  $\Delta \widehat{J_{gaba_c}}$ , l'activité de la cible au repos (i.e. avec  $I_{stim} = 0$ ) baisse, ce qui est représenté par la flèche violette ( $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}}) < 0$ ). Plus  $\Delta \widehat{J_{gaba_c}}$  augmente et plus  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}})$  est grande en valeur absolue. La courbe bleue est en effet décroissante.
- 2) La stimulation de la région de contrôle abaisse encore l'activité de la cible, ce qui est représenté par la flèche rouge ( $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) < 0$  cohérent en EI).

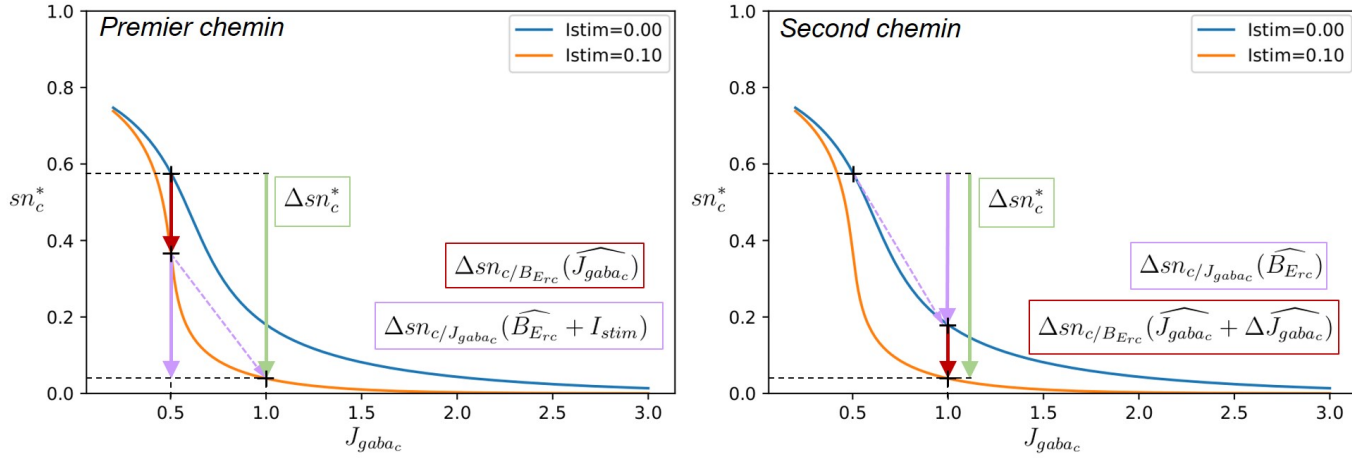


**Figure 13.19** – Deux possibilités pour calculer la variation totale d'activité de la cible soumise à une stimulation de la région de contrôle (application de  $I_{stim}$ ) et à une variation de  $J_{gaba_c}$ .  
A gauche : stimulation (flèche rouge) puis variation de  $J_{gaba_c}$  (flèche violette). A droite : variation de  $J_{gaba_c}$  (flèche violette) puis stimulation (flèche rouge).

### 13.2.3.2.3 Prendre en compte les mesures entre participants

#### 13.2.3.2.3.1 Contrôle inhibiteur : $J_{gaba_c}$ fixe pour un participant

Pour étudier la variabilité interindividuelle du contrôle inhibiteur (par exemple celui des intrusions mnésiques), nous avons proposé un modèle neural du contrôle inhibiteur. Sous l'hypothèse de ce modèle, la région de contrôle et la région cible sont connectées en EI et le contrôle inhibiteur est une pensée en variation du forçage  $B_{Erc}$  : en période de stimulation la région de contrôle subit une variation du forçage sur son pool excitateur d'une amplitude valant  $I_{stim}$ . Il n'y a pas de variation de  $J_{gaba_c}$  qui modélise la variation des activités neurales d'un sujet qui effectue une tâche de contrôle inhibiteur<sup>21</sup>. Dans le modèle, chaque participant a donc un  $J_{gaba_c}$  fixe. Pour chaque participant, on stimule la région de contrôle avec  $I_{stim}$  pour un  $J_{gaba_c}$  fixé.

Dans le chapitre 7, nous avons émis l'hypothèse d'une différence de  $J_{gaba_c}$  dans les régions cibles entre les participants. Pour étudier l'effet d'un écart de  $J_{gaba_c}$  entre participants, sur la baisse d'activité de la région cible due à une même stimulation  $I_{stim}$  de la région de contrôle, il est nécessaire de comparer entre elles les baisses d'activité  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  pour des valeurs différentes et fixées de  $J_{gaba_c}$ .

#### 13.2.3.2.3.2 Exemple avec deux participants

Supposons deux participants dont les  $J_{gaba_c}$  dans la cible valent respectivement  $\widehat{J_{gaba_c}}$  et  $\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}$ , le premier étant strictement inférieur au second (i.e.  $\Delta \widehat{J_{gaba_c}} > 0$ ). Nous avons représenté sur la figure 13.20, les deux différences d'activité à comparer.

Pour le sujet 1, la baisse d'activité due à la stimulation de la région de contrôle ne change pas (flèche rouge) :  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ .

Pour le participant 2, la diminution d'activité vaut  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}})$  (flèche grise) et non  $\Delta sn_c^*$  (flèche verte) : nous faisons varier uniquement  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}$ , pour  $J_{gaba_c} = \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}$  fixé. Même si le point d'arrivée pour construire ces différences est identique ( $sn_c^*$  stimulé avec  $J_{gaba_c} = \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}$ ), le point de départ n'est cependant pas le même. Nous devons prendre l'activité au repos de la région cible pour ce même participant, c'est-à-dire le 2, et non le 1 comme nous l'avons fait pour construire  $\Delta sn_c^*$ .

Nous souhaitons donc comparer  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}})$  par rapport à  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ , c'est-à-dire les flèches rouge et grise (cas 2 sur 13.20) et non les flèches rouge et verte (cas 1). Une fois remis cet élément en tête, je ne saurais construire à ce stade du travail un raisonnement permettant de prédire une relation d'ordre entre ces deux grandeurs.

---

21. L'amplitude de la variation du forçage  $\Delta \widehat{B_{Erc}} = I_{stim}$  est un paramètre du modèle de la tâche, quand  $\Delta J_{gaba_c}$  fait partie du métamodèle (cf 7.3).

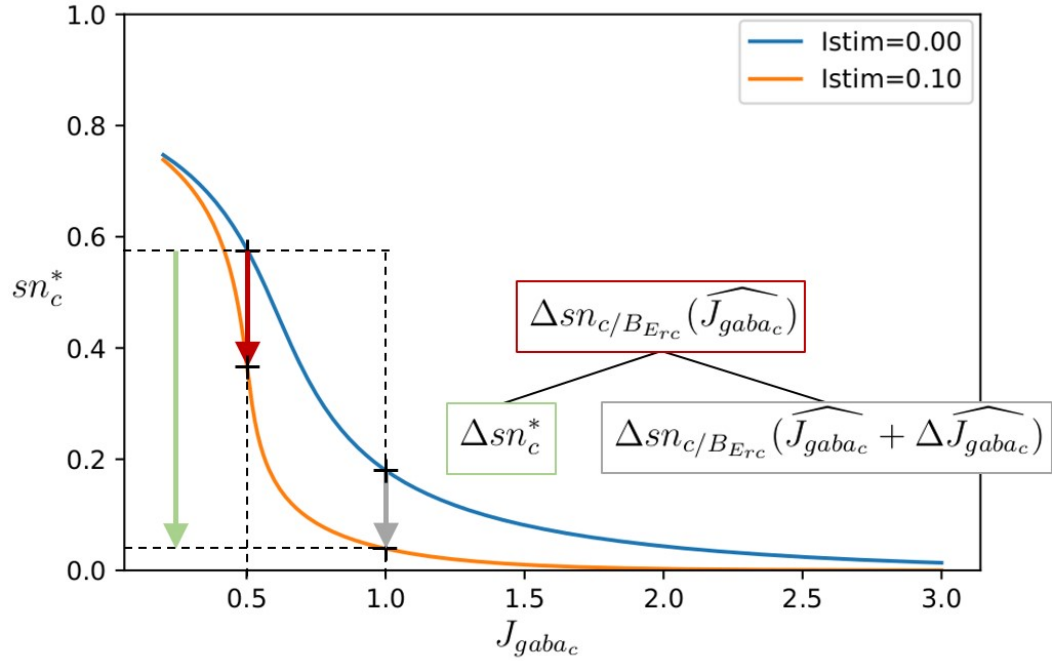


**Figure 13.20** – Récapitulatif des paires de variations à comparer pour étudier l'effet sur-inhibiteur ou sous-inhibiteur de  $J_{gaba_c}$  :

- 1) Sans prise en compte des mesures entre participants : flèches rouge et verte,
- 2) Avec prise en compte des mesures entre participants : flèches rouge et grise.

De manière plus générale, notre question peut s'énoncer de la manière suivante : comment la variation d'activité  $sn_c^*$  dans la cible due à une variation fixée du forçage  $B_{Erc}$  sur le pool excitateur de la région de contrôle dépend de  $J_{gaba_c}$  dans la cible ?

La variation d'activité  $sn_c^*$  due à une variation fixée du forçage  $B_{Erc}$ , évaluée pour une valeur fixée de  $J_{gaba_c}$ , s'écrit formellement :

$$\Delta sn_{c/B_{Erc}}(J_{gaba_c}) \equiv f_c(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}, J_{gaba_c}) - f_c(\widehat{B_{Erc}}, J_{gaba_c}) \quad (13.50)$$

Nous nous intéressons à la variation  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  en fonction de  $J_{gaba_c}$ , pour  $\widehat{B_{Erc}}$  et  $\Delta \widehat{B_{Erc}}$  fixés, dans le cas de deux régions couplées en EI.



#### 13.2.3.2.4 Effet de $J_{gaba_c}$ sur la variation d'activité de la cible $\Delta sn_{c/B_{Erc}}$ due à une stimulation de la région de contrôle

Dans le cas d'une connectivité EI, nous savons que la baisse d'activité due à la stimulation de la région de contrôle est négative. Pour une valeur de  $J_{gaba_c}$ ,  $\Delta sn_{c/B_{Erc}}(J_{gaba_c}) < 0$  traduit l'inhibition de l'activité de la région cible par la région de contrôle.

Nous allons augmenter de manière infinitésimale  $J_{gaba_c}$  et étudier si cette perturbation provoque une augmentation ou une diminution (en valeur absolue) de l'inhibition de la cible, désignées respectivement par *sur-inhibition* ou *sous-inhibition*.

##### 13.2.3.2.4.1 Définir un effet sur-inhibiteur ou sous-inhibiteur infinitésimal de $J_{gaba_c}$

Pour identifier l'effet sur-inhibiteur ou sous-inhibiteur infinitésimal d'une perturbation de  $J_{gaba_c}$  sur la baisse d'activité de la cible pendant une tâche de contrôle inhibiteur, nous nous donnons le triplet suivant (cf encadré 13.1.1.1) :

- paramètre perturbé ( $\pi$ ) :  $J_{gaba_c}$
- variable d'observation ( $\eta$ ) : variation d'activité de la cible due à une stimulation connue de la région de contrôle, i.e. une variation fixée du forçage sur le pool exciteur de la région de contrôle<sup>a</sup>, notée  $\eta = \Delta sn_{c/B_{Erc}}(J_{gaba_c})$ .
- boucle de rétroaction ( $\mathcal{B}$ ) : 2 régions couplées, une région de contrôle et une région cible pour une connectivité EI

---

a.  $B_{Erc}$  varie entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}$  ou  $\widehat{B_{Erc}} + I_{stim}$

$J_{gaba_c}$  a un effet sur-inhibiteur (resp. sous-inhibiteur) sur  $\eta = \Delta sn_{c/B_{Erc}}(J_{gaba_c})$  si une perturbation  $\delta J_{gaba_c}$  donne une perturbation négative (resp. positive) de  $\Delta sn_{c/B_{Erc}}$ . Autrement dit, si  $\frac{\partial \Delta sn_{c/B_{Erc}}(J_{gaba_c})}{\partial J_{gaba_c}} < 0$  est négative (resp. positive).

Remarque : choix de la variable d'observation

Le contrôle inhibiteur est modélisé par une variation du forçage  $B_{Erc}$  sur le pool exciteur de la région de contrôle, et non une perturbation. La variable d'observation est donc une variation  $\Delta sn_{c/B_{Erc}}$  (obtenue pour une variation connue de  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}$ ) et non une perturbation  $\delta sn_{c/B_{Erc}}$  (obtenue pour une perturbation  $\delta B_{Erc}$ ). Nous ne voulons pas seulement savoir comment la perturbation de l'activité  $sn_c^*$  de la cible due à une perturbation du forçage  $B_{Erc}$  sur le pool exciteur de la région de contrôle dépend de  $J_{gaba_c}$  dans la cible (i.e.  $\delta sn_{c/B_{Erc}}(J_{gaba_c})$  en fonction de  $J_{gaba_c}$ ), mais comment la variation de l'activité  $sn_c^*$  de la cible due à une variation fixée du forçage  $B_{Erc}$  sur le pool exciteur de la région de contrôle dépend de  $J_{gaba_c}$  dans la cible (i.e.  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  en fonction de  $J_{gaba_c}$ ).

#### 13.2.3.2.4.2 Un résultat de la thèse : la courbe en "U"

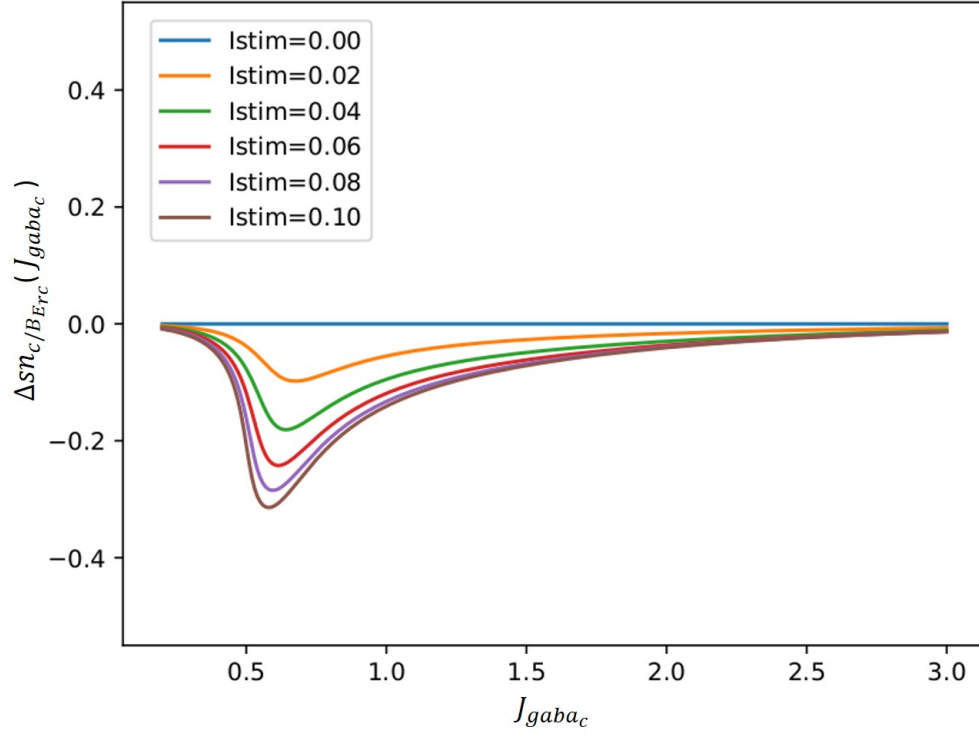


**Figure 13.21** –  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  en fonction de  $J_{gaba_c}$  dans le cas d'une architecture de connectivité EI.

Dans un système composé de 2 régions couplées en EI, nous avons calculé  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  pour différentes valeurs de  $J_{gaba_c}$  fixées (cf figure 13.21, reprise du chapitre 9). Pour cela, nous avons soustrait les  $sn_c^*$  stimulé et non stimulé pour chaque valeur de  $J_{gaba_c}$ <sup>22</sup>. Nous constatons graphiquement que  $\Delta sn_{c/B_{Erc}}$  décroît pour des valeurs de  $J_{gaba_c}$  en dessous d'un seuil (sur un intervalle  $\mathcal{J}$ ), puis croît au delà (sur un intervalle  $\mathcal{E}$ ). Le minimum de  $\Delta sn_{c/B_{Erc}}$  est atteint pour une valeur de  $J_{gaba_c}$  qui dépend de la valeur de  $\Delta \widehat{B}_{Erc} = I_{stim}$ . L'amplitude de ce minimum en dépend aussi.

Une perturbation de  $J_{gaba_c}$  a donc un effet sur-inhibiteur infinitésimal sur  $\Delta sn_{c/B_{Erc}}$  pour des valeurs de  $J_{gaba_c}$  dans  $\mathcal{J}$ , puis un effet sous-inhibiteur infinitésimal dans  $\mathcal{E}$ . Ce résultat inattendu est contradictoire avec la prédiction proposée au paragraphe 13.2.3.2.1.

<sup>22</sup>. une autre manière analytique de contruire la courbe en U est présentée à la formule 18.142 en annexes 18.4.

### Choix de vocabulaire : sur-inhibition et sous-inhibition

J'ai d'abord voulu définir un *effet inhibiteur (resp. excitateur) de  $J_{gaba_c}$  sur  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$* , si une perturbation  $\delta J_{gaba_c}$  donne une perturbation négative (resp. positive) de  $\eta^* = \Delta sn_{c/B_{Erc}}(J_{gaba_c})$ .

Je voyais alors la variation  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  comme une variable d'observation, au même titre que l'était  $sn_c^*$  à d'autres endroits de la discussion. L'effet inhibiteur ou excitateur était ainsi pensé comme le résultat négatif ou positif de la perturbation d'un paramètre sur une variable d'observation  $\eta^*$ . Une pente négative de la courbe  $\eta^* = f(J_{gaba_c})$  représentait une inhibition de  $\eta^*$  (quand  $J_{gaba_c} \in \mathcal{J}$ ), et une pente positive une excitation (quand  $J_{gaba_c} \in \mathcal{E}$ ).

Or, puisque  $\Delta sn_{c/B_{Erc}} < 0$  est la traduction formelle de l'inhibition de la cible, un collègue biologiste m'a fait remarquer qu'il était alors possible de dire : "la perturbation de  $J_{gaba_c}$  provoque une inhibition de l'inhibition" (quand  $J_{gaba_c} \in \mathcal{J}$ ), ou "la perturbation de  $J_{gaba_c}$  provoque une excitation de l'inhibition" (quand  $J_{gaba_c} \in \mathcal{E}$ ). L'emploi de ce vocabulaire crée de la confusion et ne facilite pas la communication. Nous l'abandonnons donc au profit de celui de *sur-inhibition* et *sous-inhibition*.

Le vocabulaire "effet inhibiteur ou excitateur d'un paramètre sur une variable d'observation" est restreint au cas où les variables d'observation sont directement des variables du modèle (e.g.  $\eta^* = sn_c^*$ ), et non des variations de ces variables (e.g.  $\eta^* = \Delta sn_{c/B_{Erc}}(J_{gaba_c})$  qui est la variation entre  $sn_c^*$  stimulé et non stimulé pour une valeur fixée de  $J_{gaba_c}$ ).

### 13.2.3.2.4.3 Traduction mathématique et explorations de l'origine du profil en "U"

Du point de vue analytique, pour  $\widehat{B_{Erc}}$  et  $\Delta \widehat{B_{Erc}}$  fixés :

- si  $J_{gaba_c} \in \mathcal{J}$ , la dérivée de  $\Delta sn_{c/B_{Erc}}$  par rapport à  $J_{gaba_c}$  est négative :

$$\frac{\partial \Delta sn_{c/B_{Erc}}(J_{gaba_c})}{\partial J_{gaba_c}} < 0$$

- si  $J_{gaba_c} \in \mathcal{E}$ , la dérivée de  $\Delta sn_{c/B_{Erc}}$  par rapport à  $J_{gaba_c}$  est positive :

$$\frac{\partial \Delta sn_{c/B_{Erc}}(J_{gaba_c})}{\partial J_{gaba_c}} > 0$$

Nous avons essayé de mieux comprendre l'origine de la forme en "U" des courbes  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  en fonction de  $J_{gaba_c}$  (Figure 13.21). Nous rapportons ici ce que nous avons fait, d'abord dans le cas de la connectivité EI (première exploration), puis sans rétroaction de la cible vers la région de contrôle (deuxième exploration).

## Première exploration :

Nous avons :

$$\Delta s_{n_c/B_{Erc}}(J_{gaba_c}) = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}} S_{s_{n_c/B_{Erc}}}(B_{Erc}, J_{gaba_c}) dB_{Erc} \quad (13.51)$$

Nous pouvons alors écrire la dérivée par rapport à  $J_{gaba_c}$  comme suit :

$$\frac{\partial \Delta s_{n_c/B_{Erc}}(J_{gaba_c})}{\partial J_{gaba_c}} = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}} \frac{\partial S_{s_{n_c/B_{Erc}}}(B_{Erc}, J_{gaba_c})}{\partial J_{gaba_c}} dB_{Erc} \quad (13.52)$$

Nous avons étudié le signe de la dérivée par rapport à  $J_{gaba_c}$  de la sensibilité par rapport au forçage  $B_{Erc}$ , i.e.  $\frac{\partial S_{s_{n_c/B_{Erc}}}(B_{Erc}, J_{gaba_c})}{\partial J_{gaba_c}}$ , pour des valeurs de  $B_{Erc}$  dans  $[\widehat{B_{Erc}}, \widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}]$  et pour des valeurs de  $J_{gaba_c}$  dans  $\mathcal{J}$  et  $\mathcal{E}$ . En effet, nous voulions vérifier si :

$$\forall B_{Erc} \in [\widehat{B_{Erc}}, \widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}], \forall J_{gaba_c} \in \mathcal{J}, \quad \frac{\partial S_{s_{n_c/B_{Erc}}}(B_{Erc}, J_{gaba_c})}{\partial J_{gaba_c}} < 0 \quad (13.53)$$

$$\forall B_{Erc} \in [\widehat{B_{Erc}}, \widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}], \forall J_{gaba_c} \in \mathcal{E}, \quad \frac{\partial S_{s_{n_c/B_{Erc}}}(B_{Erc}, J_{gaba_c})}{\partial J_{gaba_c}} > 0 \quad (13.54)$$

Nous avons donc tracé la sensibilité  $S_{s_{n_c/B_{Erc}}}$  en fonction de  $J_{gaba_c}$  pour plusieurs valeurs de  $B_{Erc}$ <sup>23</sup>. Nous avons constaté des changements de signe des pentes de  $S_{s_{n_c/B_{Erc}}}$  sur les intervalles  $\mathcal{J}$  et  $\mathcal{E}$ .

Pour le moment, nous ne savons pas expliquer le profil en "U" de  $\Delta s_{n_c/B_{Erc}}$  en fonction de  $J_{gaba_c}$ . Néanmoins, nous avons vu dans l'article du chapitre 9 :

1. les limites de la fonction  $\Delta s_{n_c/B_{Erc}} : \Delta s_{n_c/B_{Erc}}(J_{gaba_c})$  tend vers 0 quand  $J_{gaba_c}$  tend vers 0 ou vers  $+\infty$ .
2. le signe de la fonction pour une connectivité EI :  $\forall J_{gaba_c}, \Delta s_{n_c/B_{Erc}}(J_{gaba_c}) < 0$ .

Nous savons alors qu'il existe au moins un minimum de  $\Delta s_{n_c/B_{Erc}}$  pour  $J_{gaba_c} \in [0; +\infty[$ .

## Seconde exploration :

Nous avons également exploré si ce profil en "U" était conservé dans un système sans rétroaction de la cible vers la région de contrôle. Comme dans le cas couplé, nous avons soustrait les  $s_{n_c}^*$  stimulé et non stimulé, pour différentes valeurs de  $J_{gaba_c}$  (en prenant soin d'annuler la constante de couplage  $\kappa_{c-rc}$ ). Nous avons observé un profil en "U", même sans la rétroaction de la cible vers la région de contrôle (mais avec des amplitudes différentes du cas EI).

Dans ce scénario, la région cible se comporte au régime stationnaire comme une région isolée avec des valeurs particulières de forçage  $x_I$  appliquées sur son pool inhibiteur (cf figure 13.14).

---

23. Nous avons l'expression analytique de  $S_{s_{n_c/B_{Erc}}}$  dans l'article du chapitre 9.

La formulation au régime stationnaire du modèle de la région isolée se trouve dans l'annexe 18.1.1 ou à la section 2 des annexes de l'article du chapitre 9). Nous notons  $x_{B,I}$  la valeur du forçage sur le pool inhibiteur de la région au repos, et  $x_{stim,I}$  en période de stimulation. Nous pouvons écrire qu' au repos :  $x_{B,I} = \kappa_{c-rc} sn_{B,rc}^{*\text{dec}} + B_{I_c}$  et qu'en période de stimulation :  $x_{stim,I} = \kappa_{c-rc} sn_{stim,rc}^{*\text{dec}} + B_{I_c}$ . Entre le repos et la stimulation, le forçage sur le pool inhibiteur de la région cible varie alors de :  $\Delta x_I = x_{stim,I} - x_{B,I} = \kappa_{c-rc} (sn_{stim,rc}^{*\text{dec}} - sn_{B,rc}^{*\text{dec}})$ .

Plus généralement, il serait intéressant de voir si nous retrouvons un profil de courbe en "U" pour une région isolée, en appliquant un créneau positif sur le forçage du pool inhibiteur ( $\Delta x_I$ ), et en regardant  $\Delta sn^*$  en fonction de  $J_{gaba}$ , pour différentes valeurs de  $\Delta x_I$  fixées.

### 13.2.3.2.5 Synthèse : comment traduire formellement l'augmentation de l'inhibition de la cible en fonction de la question ?

De manière générale, pour étudier l'effet sur-inhibiteur ou sous-inhibiteur d'une augmentation de  $J_{gaba_c}$ , nous avons montré qu'il est nécessaire de préciser deux variations d'activité à comparer :

- La première est toujours celle de référence avant toute augmentation de  $J_{gaba_c}$  :  $\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}})$ .
- La deuxième, soit  $\Delta sn_c^*$ , soit  $\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}})$ , dépend de la question de recherche et plus précisément de la variable d'observation sur laquelle nous souhaitons étudier l'effet d'une augmentation de  $J_{gaba_c}$ .

Nous récapitulons ici les différents scénarios que nous avons exposés.

#### Premier cas : la traduction immédiate sans et avec rétroaction pour prédire l'effet de l'augmentation de $J_{gaba_c}$ (cf 13.2.3.2.1 et 13.2.3.2.2)

Tout ce qui suit est valable sans la rétroaction de la cible vers la région de contrôle ou avec la rétroaction (i.e. deux régions connectées en EI). Nous ne faisons pas de distinction de notation avec l'exposant <sup>dec</sup>, comme au paragraphe 13.2.3.2.1.

Nous partons de  $sn_c^*$  au repos pour  $J_{gaba_c} = \widehat{J_{gaba_c}}$ , c'est-à-dire de  $sn_c^* = f_c(\widehat{B_{E_{rc}}}, \widehat{J_{gaba_c}})$ .

Avec le premier chemin :

Nous faisons varier  $B_{E_{rc}}$  entre  $\widehat{B_{E_{rc}}}$  et  $\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}}$  : l'activité de la cible varie de  $\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}})$ . Puis nous faisons varier  $J_{gaba_c}$  entre  $\widehat{J_{gaba_c}}$  et  $\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}$  : l'activité de la cible varie en plus de  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}})$ . La variation totale vaut  $\Delta sn_c^*$  (cf figures 13.15 et 13.18).

L'effet de l'augmentation de  $J_{gaba_c}$  sur  $\Delta sn_c^*$  est déterminé par la deuxième contribution, i.e.  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{E_{rc}}} + \Delta \widehat{B_{E_{rc}}})$ . Pour étudier cet effet, nous avons regardé de manière plus

précise le signe de cette contribution. Autrement dit, pour identifier l'effet de l'augmentation de  $J_{gaba_c}$  sur la variation totale d'activité de la cible, nous avons étudié plus précisément l'effet d'une augmentation de  $J_{gaba_c}$  sur l'activité de la cible lorsque la région de contrôle est stimulée (représentée sur les courbes oranges des figures 13.15 et 13.18).

Cela est revenu à définir de manière explicite :

- le paramètre qu'on fait varier ( $\pi$ ) :  $J_{gaba_c}$
- l'intervalle de variation :  $[\widehat{J_{gaba_c}}; \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}]$
- la variable d'observation ( $\eta^*$ ) : activité de la cible pour une valeur fixée du forçage dans la région de contrôle, en période de stimulation :  $B_{Erc} = \widehat{B_{Erc}} + \Delta\widehat{B_{Erc}} = \widehat{B_{Erc}} + I_{stim}$ . Cette activité est notée :  $sn_c^* = f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, J_{gaba_c})$ .
- la boucle de rétroaction ( $\mathcal{B}$ ) : 2 régions (sans rétroaction de la cible vers la région de contrôle ou connectées en EI)

La variation  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}})$  est négative. Nous déduisons que  $J_{gaba_c}$  a un effet *inhibiteur fini* sur l'activité de la cible en période de stimulation.

Une augmentation de  $J_{gaba_c}$  produit une baisse totale d'activité de la cible  $\Delta sn_c$  plus grande que la variation de référence<sup>24</sup> *du fait de l'effet inhibiteur fini de  $J_{gaba_c}$  sur l'activité  $sn_c^*$  de la cible en période de stimulation.*

#### Avec le second chemin

D'après l'annexe 18.4, en inversant l'ordre successif de la variation des paramètres (i.e. d'abord  $J_{gaba_c}$  puis  $B_{Erc}$ ), nous montrons que l'effet de l'augmentation de  $J_{gaba_c}$  sur  $\Delta sn_c^*$  est déterminé par la première contribution :  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}})$  (cf figure 13.19 à droite).

Etudier l'effet d'une augmentation de  $J_{gaba_c}$  sur  $\Delta sn_c^*$ , revient donc à étudier uniquement le signe de la variation d'activité de la cible due à la variation de  $J_{gaba_c}$  lorsque la région de contrôle est au repos. Pour cela, nous pouvons reprendre les mêmes éléments du quadrupet précédent, à l'exception de la variable d'observation ( $\eta^*$ ) qui devient : l'activité de la cible pour une valeur fixée du forçage dans la région de contrôle, au repos :  $B_{Erc} = \widehat{B_{Erc}}$ . Cette activité est notée  $sn_c^* = f_c(\widehat{B_{Erc}}, J_{gaba_c})$ .

La variation  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}})$  est négative. Nous déduisons que  $J_{gaba_c}$  a un effet *inhibiteur fini* sur l'activité de la cible au repos.

Une augmentation de  $J_{gaba_c}$  produit une baisse totale d'activité de la cible  $\Delta sn_c$  plus grande que la variation de référence *du fait de l'effet inhibiteur fini de  $J_{gaba_c}$  sur l'activité  $sn_c^*$  de la cible au repos.*

---

24. i.e.  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$

### Conclusion :

L'effet d'une augmentation de  $J_{gaba_c}$  sur la variation totale d'activité de la cible  $\Delta sn_c^*$  est déterminé par l'effet d'une augmentation de  $J_{gaba_c}$  sur l'activité de la cible  $sn_c^*$  pour une valeur fixée du forçage dans la région de contrôle (en période de stimulation pour le premier chemin, ou hors stimulation pour le second chemin). C'est la variable d'observation implicite à partir de laquelle nous avons prédit l'effet de sur-inhibition de  $J_{gaba_c}$  : **l'activité de la cible au régime stationnaire ( $sn_c^*$ ) pour une valeur fixée du forçage dans la région de contrôle ( $B_{Erc}$ )**.

#### Quelle est la question ?

*En choisissant  $\Delta sn_c^*$ , à comparer à  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ , nous avons étudié l'effet d'une augmentation de  $J_{gaba_c}$  sur l'activité de la cible pour une valeur fixée du forçage  $B_{Erc}$  dans la région de contrôle.*

*$\eta^* = sn_c^* = f_c(B_{Erc}, J_{gaba_c})$  est la variable d'observation sur laquelle nous souhaitons étudier l'effet de l'augmentation de  $J_{gaba_c}$ .*

La première traduction immédiate portait sur un système découplé sans rétroaction de la cible vers la région de contrôle (cf 13.2.3.2.1). Nous avons montré que l'effet sur-inhibiteur de  $J_{gaba_c}$ , prédit dans le système découplé, est conservé même avec la rétroaction de la cible vers la région de contrôle (cf 13.2.3.2.2).

L'effet sur-inhibiteur de  $J_{gaba_c}$  dans ce contexte :

- permet de conclure que toute chose égale par ailleurs, l'augmentation de  $J_{gaba_c}$  chez le participant 2 rend plus faible l'activité de la cible  $sn_c^*$  par rapport au participant 1 en période de stimulation ( $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}) < 0$ ) et au repos ( $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}}) < 0$ ).
- ne nous dit rien sur l'effet sur-inhibiteur ou sous-inhibiteur de  $J_{gaba_c}$  quand nous considérons une variation  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$  par participant.

### **Deuxième cas : prendre en compte des mesures entre participants (13.2.3.2.3)**

#### Quelle est la question ?

*En choisissant  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}})$ , à comparer à  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ , nous avons étudié comment la variation d'activité de la cible due à une même stimulation de la région de contrôle, dépend d'un écart de  $J_{gaba_c}$ .*

*$\eta^* = \Delta sn_{c/B_{Erc}}(J_{gaba_c})$  est la variable d'observation sur laquelle nous souhaitons étudier l'effet de l'augmentation de  $J_{gaba_c}$ .*

### 13.2.4 Bilan

Au regard du titre de ma thèse,

*Etude de la balance Excitation/Inhibition de régions cérébrales impliquées dans une tâche de contrôle inhibiteur : modélisation de données obtenues en Imagerie par Résonance Magnétique fonctionnelle et inversion*

une grande partie de mon travail de physicienne en contact avec des neuroscientifiques expérimentateurs a consisté à préciser et traduire une question posée depuis les neurosciences expérimentales en questions posées depuis un modèle proposé par la communauté des neurosciences théoriques.

Pour étudier l'effet d'une augmentation de  $J_{gaba_c}$  sur l'inhibition de l'activité de la cible, le choix de la variable d'observation est important et dépend de la question de recherche. Le rendre explicite a constitué une étape importante du travail de précision et de traduction de la question posée depuis les neurosciences expérimentales (en langage naturel) vers des questions posées depuis le modèle (en langage naturel et traduites en langage mathématique). J'ai montré que dans un système composé de deux régions couplées, l'architecture de connectivité EI permet un abaissement de l'activité de la cible quand la région de contrôle est stimulée. J'ai montré l'importance de garder explicite la boucle de rétroaction pour ne pas oublier l'effet retour de la cible vers la région de contrôle et j'ai explicité en quoi un raisonnement causal inspiré de la métaphore de la balance E/I ne correspond pas toujours aux effets non linéaires du modèle.





# Chapitre 14

## Perspectives

*Ce chapitre est divisé en deux parties. A la section 14.1 je partage plusieurs pistes pour poursuivre le travail de modélisation. Les perspectives de l'inversion sont données à la section 14.2. Je présente d'abord l'intérêt de la méthode actuelle pour étudier le contrôle inhibiteur des intrusions mnésiques dans le trouble de stress post-traumatique, ou pour l'étude en IRMf d'un autre type de contrôle inhibiteur (e.g. moteur). J'expose enfin les limites et perspectives de chaque étape de la procédure d'inversion présentée au chapitre 10.*

### 14.1 Modèle direct

#### 14.1.1 Extension à plus de régions

Dans un système composé de deux régions couplées, une région de contrôle et une région cible, nous avons construit analytiquement les sensibilités des variables d'état au forçage sur le pool exciteur de la région de contrôle<sup>1</sup>.

Il serait intéressant d'augmenter le nombre de régions en considérant deux sous-systèmes de régions : un sous-système de régions de contrôle et un sous-système de régions cibles (e.g. onze régions de contrôle et quatre régions cibles comme dans le chapitre 11). La méthode d'analyse par emboîtement, utilisée au chapitre 9, pourrait aider à construire les sensibilités des variables d'état dans toutes les régions par rapport aux forçages sur les pools excitateurs des régions de contrôle.

Par rapport à deux régions, le passage à deux sous-systèmes fait intervenir en plus la dépendance à une matrice de connectivité structurale qui pondère les couplages entre régions. Ce travail analytique pourrait éclairer la question des connexions structurales les plus importantes dans la communication entre régions de contrôle et régions cibles.

#### 14.1.2 Explorer d'autres architectures de connectivité

Dans notre nouveau modèle de connectivité (chapitre 9), une fibre part du pool exciteur d'une région  $j$  et peut se connecter soit au pool exciteur soit au pool inhibiteur d'une autre région  $i$ , de manière exclusive. Parmi le nombre total de fibres qui partent de la région  $j$  vers

---

1. en particulier les sensibilités des  $sn^*$ , mais nous avons également celles des  $sg^*$

la région  $i$ ,  $k_{E_{ij}}$  désigne la fraction de fibres qui connectent le pool excitateur de la région  $i$ . Par définition,  $k_{E_{ij}}$  appartient à l'intervalle  $[0; 1]$ .

Nous avons étudié le cas particulier de deux régions et avons écrit les sensibilités de  $sn_{rc}^*$  et  $sn_c^*$  par rapport au forçage  $B_{E_{rc}}$  pour toutes valeurs de  $k_{E_{rc-c}}$  et  $k_{E_{c-rc}}$  dans  $[0; 1]$ . Nous avons ensuite choisi de fixer  $k_{E_{rc-c}}$  et  $k_{E_{c-rc}}$  en prenant les bornes de cet intervalle. Nous avons ainsi pu étudier les quatre architectures de connectivité présentées dans l'article :

E-E ( $k_{E_{rc-c}} = 1, k_{E_{c-rc}} = 1$ ), E-I ( $k_{E_{rc-c}} = 1, k_{E_{c-rc}} = 0$ ),

I-E ( $k_{E_{rc-c}} = 0, k_{E_{c-rc}} = 1$ ) et I-I ( $k_{E_{rc-c}} = 0, k_{E_{c-rc}} = 0$ ).

Il serait intéressant d'explorer d'autres valeurs intermédiaires de  $k_{E_{rc-c}}$  et  $k_{E_{c-rc}}$  qui permettraient d'avoir une baisse de l'activité de la région cible quand l'activité de la région de contrôle augmente. Une structure de connectivité dans laquelle des faisceaux de fibres partant de  $j$  se partageraient entre pools excitateur et inhibiteur de la région  $i$  semble être un scénario moins caricatural et sûrement plus pertinent pour décrire le système biologique.

### 14.1.3 Modéliser d'autres conditions expérimentales

Le modèle neural de contrôle inhibiteur proposé dans la thèse permet d'étudier la condition No-Think dans une tâche de suppression mnésique. Il serait intéressant de modéliser d'autres conditions expérimentales et le travail présenté dans cette thèse donne quelques pistes.

#### 14.1.3.1 Think

La condition Think est caractérisée par une augmentation de l'activité des régions cibles<sup>2</sup> (cf annexe 15.1 figure 15.3). Par ailleurs, [Schmitz et al., 2017] soulignent le rôle de la balance E/I sur les variations d'activité BOLD de l'hippocampe pendant la condition Think. Chez des participants sains, plus la concentration de GABA dans l'hippocampe (mesurée par spectroscopie au repos) est faible et plus l'augmentation d'activité de l'hippocampe pendant la condition Think est grande.

Grâce au travail de thèse, nous savons que pour n'importe quelle architecture de connectivité :

- une augmentation du forçage  $B_{E_c}$  dans la région cible produit une augmentation de l'activité de la cible. En effet, nous pouvons toujours nous ramener à un cas symétrique de stimulation de la région de contrôle et les résultats du chapitre 9 permettent de déduire le comportement du modèle pour une stimulation de la cible.

Comme pour le contrôle inhibiteur, des valeurs différentes et fixes entre participants de  $J_{gaba_c}$  pourraient être à l'origine d'une variabilité interindividuelle de l'augmentation d'activité des cibles en condition Think.

- une diminution infinitésimale de  $J_{gaba_c}$  (i.e.  $\delta J_{gaba_c} < 0$ ) produit une augmentation infinitésimale de l'activité de la cible (i.e.  $\delta sn_c / J_{gaba_c} > 0$ ), d'après l'annexe 18.3. Une variation négative de  $J_{gaba_c}$ , modélisée au cours du temps par un créneau descendant

---

2. En toute rigueur, ce vocabulaire est uniquement valable dans le cas du contrôle inhibiteur. Nous désignons ici les régions, précédemment identifiées comme "cibles" quand nous parlions de contrôle inhibiteur.

par exemple, permet donc d'obtenir une variation positive de l'activité de la cible ( $\Delta sn_c / J_{gaba_c} > 0$ ).

Ceci modéliserait une levée d'inhibition dans la région cible (baisse des poids synaptiques GABAergiques entre les interneurons inhibiteurs et les neurones excitateurs), conduisant à une augmentation de l'activité de la cible.

Des différences de poids inhibiteur dans l'hippocampe au repos pourraient être prises en compte dans le modèle en considérant différentes valeurs basales de  $J_{gaba_c}$ , qui subirait une variation pendant une stimulation Think.

### 14.1.3.2 No Think intrusive

Par rapport à la condition No Think non intrusive, la condition No Think intrusive conduit à une augmentation supplémentaire d'activité de la région de contrôle et à une diminution supplémentaire d'activité de la région cible (cf annexe 15.1 et section 5.5.2). [Benoit et al., 2015] indiquent qu'une activation de l'hippocampe pourrait être due à une intrusion, et que l'hippocampe transmettrait un signal à la région de contrôle pour augmenter davantage son activité. [Levy and Anderson, 2012] expliquent que ces dynamiques (d'abord l'activation de l'hippocampe au moment d'une intrusion, puis le recrutement de la région de contrôle) ne seraient pas mesurables étant donné la fréquence d'échantillonnage des mesures BOLD (cf section 4.3.5). Nous constaterions alors uniquement le résultat produit : augmentation accrue de l'activité de la région de contrôle et baisse accrue de l'activité de la cible.

Si nous partons du modèle de contrôle inhibiteur présenté dans la thèse, dans le cas d'une connectivité EI, la stimulation de la région de contrôle (se traduisant par une variation du forçage  $B_{Erc}$ ) produit une *excitation* de l'activité de la région de contrôle et une *inhibition* de l'activité de la cible. Ce scénario pourrait constituer un modèle pour la condition No Think non intrusive, l'activité de la cible maintenue suffisamment basse ne conduisant pas à la survenue d'une intrusion.

Pour modéliser la condition No Think intrusive, nous pourrions ajouter à cette stimulation de la région de contrôle, la variation concomitante d'un autre paramètre  $\pi_{int_c}$  dans la cible (modélisant une intrusion). Puisqu'il ne serait pas possible de séparer avec l'IRMf les effets produits par deux variations se produisant l'une après l'autre, cette hypothèse de modélisation peut faire sens. La variation supplémentaire de ce paramètre  $\pi_{int_c}$  provoquerait une *sur-excitation* de la région de contrôle et une *sur-inhibition* de la région cible, pour lesquelles nous pouvons proposer une traduction formelle, grâce au travail d'analyse mené à la section 13.2.3.2.

Notons  $\Delta sn_{rc}^*$  (resp.  $\Delta sn_c^*$ ) la variation totale de l'activité de la région de contrôle (resp. de la cible) qui résulte de la variation de ces deux paramètres. Elle peut s'écrire comme une somme de deux variations :  $\Delta sn_i^* = \Delta sn_{i/B_{Erc}}(\widehat{\pi_{int_c}}) + \Delta sn_{i/\pi_{int_c}}(\widehat{B_{Erc}} + I_{stim})$ <sup>3</sup>. Nous voudrions que la variation de  $\pi_{int_c}$  en même temps qu'une stimulation de la région de contrôle conduise à :

- augmenter la variation totale d'activité de la région de contrôle par rapport à celle

3. Nous pouvons nous servir du travail présenté en annexe 18.4 (avec un autre choix que  $J_{gaba_c}$  pour le second paramètre), pour écrire formellement les variations totales d'activité  $\Delta sn_i^*$

uniquement due à sa stimulation :  $0 < \Delta sn_{rc/B_{Er_c}} < \Delta sn_{rc}^*$ . Pour cela, il faudrait que  $\Delta sn_{rc/\pi_{int_c}} > 0$

- diminuer la variation totale d'activité de la cible par rapport à celle uniquement due à la stimulation de la région de contrôle :  $0 > \Delta sn_{c/B_{Er_c}} > \Delta sn_c^*$ . Autrement dit, nous voudrions que  $\Delta sn_{c/\pi_{int_c}} < 0$ .

### 14.1.3.3 Les pistes avec DCM

Nous notons que dans DCM, une condition expérimentale est modélisée par des variations des coefficients de couplage (en plus des variations de forçages). Nous pourrions nous référer au paragraphe 5.3.1.2. Il pourrait être intéressant d'explorer dans notre modèle neural, l'effet d'une modification des valeurs des coefficients de couplages entre régions pour une architecture de connectivité donnée.

### 14.1.4 Un modèle neural dynamique ?

Nous avons décrit au chapitre 3 deux types d'inhibition entre neurones : phasique et tonique. La seconde assure une transmission plus lente de l'influx nerveux, de l'ordre de la dizaine de secondes, et serait particulièrement impliquée dans les processus de contrôle inhibiteur (section 3.2.3).

Nous nous sommes basés sur le modèle neural proposé par [Naskar et al., 2021], qui présente des temps d'installation au régime stationnaire de l'ordre de la seconde, très courts devant les durées des stimulations et des temps d'installation du BOLD. Ceci a d'ailleurs permis de formuler le modèle neural directement au régime stationnaire (cf chapitre 7).

L'étude des temps caractéristiques dans une région (cf article chapitre 9) permet d'identifier des paramètres du modèle pour augmenter la durée d'installation des variables d'état au régime stationnaire et passer à un modèle avec de la dynamique. Un modèle neural dynamique, comme c'est le cas pour le modèle utilisé dans DCM, peut sembler plus pertinent pour rendre compte des phénomènes biologiques en lien avec l'inhibition tonique. Il serait de plus possible de modéliser, au cours du temps, une augmentation d'activité de la cible (pendant une intrusion) qui causerait ensuite une augmentation d'activité de la région de contrôle (recrutement du contrôle) puis une diminution d'activité de la cible.

### 14.1.5 Quel forçage en entrée du modèle hémodynamique ?

Dans le modèle hémodynamique utilisé dans la thèse (cf 6.3), l'activité neurale excitatrice  $sn(t)$  de chaque région permet d'appliquer un forçage sur les variables hémodynamiques (comme pour [Naskar et al., 2021] et [Wang et al., 2019]). L'étude du passage entre les niveaux de description neurale et BOLD a permis de stationnariser le modèle neural :  $sn^*$  sert de forçage hémodynamique.

Le modèle BOLD, permet d'écrire pour chaque région une observable  $y(t)$  en fonction des variables hémodynamiques. Pour chaque valeur de l'activité  $sn^*$  entre 0 et 1, il existe un unique point fixe  $y^*$ , entre 0 et 0.04, vers lequel tend le signal BOLD  $y(t)$ . Quand  $sn^*$  diminue,

le signal BOLD  $y(t)$  diminue, tout en restant positif<sup>4</sup>. Quand nous reprenons l'historique qui conduit à la formulation de ce modèle BOLD (cf section 5.3.2.2),  $y$  semble modéliser non pas un signal BOLD mais plutôt une variation du signal BOLD par rapport au repos, normalisée par rapport au repos.

Si nous conservons l'idée qu'à une valeur de  $sn^*$  correspond une valeur du signal BOLD, il est possible de se demander si le forçage en entrée du modèle hémodynamique ne devrait pas plutôt s'exprimer comme une variation de  $sn^*$  par rapport au repos, normalisée par rapport au repos :  $\frac{sn_{stim}^* - sn_B^*}{sn_B^*}$ . [Stephan et al., 2004] soulignent que la nature de l'activité neurale dans le modèle hémodynamique proposé par [Friston et al., 2000] n'est pas précisée, notamment parce que la manière dont l'activation des neurones déclenchent des changements hémodynamiques est une question ouverte.

### 14.1.6 Prendre en compte des régions d'une autre nature que les régions de contrôle et les régions cibles : les régions relais

Une tâche de contrôle inhibiteur implique des variations d'activité dans des régions de contrôle et des régions cibles, mais aussi dans des régions de relais (principalement sous-corticales, cf section 2.3). Nous avons modélisé le comportement d'un système de deux régions couplées, composé d'une région de contrôle et d'une région cible. Une étape supplémentaire du travail pourrait consister en l'ajout d'une région relais pour constituer un système de trois régions couplées.

Il serait intéressant de préciser au niveau biologique la différence de comportement entre une région relais et une région cible dans le cerveau, pour pouvoir modéliser le comportement de la région relais. Pour le moment, la région cible et le relais peuvent posséder la même propriété de n'être pas stimulées (contrairement à la région de contrôle). Nous proposons d'être indifférent au signe de la variation de la région relais, alors que nous souhaitons une variation d'activité négative dans la cible et positive dans la région de contrôle.

Nous pouvons étudier les variations d'activité de la région cible et de la région de contrôle due à une stimulation de la région de contrôle, à travers cette nouvelle boucle de rétroaction (caractérisée par l'ajout de la région relais). Ceci permettrait d'éclairer ce qu'apporte la présence de la région relais par rapport au système uniquement composé d'une région de contrôle et d'une cible. Dans quelle mesure le relais modifie la transmission entre la région de contrôle et la cible ?

Dans une approche analytique, nous pourrions utiliser la même méthode d'emboîtement pour exprimer dans le système couplé les sensibilités des  $sn_i^*$  (contrôle, cible et relais) par rapport au forçage  $B_{Er_c}$  de la région de contrôle en fonction des sensibilités des régions découplées<sup>5</sup>, pour n'importe quel type de connectivité entre les trois régions.

Nous pourrions étudier différentes architectures de connectivité pour connecter les régions entre elles de manière pertinente. Une analyse d'ordre de grandeur pourrait permettre de mieux comprendre dans quelle mesure les couplages avec la région relais modifient les valeurs des sensibilités couplées par rapport au cas de deux régions. Pour cela, il est nécessaire de

4. sauf éventuellement pendant un undershoot si  $y^*$  est proche de 0

5. on annule tous les couplages entre régions pour exprimer les sensibilités découplées en fonction des sensibilités des régions isolées évaluées à des valeurs particulières de forçages

se donner une matrice de connectivité structurale ( $C_{ij}$ ) pour fixer des coefficients de couplage ( $\kappa_{ij}$ ) entre contrôle-cible, contrôle-relais et cible-relais. Quelle matrice de connectivité structurale fait sens ? Quel est le poids relatif de ces coefficients de couplage ? Cette question n'émerge qu'à partir de trois régions puisque chaque élément  $C_{ij}$  de la matrice de connectivité structurale est le rapport du nombre de fibres entre les régions  $i$  et  $j$ , divisé par le plus grand nombre de fibres mesuré parmi les régions du système considéré. Dans le cas de deux régions nous avons seulement :  $C_{12} = C_{21} = 1$ .

### 14.1.7 Augmenter l'inhibition dans un contexte de contrôle exciteur

Nous avons considéré que pendant une tâche comme le Think/No Think, l'augmentation d'activité de la région de contrôle s'accompagne d'une baisse de l'activité dans la région cible. Nous nous sommes placés dans cette perspective pour commencer à travailler, sans chercher à apporter un éclairage supplémentaire pour savoir si les mesures BOLD effectuées au cours d'une telle tâche correspondent davantage à une baisse d'activité de la cible ou à une moindre excitation.

Depuis le modèle, nous avons défini le contrôle inhibiteur par la chaîne de causalité suivante où nous prenons en compte la rétroaction de la région cible vers la région de contrôle (cf 13.2.2) :

$$\text{Contrôle Inhibiteur : } I_{stim} \Rightarrow (\Delta sn_{rc}^* > 0, \Delta sn_c^* < 0) \quad (14.1)$$

Dans ce contexte, augmenter l'inhibition signifie augmenter la baisse d'activité de la cible, c'est-à-dire rendre sa variation  $\Delta sn_c^* < 0$  plus grande en valeur absolue.

Les résultats de la première étape de l'inversion sur un participant montrent  $\Delta sn_{rc}^* > 0$  et  $\Delta sn_c^* < 0$ , ce qui va dans le sens du contrôle inhibiteur tel que nous venons de le définir. Il est nécessaire de vérifier que ce résultat se conserve quand on augmente le nombre de participants.

Sans interroger la nature de ce que nous mesurons (baisse d'activité ou moindre excitation de la cible), il me paraît utile de définir et préciser ce que signifierait une moindre excitation dans le cadre de notre modèle. Pour cela, nous revenons d'abord sur la définition du *contrôle*, puis poursuivons avec celle de *contrôle exciteur*. Nous nous servirons ensuite du concept de *moindre excitation* pour traduire ce que signifie *augmenter l'inhibition dans le contexte du contrôle exciteur*. Ce travail s'inscrit dans la continuité des éléments développés au chapitre 13 pour le contrôle inhibiteur. Il pourra être utile dans la perspective d'étudier les données mesurées pendant la tâche sous l'angle d'une moindre excitation de l'activité des cibles.

#### 14.1.7.1 Contrôle

Le contrôle peut être défini par l'augmentation d'activité d'une région due à une variation du forçage sur son pool exciteur. Il peut être décrit par la chaîne de causalité :

$$\text{Contrôle : } I_{stim} \Rightarrow \Delta sn_{rc}^* > 0 \quad (14.2)$$

Il peut être de deux types : 'inhibiteur' ou 'excitateur'. Cette région, dite de contrôle, est couplée à une région cible. Le qualificatif 'inhibiteur' ou 'excitateur' fait référence au signe de la variation de l'activité de la région cible due à la variation du forçage dans la région de contrôle. Quand la variation d'activité de la région cible est négative ( $\Delta sn_{rc}^* < 0$ ), on parle de contrôle inhibiteur.

#### 14.1.7.2 Contrôle Excitateur

Le contrôle excitateur est défini en neurosciences par l'augmentation d'activité de la région cible due à l'augmentation d'activité de la région de contrôle<sup>6</sup>. On parle alors de "upregulation" de la région cible par la région de contrôle (i.e. augmentation de l'activité de la cible par rapport à un état de base). Comme pour la "downregulation", on ne pense pas nécessairement la rétroaction de la cible vers la région de contrôle. En prenant en compte la rétroaction de la région cible vers la région de contrôle, nous pouvons définir le contrôle excitateur par la chaîne de causalité suivante :

$$\text{Contrôle Excitateur : } I_{stim} \Rightarrow (\Delta sn_{rc}^* > 0, \Delta sn_c^* > 0) \quad (14.3)$$

#### 14.1.7.3 Définition et traduction du concept de moindre excitation dans le modèle

Une excitation est une variation positive de l'activité d'une région par rapport à un niveau d'activité de référence (au régime stationnaire). Elle se traduit par  $\Delta sn_i^* > 0$  pour une région  $i$ .

Une moindre excitation est une baisse de cette variation positive : elle reste positive mais devient plus petite en valeur absolue. Le concept de moindre excitation est donc une comparaison entre deux variations positives de l'activité d'une même région : celle d'intérêt plus petite qu'une autre de référence. Cette dernière n'est pas forcément toujours énoncée de manière explicite.

En toute généralité, cette baisse de la variation positive peut être envisagée comme le résultat d'une variation d'un ou deux paramètres, directement dans la région qui subit une moindre excitation (par un mécanisme interne à la région), ou dans une autre région avec laquelle elle est couplée (e.g. cas de 2 régions couplées, une région de contrôle et une région cible). C'est ce que nous détaillons ici.

##### 14.1.7.3.1 Cas de la variation d'un paramètre $\pi_E$

Supposons un paramètre  $\pi_E$  dans un intervalle  $\mathcal{E}$  ayant un effet excitateur infinitésimal sur l'activité  $sn_c^*$  de la cible, étant donnée une certaine boucle de rétroaction. Pour toute valeur de  $\pi_E$  dans  $\mathcal{E}$ , une perturbation  $\delta\pi_E$  cause une perturbation de l'activité de la cible  $\delta sn_{c/\pi_E}^* > 0$  : l'activité de la cible augmente. Nous pouvons écrire cette perturbation en fonction de la sensibilité de  $sn_c^*$  par rapport à  $\pi_E$ , qui est positive.

Supposons que nous faisons varier  $\pi_E$  entre  $\widehat{\pi_E}$  et  $\widehat{\pi_E} + \Delta\widehat{\pi_E}$  (contenus dans  $\mathcal{E}$ ). Nous avons vu que la variation d'activité de la cible  $\Delta sn_{c/\pi_E}$  peut s'écrire comme une intégrale de

---

6. et non due à la cible elle-même via des mécanismes neuraux internes à la région



la sensibilité de  $sn_c^*$  par rapport à  $\pi_E$ .  $\Delta sn_{c/\pi_E}$  est alors strictement positive est augmentée avec la borne supérieure de l'intégrale, i.e. avec  $\Delta \widehat{\pi_E}$ . Nous avons parlé d'un effet exciteur fini de  $\pi_E$  sur l'activité  $sn_c^*$  de la cible entre  $\widehat{\pi_E}$  et  $\widehat{\pi_E} + \Delta \widehat{\pi_E}$ , étant donné la boucle de rétroaction.

Une variation  $\Delta \widehat{\pi_E}$  plus petite du paramètre conduit bien à une variation plus petite de la variable d'observation ( $sn_c^*$ ).

Exemples :

- Dans la région de contrôle,  $I_{stim}$  plus faible<sup>7</sup> cause une variation positive plus faible de l'activité  $sn_{rc}^*$  de la région de contrôle. Ceci reste vrai si nous considérons la région isolée, découplée d'une cible (absence de rétroaction de la cible vers la région de contrôle) ou couplée avec une cible (pour n'importe quelle architecture de connectivité). Nous pouvons nous reporter à l'annexe 18.2, figures 18.6 et 18.5.
- Dans le cas d'une connectivité EE ou IE uniquement,  $I_{stim}$  plus faible dans la région de contrôle cause une variation positive plus faible de l'activité  $sn_c^*$  de la région cible.

#### 14.1.7.3.2 Cas de la variation de deux paramètres $\pi_E$ et $\pi_I$

Ce paragraphe est rédigé dans la continuité de la section 13.2.3.2. Nous exposons un cas général avec deux paramètres  $\pi_E$  et  $\pi_I$ . Pour faciliter la lecture, nous présentons un exemple dans un système de deux régions couplées en E-E. Nous avons représenté l'activité de la cible  $sn_c^*$ , en fonction de la stimulation de la région de contrôle ( $\pi_E \equiv B_{E_{rc}}$ ), pour différentes valeurs de  $\pi_I \equiv J_{gabac}$  (figures 14.1 et 14.2).

Supposons un paramètre  $\pi_E$  dans un intervalle  $\mathcal{E}$  ayant un effet exciteur infinitésimal sur l'activité  $sn_c^*$  de la cible, étant donnée une certaine boucle de rétroaction.

Toute variation de  $\pi_E$  entre  $\widehat{\pi_E}$  et  $\widehat{\pi_E} + \Delta \widehat{\pi_E}$  (contenus dans  $\mathcal{E}$ ) produit une variation positive de l'activité  $sn_c^*$  de la cible :  $\Delta sn_{c/\pi_E} > 0$ . On parle d'excitation de la cible. Par exemple, l'application de  $I_{stim} = \Delta \widehat{\pi_E} = 0.1$  conduit à augmenter  $sn_c^*$  (flèche rouge vers le haut sur la figure 14.1).

Supposons un paramètre  $\pi_I$  dans un intervalle  $\mathcal{J}$ , ayant un effet inhibiteur infinitésimal sur une variable d'observation ( $\eta$ , que nous allons préciser), étant donnée la même boucle de rétroaction. Pour toute valeur de  $\pi_I$  dans  $\mathcal{J}$ , une perturbation  $\delta \pi_I$  cause une perturbation de la variable d'observation  $\delta \eta / \pi < 0$  : la variable d'observation  $\eta$  diminue.

La variable d'observation  $\eta$  dépend de la question de recherche. Elle peut être :

- l'activité de la cible  $sn_c^*$ , évaluée pour une valeur particulière de  $\pi_E$  ( $\widehat{\pi_E}$  ou  $\widehat{\pi_E} + \Delta \widehat{\pi_E}$ )
- la variation d'activité de la cible due à la variation de  $\pi_E$  :  $\Delta sn_{c/\pi_E}$

##### 14.1.7.3.2.1 Activité de la cible $sn_c^*$

Nous faisons varier les deux paramètres  $\pi_E$  (entre  $\widehat{\pi_E}$  et  $\widehat{\pi_E} + \Delta \widehat{\pi_E}$  contenus dans  $\mathcal{E}$ ) et  $\pi_I$  (entre  $\widehat{\pi_I}$  et  $\widehat{\pi_I} + \Delta \widehat{\pi_I}$  contenus dans  $\mathcal{J}$ ) en même temps. Nous nous intéressons à la

---

7. La région s'active moins par elle-même

variation totale de l'activité de la cible  $sn_c^*$  due à ces deux variations, notée  $\Delta sn_c^*$ . Celle-ci s'écrit comme une somme de deux variations :

$$\Delta sn_c^* = \Delta sn_{c/\pi_E}(\widehat{\pi_I}) + \Delta sn_{c/\pi_I}(\widehat{\pi_E} + \Delta \widehat{\pi_E}) \quad (14.4)$$

$$= \Delta sn_{c/\pi_I}(\widehat{\pi_E}) + \Delta sn_{c/\pi_E}(\widehat{\pi_I} + \Delta \widehat{\pi_I}) \quad (14.5)$$

$\Delta sn_{c/\pi_E}(\pi_I)$  désigne la variation de l'activité de la cible uniquement due à la variation de  $\pi_E$ , évaluée pour une valeur particulière de  $\pi_I$ .  $\Delta sn_{c/\pi_I}(\pi_E)$  désigne la variation de l'activité de la cible uniquement due à la variation de  $\pi_I$ , évaluée pour une valeur particulière de  $\pi_E$ . Nous représentons sur la figure 14.1 le chemin correspondant à l'égalité (14.4). Il est possible de raisonner avec le deuxième chemin (cf 13.2.3.2).

$\pi_E$  a un effet excitateur sur  $sn_c^*$ <sup>8</sup> :  $\Delta sn_{c/\pi_E} > 0$  (flèche rouge sur la figure 14.1). Comme  $\pi_I$  a un effet inhibiteur sur  $sn_c^*$  :  $\Delta sn_{c/\pi_I} < 0$  (flèche violette sur la même figure). La variation de l'activité de la cible due à la variation  $\pi_I$  s'ajoute à l'excitation de la cible due à la variation de  $\pi_E$  et a pour effet de la réduire. La variation de  $\pi_I$  "freine" l'effet de l'unique variation de  $\pi_E$  sur l'activité de la cible. Si la somme de ces deux variations,  $\Delta sn_c$ , reste positive, nous avons toujours une excitation de la cible  $\Delta sn_c > 0$ , mais une moindre excitation<sup>9</sup> par rapport à celle obtenue uniquement avec la variation de  $\pi_E$  ( $\Delta sn_c < \Delta sn_{c/\pi_E}$ ). Par exemple sur la figure 14.1, la flèche verte est plus courte que la rouge.

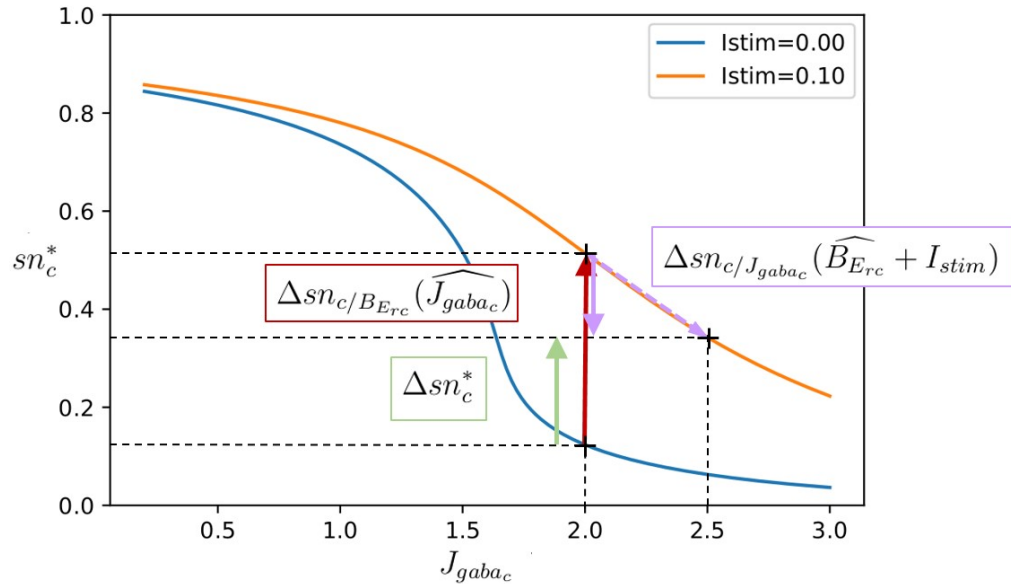


**Figure 14.1** – La variation de  $B_{Erc}$  conduit à une augmentation de l'activité de la région cible :  $\Delta sn_{c/B_{Erc}} > 0$ . On parle d'excitation de la région cible, représentée par le flèche rouge vers le haut. La variation supplémentaire de  $J_{gaba_c}$  réduit cette augmentation (flèche violette). On parle de moindre excitation.

Notons que la somme  $\Delta sn_c$  peut devenir négative, si  $\Delta sn_{c/\pi_I} < 0$  (flèche violette) compense  $\Delta sn_{c/\pi_E} > 0$  (flèche rouge). On ne parle plus de moindre excitation puisque la variation

8. pour  $\pi_I = \widehat{\pi_I}$  ou  $\pi_I = \widehat{\pi_I} + \Delta \widehat{\pi_I}$

9. ou une "sous-excitation", puisque nous avons employé "sur-inhibition" dans la section 13.2.3.2

résultante  $\Delta sn_c$  n'est alors plus positive.

#### 14.1.7.3.2.2 Variation d'activité de la cible due à la variation de $\pi_E$ : $\Delta sn_{c/\pi_E}$

Nous regardons comment la variation produite par  $\pi_E$ , i.e.  $\Delta sn_{c/\pi_E}(\pi_I)$ , dépend de la valeur de  $\pi_I$ . Comme  $\pi_E$  a un effet excitateur sur  $sn_c^*$  :  $\Delta sn_{c/\pi_E} > 0$  (flèches rouge et grise vers le haut sur la figure 14.2). Si augmenter  $\pi_I$  réduit  $\Delta sn_{c/\pi_E}$  (flèche grise plus courte que flèche rouge sur la figure 14.2), qui reste positive, on obtient une moindre excitation de la cible.

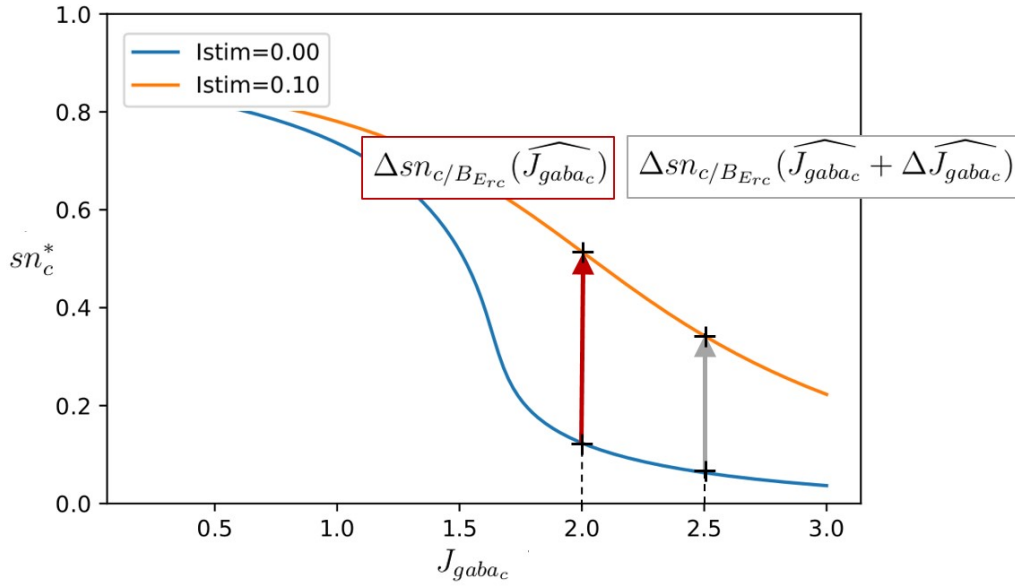


**Figure 14.2** – Comparaison de la variation d'activité de la cible due à la stimulation de la région de contrôle,  $\Delta sn_{c/B_{Erc}}(J_{gaba_c})$ , pour deux valeurs de  $J_{gaba_c}$ . La flèche rouge représente une excitation :  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) > 0$ . La seconde flèche grise est plus courte que la rouge : augmenter  $J_{gaba_c}$  par rapport à  $\widehat{J_{gaba_c}}$  a eu pour effet de réduire l'excitation. On parle de moindre excitation.

#### 14.1.7.4 Que signifie augmenter l'inhibition dans le contexte d'un contrôle excitateur ?

Dans le contexte du contrôle excitateur, augmenter l'inhibition de la cible signifie abaisser la variation positive de l'activité de la cible  $sn_c^*$  (i.e. moins augmenter). On peut alors parler d'abaissement de l'excitation ou de moindre excitation de l'activité de la cible.

Le contrôle excitateur est une pensée en variation du forçage de la région de contrôle ( $I_{stim}$ ). Pour augmenter l'inhibition de la cible dans ce contexte, nous n'allons pas juste abaisser  $I_{stim}$ , mais plutôt chercher à faire varier un autre paramètre du modèle. La moindre excitation pourrait alors être modélisée à partir du cas de la variation de deux paramètres expliquée ci-dessus.

#### 14.1.7.5 Comment le travail dans la thèse permet déjà d'explorer cette question ?

Dans la thèse, nous n'avons pas modélisé une moindre excitation de l'activité des cibles. Cependant, le travail présenté dans l'article permettrait d'explorer cette question.

##### 14.1.7.5.1 Un modèle du contrôle excitateur

En effet, si nous considérons une région de contrôle et une région cible connectée en EE, nous avons construit analytiquement toutes les sensibilités couplées par rapport à  $B_{Er_c}$ . Nous avons montré qu'elles sont positives. Pour différentes valeurs de  $J_{gaba_c}$  fixées, des variations du forçage  $B_{Er_c}$  donnent donc des variations d'activité de la région de contrôle  $\Delta sn_{rc/B_{Er_c}}(J_{gaba_c})$  et de la cible  $\Delta sn_{c/B_{Er_c}}(J_{gaba_c})$  également positives (constat fait par simulations numériques et confirmé par construction de la variation comme intégrale de la sensibilité). La connectivité EE permet alors de modéliser un contrôle excitateur.

##### 14.1.7.5.2 Effet d'une augmentation de $J_{gaba_c}$ sur $\Delta sn_{c/B_{Er_c}}$ dans le cas d'un contrôle excitateur

Nous avons ensuite étudié l'effet d'une augmentation de  $J_{gaba_c}$  dans la cible :

- paramètre perturbé ( $\pi$ ) :  $J_{gaba_c}$
- variable d'observation ( $\eta$ ) : variation d'activité de la cible due à une variation fixée du forçage<sup>10</sup> dans la région de contrôle, notée  $\Delta sn_{c/B_{Er_c}}(J_{gaba_c})$ .
- la boucle de rétroaction ( $\mathcal{B}$ ) : 2 régions couplées, une région de contrôle et une région cible pour une connectivité EE

Nous avons tracé en passant par des simulations le graphe de  $\Delta sn_{c/B_{Er_c}}(J_{gaba_c})$  en fonction de  $J_{gaba_c}$  et regardé les pentes des courbes (cf figure 17.C de l'article du chapitre 9).  $\Delta sn_{c/B_{Er_c}}(J_{gaba_c})$  reste positif, croît d'abord en fonction de  $J_{gaba_c}$  sur un intervalle  $\mathcal{E}$  puis décroît sur  $\mathcal{J}$ . Dans  $\mathcal{E}$ , augmenter  $J_{gaba_c}$  conduit à une sur-excitation alors que dans  $\mathcal{J}$ , augmenter  $J_{gaba_c}$  conduit à une moindre excitation (ou sous-excitation).

---

10. entre  $\widehat{B_{Er_c}}$  et  $\widehat{B_{Er_c}} + \Delta \widehat{B_{Er_c}} = \widehat{B_{Er_c}} + I_{stim}$

## 14.2 Inversion

### 14.2.1 Perspective d'application de la méthode en l'état actuel

#### 14.2.1.1 Etude de pathologies psychiatriques comme le Trouble de Stress Post-Traumatique

##### 14.2.1.1.1 Enjeux cliniques de l'étude du contrôle inhibiteur

Les capacités de contrôle inhibiteur sont étudiées dans diverses pathologies psychiatriques comme le trouble déficitaire de l'attention avec ou sans hyperactivité (TDA/H), schizophrénie, trouble du spectre autistique (TSA), addictions, trouble obsessionnel compulsif (TOC), trouble de stress post-traumatique (TSPT). En laboratoire, des déficits ont été mesurés dans des tâches standards de contrôle inhibiteur, chez des groupes de patients par rapport à des groupes de sujets contrôles. Plus généralement, certains symptômes caractéristiques de ces pathologies résulteraient d'un déficit de contrôle inhibiteur ([Bari and Robbins, 2013]). Les pensées intrusives font notamment partie des symptômes de la schizophrénie (hallucinations), du TOC (obsessions), de la dépression (ruminations) ou du TSPT (intrusions mnésiques) ([Schmitz et al., 2017], [Anderson et al., 2016]).

Améliorer notre compréhension des mécanismes cérébraux qui sous-tendent le contrôle inhibiteur, dans les cas sains et pathologiques, pourrait permettre à plus long terme d'améliorer le diagnostic, les traitements et la prévention de ces pathologies ([Bari and Robbins, 2013]).

##### 14.2.1.1.2 Symptômes intrusifs dans le Trouble de Stress Post-Traumatique

Après exposition à un événement traumatique, un individu peut présenter un ensemble de symptômes caractéristiques d'un état de stress aigu. Cette réaction physiologique et psychologique est considérée comme normale jusqu'à un mois après l'exposition. Au-delà d'un mois, elle devient pathologique et le diagnostic est remplacé par celui d'un trouble de stress post-traumatique (TSPT). Celui-ci peut apparaître à tout âge, y compris durant l'enfance. Le manuel diagnostique et statistique des troubles mentaux (DSM-5) définit les critères actuellement admis pour le diagnostic d'un trouble de stress post-traumatique<sup>11</sup>. Les intrusions constituent le symptôme prédominant dans la pathologie, qui maintient et aggrave les autres symptômes : les patients revivent de manière récurrente, involontaire et impuissante certains éléments vifs et détaillés de l'événement traumatique.

Ces intrusions s'accompagnent d'un sentiment de détresse avec des réactions émotionnelles (peur, horreur, grande tristesse, culpabilité, honte, colère...) et/ou physiques intenses (difficultés à respirer, rythme cardiaque accéléré, vertiges, transpiration...), conférant aux symptômes intrusifs un caractère fortement envahissant dans le quotidien des patients. Elles peuvent se produire lors de l'exposition à des indices internes (pensées, émotions) ou externes (caractéristiques de l'environnement) qui rappellent un aspect de l'événement traumatique.

---

11. Les modalités d'expositions ainsi que quatre familles de symptômes y sont précisées : symptômes intrusifs, d'évitement, d'hyper-réactivité ainsi que des altérations négatives de la cognition et de l'humeur. La personne doit présenter des symptômes dans chacune des quatre familles.

Le patient peut vivre des réactions dissociatives (flashbacks) pendant lesquelles il se sent et/ou agit comme si l'événement traumatique se reproduisait ou allait se reproduire. De telles réactions peuvent survenir sur un continuum, l'expression la plus extrême étant une perte totale de la conscience de l'environnement au moment présent. Enfin des intrusions peuvent se manifester sous forme de cauchemars répétitifs dans lesquels le contenu et/ou l'affect sont liés à l'événement traumatique.

D'abord uniquement étudié comme un trouble de la mémoire ([Brewin et al., 2010]), le TSPT pourrait aussi être un trouble du contrôle inhibiteur de la mémoire ([Catarino et al., 2015], [Mary et al., 2020], [Leone et al., 2022]), les patients ne parvenant pas à réguler la survenue d'intrusions mnésiques<sup>12</sup>.

Des déficits de contrôle inhibiteur au niveau comportemental et cérébral sont mesurés chez des patients avec TSPT pendant une tâche de Think/No Think ([Mary et al., 2020], [Leone et al., 2022]).

Par ailleurs, [Rosso et al., 2017] mesurent par MRS une concentration plus élevée en glutamate dans l'hippocampe droit chez des patients avec TSPT par rapport à des sujets contrôles. Ces mesures sont mises en corrélation avec les symptômes intrusifs. Plus la concentration en glutamate hippocampique est élevée et plus les symptômes intrusifs sont importants. Cette mesure élevée semble indiquer une modification de l'état de la balance E/I dans l'hippocampe chez les patients TSPT.

#### 14.2.1.1.3 Liens avec le travail présenté dans la thèse

Sous l'hypothèse de notre modèle, nous avons vu que les propriétés E/I locales (représentées par les paramètres  $I_{stim_i}$  et  $J_{gaba_i}$ ) influencent l'effet inhibiteur de régions de contrôle sur des régions cibles pendant une tâche de contrôle inhibiteur (i.e. l'amplitude de la baisse d'activité des régions cibles). Pour mieux comprendre si des modifications de ces propriétés E/I peuvent être en lien avec un dysfonctionnement du contrôle inhibiteur de la mémoire, il serait alors intéressant de poursuivre le travail d'inversion sur les données de tous les participants de l'étude REMEMBER, puis de conduire des analyses statistiques pour comparer les valeurs des paramètres estimés entre les groupes (contrôles, exposés avec et sans TSPT).

#### 14.2.1.2 D'autres types de données

Nous avons mené une inversion avec des données BOLD mesurées pendant une tâche de contrôle inhibiteur des intrusions mnésiques. Nous aurions aussi pu travailler avec des données qui mesurent une autre catégorie de contrôle inhibiteur (cf chapitre 2, section 2.2), par exemple le contrôle inhibiteur moteur avec une tâche de go/no go.

[Falconer et al., 2008] mesure une moindre augmentation d'activité dans des régions de contrôle préfrontales pendant une tâche de go/no go chez des sujets avec TSPT. A plus long

---

12. A l'inverse, il est intéressant de noter qu'un contrôle inhibiteur exacerbé de la mémoire après un événement traumatique pourrait être lié au développement d'une amnésie dissociative ([Anderson and Hanslmayr, 2014])

terme, utiliser des données mesurant différents types de contrôle inhibiteur sur des populations pathologiques permettrait d'étudier s'il existe une atteinte plus globale des mécanismes de contrôle inhibiteur dans ces pathologies.

### 14.2.2 Première étape de la procédure d'inversion : variabilité de la réponse hémodynamique ?

Nous avons fixé les valeurs des paramètres contenus dans les modèles hémodynamiques et BOLD de manière identique dans toutes les régions.

Le modèle hémodynamique est un modèle déterministe (sans bruit). La réponse hémodynamique à une même activité neurale ne varie donc pas d'une région à l'autre. Si nous inversons les données d'autres participants, cette réponse ne variera pas non plus entre les participants. Nous n'avons pris en compte aucune variabilité de la réponse hémodynamique.

Le modèle BOLD est bruité. En l'absence du bruit, la réponse BOLD est aussi invariable entre régions et entre participants.

Ce choix présente des limites puisque la variabilité des réponses entre régions et entre participants a été rapportée dans plusieurs études ([Buckner et al., 1996], [Aguirre et al., 1998], [Griffeth et al., 2011], [Buxton, 2012]). La toolbox VBA que nous avons utilisée propose d'ailleurs déjà la possibilité d'inverser certains paramètres hémodynamiques et BOLD ( $\alpha$ ,  $\gamma$ ,  $\tau$ ,  $\rho$ ,  $\epsilon$ ). Nous avons néanmoins travaillé dans un premier temps sans considérer cette variabilité et nous justifions ce choix.

La procédure d'inversion nous permet d'estimer pour chaque région  $i$  les  $sn_{B_i}^*$  et  $sn_{stim_i}^*$  (première étape) et les paramètres neuraux  $I_{stim_i}$  et  $J_{gaba_i}$  (deuxième étape). Notre objectif est de pouvoir poser une interprétation robuste des résultats en sortie d'inversion, c'est-à-dire savoir dire comment les valeurs des paramètres inversés ont un effet sur le comportement du signal BOLD. Pour cela, nous avons évidemment besoin de comprendre en amont comment varie le signal BOLD en fonction des paramètres que l'on inverse.

En fixant dans chaque région tous les paramètres BOLD sauf  $sn^*$ , nous savons qu'une augmentation de  $sn^*$  est liée à une augmentation du BOLD au régime stationnaire  $y^*$  et inversement (cf section 7.1.2). Pour chaque région, nous savons alors interpréter une relation d'ordre/différence entre les valeurs inversées de  $sn_B$  et  $sn_{stim}$  en terme de variation du signal BOLD. Nous pouvons même envisager aisément des comparaisons de  $sn_B$  et  $sn_{stim}$  entre régions puisque la relation bijective entre  $sn^*$  et  $y^*$  est identique entre les régions.

Dans le cas où nous ne fixons pas les autres paramètres hémodynamiques et BOLD, nous ajoutons des degrés de libertés pour la variation du signal BOLD. Une valeur plus élevée de  $sn^*$  ne correspond plus nécessairement à une valeur plus grande du signal BOLD au régime stationnaire  $y^*$ . L'interprétation des effets que produisent des variations de  $sn^*$  sur le signal BOLD  $y^*$  est possiblement plus complexe et demande à être étudiée pour que nous puissions interpréter une différence entre les valeurs inversées de  $sn_B$  et  $sn_{stim}$  en terme de variation du signal BOLD.

Nous remontons ensuite aux paramètres neuraux  $I_{stim_i}$  et  $J_{gaba_i}$  à partir des valeurs de  $sn_B$  et  $sn_{stim}$  dans toutes les régions. Pour pouvoir interpréter l'effet des valeurs des paramètres neuraux inversés sur le signal BOLD, nous avons besoin de comprendre comment les  $sn_B$



et  $sn_{stim}$  se comportent en fonction des paramètres neuraux (fait uniquement pour deux régions), et à nouveau comment le signal BOLD se comporte en fonction des  $sn_B$  et  $sn_{stim}$ . En conclusion, en l'état actuel de nos connaissances sur le comportement du modèle, l'inversion de paramètres hémodynamiques et BOLD rend incompréhensibles les résultats en sortie d'inversion, d'où notre choix de ne pas prendre en compte pour le moment la variabilité des réponses hémodynamiques et BOLD. Dans ce sens, nous pouvons citer [Buxton, 2012] : "*a very detailed model may become less useful for modeling data because so much that is included in the model is not well understood.*"

Pour pouvoir interpréter les  $sn_{B_i}$ ,  $sn_{stim_i}$ , et en particulier  $I_{stim_i}$  et  $J_{gaba_i}$  en sortie d'inversion, tout en estimant des paramètres hémodynamiques et BOLD supplémentaires, il semble alors nécessaire d'analyser le comportement du modèle BOLD pour une région quand on intègre la variabilité de ces nouveaux paramètres. Plus précisément, pour  $sn^* \in [0; 1]$ , quelle variation du signal BOLD  $y^*$  produit la variation de ces nouveaux paramètres ? Est-ce que la variation de  $y^*$  produite par la variation de ces paramètres est toujours plus petite que celle produite par une variation de  $sn^*$  ? Si non, est-ce que la variation de  $y^*$  produite par la variation de ces paramètres est plus petite que celle produite par la différence entre  $sn_B$  et  $sn_{stim}$  inversés ? Si c'est le cas, alors étant donné les valeurs de  $sn_B$  et  $sn_{stim}$  estimées, une augmentation de  $sn^*$  entre le repos et la stimulation correspond bien à une augmentation de la réponse BOLD.

## 14.2.3 Deuxième étape de la procédure d'inversion

### 14.2.3.1 Interprétation des paramètres ?

Dans un système composé de deux régions uniquement, nous avons étudié comment  $I_{stim}$  dans la région de contrôle et  $J_{gaba_c}$  dans la cible ont un effet sur l'activité de la région de contrôle  $sn_{rc}^*$  et de la cible  $sn_c^*$  au régime stationnaire, au repos et en période de stimulation. Nous n'avons pas étudié le comportement des  $sn_i^*$  dans un système composé de 15 régions (11 régions de contrôle et 4 cibles) avec deux degrés de liberté par région ( $I_{stim_i}$  et  $J_{gaba_i}$ ). Ceci rend pour le moment difficile la compréhension des valeurs des paramètres neuraux inversés pour un participant en sortie de la deuxième étape de la procédure d'inversion.

A plus long terme, nous envisageons d'inverser les données de tous les participants de l'étude REMEMBER. Afin de pouvoir proposer une interprétation clinique des résultats en sortie d'inversion sur tous les participants, la relation non linéaire entre la baisse d'activité des cibles et les paramètres  $J_{gaba}$  dans les cibles est une question importante à explorer. Nous l'avons fait dans un système composé de 2 régions uniquement, avec 2 paramètres libres ( $I_{stim}$  dans la région de contrôle et  $J_{gaba_c}$  dans la cible). Il serait nécessaire de le poursuivre dans un système composé de plus de régions (ici 15 régions) et donc avec plus de paramètres libres.

### 14.2.3.2 Propagation d'incertitudes

La première étape de l'inversion (section 10.3) nous permet d'estimer les moments de la loi gaussienne a posteriori que suit le vecteur  $\vec{\theta}$  contenant tous les paramètres transformés



$s\tilde{n}_{B_i}$  et  $sn_{stim_i}$ . Nous avons un vecteur de moyennes et une matrice de covariance. Il serait intéressant de pouvoir propager ces incertitudes pour avoir un intervalle de confiance sur les paramètres  $I_{stim_i}$  et  $J_{gaba_i}$ .

Nous pourrions travailler à partir du modèle neural réécrit en fonction des paramètres transformés  $s\tilde{n}_{B_i}$  et  $sn_{stim_i}$ , pour nous permettre de continuer à travailler avec des gaussiennes. S'il est possible d'écrire algébriquement la dépendance des paramètres  $I_{stim_i}$  et  $J_{gaba_i}$  aux  $s\tilde{n}_{B_i}$  et  $sn_{stim_i}$ , alors il devrait être possible de calculer la densité de probabilité que suivent ces paramètres quand on connaît la densité de probabilité que suivent  $s\tilde{n}_{B_i}$  et  $sn_{stim_i}$ . Sinon, nous pouvons nous tourner plutôt vers des méthodes numériques de propagation d'incertitudes.

### 14.2.3.3 Sources d'incertitudes ?

Construire une barre d'erreur sur les paramètres neuraux inversés ouvre la question de l'identification des sources d'incertitudes.

Nous venons d'aborder l'idée de construire une barre d'erreur sur les paramètres neuraux inversés à partir des incertitudes sur les  $s\tilde{n}_{B_i}$  et  $sn_{stim_i}$ , et donc les  $sn_{B_i}$  et  $sn_{stim_i}$ .

L'estimation des paramètres neuraux à partir des  $sn_{B_i}$  et  $sn_{stim_i}$  repose entre autres choses sur une matrice de connectivité structurale  $\overline{C}$  mesurée à partir des données de participants. Ces mesures obtenues par Imagerie du Tenseur de Diffusion (DTI) sont bruitées et constituent donc une source d'incertitude pour l'estimation des paramètres neuraux. Comment construire des incertitudes sur le nombre de fibres mesurées entre régions (i.e. sur la matrice de connectivité structurale) ?

# Sixième partie

## Bibliographie



# Bibliographie

- [Abbott and Chance, 2005] Abbott, L. F. and Chance, F. S. (2005). Drivers and modulators from push-pull and balanced synaptic input. In Progress in Brain Research, volume 149, pages 147–155.
- [Aguirre et al., 1998] Aguirre, G. K., Zarahn, E., and D’Esposito, M. (1998). The variability of human, bold hemodynamic responses. NeuroImage, 8 :360–369.
- [Ahmed et al., 2020] Ahmed, M. H., Ghatge, M. S., and Safo, M. K. (2020). Hemoglobin : Structure, function and allostery. Subcellular Biochemistry, 94 :345–382.
- [Anderson et al., 2016] Anderson, M. C., Bunce, J. G., and Barbas, H. (2016). Prefrontal–hippocampal pathways underlying inhibitory control over memory. Neurobiology of Learning and Memory, 134 :145–161.
- [Anderson and Green, 2001] Anderson, M. C. and Green, C. (2001). Suppressing unwanted memories by executive control. Nature, 410 :366–369.
- [Anderson and Hanslmayr, 2014] Anderson, M. C. and Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. Trends in Cognitive Sciences, 18 :279–292.
- [Anderson and Levy, 2009] Anderson, M. C. and Levy, B. J. (2009). Suppressing unwanted memories. Current Directions in Psychological Science, 18 :189–194.
- [Anderson et al., 2004] Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., and Gabrieli, J. D. E. (2004). Neural systems underlying the suppression of unwanted memories. Science, 303 :232–235.
- [Andreelli and Mosbah, 2014] Andreelli, F. and Mosbah, H. (2014). Irm fonctionnelle cérébrale : les principes. Medecine des Maladies Metaboliques, 8 :13–19.
- [Apšvalka et al., 2022] Apšvalka, D., Ferreira, C. S., Schmitz, T. W., Rowe, J. B., and Anderson, M. C. (2022). Dynamic targeting enables domain-general inhibitory control over action and thought by the prefrontal cortex. Nature Communications, 13.
- [Bari and Robbins, 2013] Bari, A. and Robbins, T. W. (2013). Inhibition and impulsivity : Behavioral and neural basis of response control. Progress in Neurobiology, 108 :44–79.
- [Beal, 1998] Beal, M. J. (1998). Variational algorithms for approximate bayesian inference. Physics.
- [Bechara et al., 1994] Bechara, A., Damasio, A. R., Damasio, H., and Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition, 50 :7–15.
- [Benoit and Anderson, 2012] Benoit, R. G. and Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. Neuron, 76 :450–460.

- [Benoit et al., 2015] Benoit, R. G., Hulbert, J. C., Huddleston, E., and Anderson, M. C. (2015). Adaptive top-down suppression of hippocampal activity and the purging of intrusive memories from consciousness. Journal of Cognitive Neuroscience, 27 :96–111.
- [Bergeron et al., 2014] Bergeron, M. J., Castonguay, A., and Koninck, Y. D. (2014). Kcc2 : nouvelle cible thérapeutique pour le traitement de pathologies neurologiques. Medecine/Sciences, 30 :514–517.
- [Bonelli and Cummings, 2007] Bonelli, R. M. and Cummings, J. L. (2007). Frontal-subcortical circuitry and behavior. Dialogues in Clinical Neuroscience, 9 :141–151.
- [Boynton et al., 1996] Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. The Journal of Neuroscience, 16 :4207–4221.
- [Brewin et al., 2010] Brewin, C. R., Gregory, J. D., Lipton, M., and Burgess, N. (2010). Intrusive images in psychological disorders : Characteristics, neural mechanisms, and treatment implications. Psychological Review, 117 :210–232.
- [Broto, 2020] Broto, M. (2020). Proposition d’une prise en charge axée sur l’inhibition chez un adolescent présentant un tda/h. Université Toulouse III, [http://medecine.univ-tlse3.fr/medias/fichier/broto-marie\\_1613030808793-pdf](http://medecine.univ-tlse3.fr/medias/fichier/broto-marie_1613030808793-pdf).
- [Brunel and Wang, 2001] Brunel, N. and Wang, X.-J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. Journal of Computational Neuroscience, 11 :63–85.
- [Buckner, 1998] Buckner, R. L. (1998). Event-related fmri and the hemodynamic response. Hum. Brain Mapping, 6 :373–377.
- [Buckner et al., 1996] Buckner, R. L., Bandettini, P. A., O’Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proceedings of the National Academy of Sciences, 93 :14878–14883. doi : 10.1073/pnas.93.25.14878.
- [Buxton, 2012] Buxton, R. B. (2012). Dynamic models of bold contrast. NeuroImage, 62 :953–961.
- [Buxton et al., 1998] Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation : The balloon model. Magnetic Resonance in Medicine, 39(6) :855–864.
- [Catarino et al., 2015] Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., and Anderson, M. C. (2015). Failing to forget : Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. Psychological Science, 26 :604–616.
- [Cerf et al., 2010] Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R. Q., Koch, C., and Fried, I. (2010). On-line, voluntary control of human temporal lobe neurons. Nature, 467 :1104–1108.
- [Cole et al., 2014] Cole, M. W., Repovš, G., and Anticevic, A. (2014). The frontoparietal control system : A central role in mental health. Neuroscientist, 20 :652–664.

- [Collinson et al., 2002] Collinson, N., Kuenzi, F. M., Jarolimek, W., Maubach, K. A., Cothliff, R., Sur, C., Smith, A., Otu, F. M., Howell, O., Atack, J. R., McKernan, R. M., Seabrook, G. R., Dawson, G. R., Whiting, P. J., and Rosahl, T. W. (2002). Enhanced learning and memory and altered gabaergic synaptic transmission in mice lacking the  $\alpha 5$  subunit of the gabaareceptor. *Journal of Neuroscience*, 22(13) :5572–5580.
- [Corson, 2010] Corson, N. (2010). *Dynamique d’un modèle neuronal, synchronisation et complexité*. PhD thesis, Université du Havre.
- [Crestani et al., 2002] Crestani, F., Keist, R., Fritschy, J.-M., Benke, D., Vogt, K., Prut, L., Blüthmann, H., Möhler, H., and Rudolph, U. (2002). Trace fear conditioning involves hippocampal  $\gamma$ -gaba receptors. *Proceedings of the National Academy of Sciences*, 99 :8980–8985.
- [Czeh et al., 2015] Czeh, B., Varga, Z. K., Henningsen, K., Kovács, G. L., Miseta, A., and Wiborg, O. (2015). Chronic stress reduces the number of gabaergic interneurons in the adult rat hippocampus, dorsal-ventral and region-specific differences. *Hippocampus*, 25 :393–405.
- [Dale and Buckner, 1997] Dale, A. M. and Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fmri. *Hum. Brain Mapping*, 5 :329–340.
- [Daunizeau et al., 2014] Daunizeau, J., Adam, V., and Rigoux, L. (2014). Vba : A probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, 10.
- [Daunizeau et al., 2011] Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic causal modelling : A critical review of the biophysical and statistical foundations. *NeuroImage*, 58 :312–322.
- [Daunizeau et al., 2009] Daunizeau, J., Friston, K. J., and Kiebel, S. J. (2009). Variational bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D : Nonlinear Phenomena*, 238 :2089–2118.
- [de Marco et al., 2006] de Marco, G., (Paris, C., Guillemin, R., Vallée, J.-N., (Amiens, S., Lehmann, P., (Paris, M., and (Paris, J. (2006). L’irm fonctionnelle (irmf) 1re partie : Bases physiques et application aux tumeurs cérébrales. *Neurologies*, Vol. 9.
- [Deco et al., 2018] Deco, G., Cruzat, J., Cabral, J., Knudsen, G. M., Carhart-Harris, R. L., Whybrow, P. C., Logothetis, N. K., and Kringelbach, M. L. (2018). Whole-brain multi-modal neuroimaging model using serotonin receptor maps explains non-linear functional effects of lsd. *Current Biology*, 28 :3065–3074.e6.
- [Deco and Jirsa, 2012] Deco, G. and Jirsa, V. K. (2012). Ongoing cortical activity at rest : Criticality, multistability, and ghost attractors. *Journal of Neuroscience*, 32 :3366–3375.
- [Deco et al., 2014] Deco, G., Ponce-Alvarez, A., Hagmann, P., Romani, G. L., Mantini, D., and Corbetta, M. (2014). How local excitation-inhibition ratio impacts the whole brain dynamics. *Journal of Neuroscience*, 34 :7886–7898.
- [Deco et al., 2013] Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G. L., Hagmann, P., and Corbetta, M. (2013). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience*, 33 :11239–11252.

- [Depue et al., 2007] Depue, B. E., Curran, T., and Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. Science, 317 :215–219.
- [Depue et al., 2016] Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F., and Banich, M. T. (2016). The organization of right prefrontal networks reveals common mechanisms of inhibitory regulation across cognitive, emotional, and motor processes. Cerebral Cortex, 26 :1634–1646.
- [Destexhe et al., 1994] Destexhe, A., Mainen, Z. F., and Sejnowski, T. J. (1994). Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. Journal of Computational Neuroscience, 1 :195–230.
- [Di et al., 2021] Di, X., Zhang, Z., and Biswal, B. B. (2021). Understanding psychophysiological interaction and its relations to beta series correlation. Brain Imaging and Behavior, 15 :958–973.
- [Diamond, 2013] Diamond, A. (2013). Executive functions. Annual Review of Psychology, 64 :135–168.
- [Drachman, 2005] Drachman, D. A. (2005). Do we have brain to spare ? Neurology, 64 :2004–2005.
- [Eichenbaum et al., 2007] Eichenbaum, H., Yonelinas, A. P., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. Annual Review of Neuroscience, 30 :123–152.
- [Eigsti et al., 2006] Eigsti, I.-M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M. B., Davidson, M. C., Aber, J. L., and Casey, B. J. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. Psychological Science, 17 :478–484. PMID : 16771797.
- [Eldridge et al., 2000] Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., and Engel, S. A. (2000). Remembering episodes : a selective role for the hippocampus during retrieval. Nature Neuroscience, 3 :1149–1152.
- [Falconer et al., 2008] Falconer, E., Bryant, R., Felmingham, K. L., Kemp, A. H., Gordon, E., Peduto, A., Olivieri, G., Williams, L. M., Brain, W., and Falconer, E. (2008). The neural networks of inhibitory control in posttraumatic stress disorder. J Psychiatry Neurosci, 33.
- [Farrant and Nusser, 2005] Farrant, M. and Nusser, Z. (2005). Variations on an inhibitory theme : Phasic and tonic activation of gaba a receptors. Nature Reviews Neuroscience, 6 :215–229.
- [Fatima et al., 2020] Fatima, A., Nasreddine, B., and Nadia, B. (2020). Analyse et conception dans la bande ISM de structures coaxiales TEM et quasi-TEM pour des applications médicales et biologiques. PhD thesis, Université Aboubakr Belkaïd–Tlemcen.
- [Fawcett and Hulbert, 2020] Fawcett, J. M. and Hulbert, J. C. (2020). The many faces of forgetting : Toward a constructive view of forgetting in everyday life. Journal of Applied Research in Memory and Cognition, 9 :1–18.
- [Friston et al., 1998] Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-related fmri : Characterizing differential responses. NeuroImage, 7 :30–40.

- [Friston, 2011] Friston, K. J. (2011). Functional and effective connectivity : A review. Brain Connectivity, 1 :13–36.
- [Friston et al., 2003] Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. NeuroImage, 19 :1273–1302.
- [Friston et al., 2000] Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. (2000). Non-linear responses in fmri : The balloon model, volterra kernels, and other hemodynamics. NeuroImage, 12 :466–477.
- [Frässle et al., 2017] Frässle, S., Lomakina, E. I., Razi, A., Friston, K. J., Buhmann, J. M., and Stephan, K. E. (2017). Regression dcm for fmri. NeuroImage, 155 :406–421.
- [Gagnepain et al., 2014] Gagnepain, P., Henson, R. N., and Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. Proceedings of the National Academy of Sciences of the United States of America, 111.
- [Gagnepain et al., 2017] Gagnepain, P., Hulbert, J., and Anderson, M. C. (2017). Parallel regulation of memory and emotion supports the suppression of intrusive memories. Journal of Neuroscience, 37 :6423–6441.
- [Gilson et al., 2019] Gilson, M., Zamora-López, G., Pallarés, V., Adhikari, M. H., Senden, M., Campo, A. T., Mantini, D., Corbetta, M., Deco, G., and Insabato, A. (2019). Model-based whole-brain effective connectivity to study distributed cognition in health and disease. Network Neuroscience, 4 :338–373.
- [Gitelman et al., 2003] Gitelman, D. R., Penny, W. D., Ashburner, J., and Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fmri : The importance of hemodynamic deconvolution. NeuroImage, 19 :200–207.
- [Godefroy, 2009] Godefroy, O. (2009). Fonctions de contrôle frontales et syndromes dysexécutifs : quelles délimitations et quelles avancées ? Revue de neuropsychologie, 1 :12.
- [Griffeth et al., 2011] Griffeth, V. E., Perthen, J. E., and Buxton, R. B. (2011). Prospects for quantitative fmri : Investigating the effects of caffeine on baseline oxygen metabolism and the response to a visual stimulus in humans. NeuroImage, 57 :809–816.
- [Grubb et al., 1974] Grubb, R. L., Raichle, M. E., Eichling, J. O., and Ter-Pogossian, M. M. (1974). The effects of changes in PaCO<sub>2</sub> cerebral blood volume, blood flow, and vascular mean transit time. Stroke, 5(5) :630–639.
- [Gu et al., 2017] Gu, S., Betzel, R. F., Mattar, M. G., Cieslak, M., Delio, P. R., Grafton, S. T., Pasqualetti, F., and Bassett, D. S. (2017). Optimal trajectories of brain state transitions. NeuroImage, 148 :305–317.
- [Heinzle et al., 2016] Heinzle, J., Koopmans, P. J., den Ouden, H. E., Raman, S., and Stephan, K. E. (2016). A hemodynamic model for layered bold signals. NeuroImage, 125 :556–570.
- [Hunt et al., 2023] Hunt, K. J., Knight, L. K., and Depue, B. E. (2023). Related neural networks underlie suppression of emotion, memory, motor processes as identified by data-driven analysis. BMC Neuroscience, 24.



- [Keribin, 2009] Keribin, C. (2009). Les méthodes bayésiennes variationnelles et leur application en neuroimagerie : une étude de l'existant. Technical report, Institut National de Recherche en Informatique et en Automatique, Saclay. Ile-De-France.
- [Kwong et al., 1992] Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I., Weisskoff, R. M., Poncelet, B. P., Kennedy, D., Hoppel, B., Cohen, M., and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proceedings of the National Academy of Sciences of the United States of America, 89 :5675–5679.
- [Lefebvre, 2017] Lefebvre, P. (2017). Elastographie par Résonance Magnétique : Nouvelle méthode d'acquisition fondée sur le contrôle optimal et comparaison de l'ERM avec une technique de rhéologie haute-fréquence. PhD thesis, Université de Lyon - INSA.
- [Leone, 2021] Leone, G. (2021). Predicting the future to control the past : predictive computations underlying the brain dynamics of memory control and their role in understanding post-traumatic stress disorder evolution following the 13th November terrorist attacks. PhD thesis, NIMH - Neuropsychologie et imagerie de la mémoire humaine. 2021NORMC025.
- [Leone et al., 2022] Leone, G., Postel, C., Mary, A., Fraisse, F., Vallée, T., Viader, F., de La Sayette, V., Peschanski, D., Dayan, J., Eustache, F., and Gagnepain, P. (2022). Altered predictive control during memory suppression in ptsd. Nature Communications, 13.
- [Levy and Anderson, 2012] Levy, B. J. and Anderson, M. C. (2012). Purging of memories from conscious awareness tracked in the human brain. Journal of Neuroscience, 32 :16785–16794.
- [Lezak, 1982] Lezak, M. D. (1982). The problem of assessing executive functions. International Journal of Psychology, 17 :281–297.
- [Lezak, 1994] Lezak, M. D. (1994). Domains of behavior from a neuropsychological perspective : the whole story. In Nebr Symp Motiv, volume 41, pages 23–55.
- [Li et al., 1998] Li, D., Waight, D. J., and Wang, Y. (1998). In vivo correlation between blood t2\* and oxygen saturation. Journal of magnetic resonance imaging : JMRI, 8 :1236—1239.
- [Lindquist et al., 2009] Lindquist, M. A., Loh, J. M., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fmri : efficiency, bias and mis-modeling. NeuroImage, 45.
- [Maguire, 2014] Maguire, J. (2014). Stress-induced plasticity of gabaergic inhibition. Frontiers in Cellular Neuroscience, 8.
- [Makkar et al., 2010] Makkar, S. R., Zhang, S. Q., and Cranney, J. (2010). Behavioral and neural analysis of gaba in the acquisition, consolidation, reconsolidation, and extinction of fear memory. Neuropsychopharmacology, 35 :1625–1652.
- [Mandeville et al., 1999] Mandeville, J. B., Marota, J. J. A., Ayata, C., Zaharchuk, G., Moskowitz, M. A., Rosen, B. R., and Weisskoff, R. M. (1999). Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. Journal of Cerebral Blood Flow and Metabolism, 19 :679–689. PMID : 10366199.

- [Mary et al., 2020] Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., Vallée, T., Klein-Peschanski, C., Viader, F., de la Sayette, V., Peschanski, D., Eustache, F., and Gagnepain, P. (2020). Resilience after trauma : The role of memory suppression. Science, 367.
- [McEwen et al., 2016] McEwen, B. S., Nasca, C., and Gray, J. D. (2016). Stress effects on neuronal structure : Hippocampus, amygdala, and prefrontal cortex. Neuropsychopharmacology, 41 :3–23.
- [McLaren et al., 2012] McLaren, D. G., Ries, M. L., Xu, G., and Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gppi) : A comparison to standard approaches. NeuroImage, 61 :1277–1286.
- [Melville et al., 1975] Melville, D., Paul, F., and Roath, S. (1975). Direct magnetic separation of red cells from whole blood. Nature, 255 :706.
- [Menon et al., 1995] Menon, R. S., Ogawa, S., Hu, X., Strupp, J. P., Anderson, P., and Uğurbil, K. (1995). Bold based functional mri at 4 tesla includes a capillary bed contribution : Echo-planar imaging correlates with previous optical imaging using intrinsic signals. Magnetic Resonance in Medicine, 33(3) :453–459.
- [Miyake et al., 2000] Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks : A latent variable analysis. Cognitive Psychology, 41 :49–100.
- [Mueller et al., 2013] Mueller, S., Wang, D., Fox, M., Yeo, B., Sepulcre, J., Sabuncu, M., Shafee, R., Lu, J., and Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. Neuron, 77 :586–595. doi : 10.1016/j.neuron.2012.12.028.
- [Munakata et al., 2011] Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., and O'Reilly, R. C. (2011). A unified framework for inhibitory control. Trends in Cognitive Sciences, 15 :453–459.
- [Naskar et al., 2021] Naskar, A., Vattikonda, A., Deco, G., Roy, D., and Banerjee, A. (2021). Multiscale dynamic mean field (mdmf) model relates resting-state brain dynamics with local cortical excitatory–inhibitory neurotransmitter homeostasis. Network Neuroscience, 5 :757–782.
- [Obata et al., 2004] Obata, T., Liu, T. T., Miller, K. L., Luh, W. M., Wong, E. C., Frank, L. R., and Buxton, R. B. (2004). Discrepancies between bold and flow dynamics in primary and supplementary motor areas : Application of the balloon model to the interpretation of bold transients. NeuroImage, 21 :144–153.
- [Ogawa et al., 1993] Ogawa, S., Menon, R., Tank, D., Kim, S., Merkle, H., Ellermann, J., and Ugurbil, K. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. a comparison of signal characteristics with a biophysical model. Biophysical Journal, 64(3) :803–812.
- [Ogawa et al., 1992] Ogawa, S. T., Tank, D., Menon, R., Ellermann, J., Kim, S.-G., Merkle, H., and KM, U. (1992). Intrinsic signal changes accompanying sensory stimulation : functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences of the United States of America, 89 :5951–5955.

- [O'Reilly et al., 2012] O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., and Johansen-Berg, H. (2012). Tools of the trade : Psychophysiological interactions and functional connectivity. Social Cognitive and Affective Neuroscience, 7 :604–609.
- [Parkes et al., 2018] Parkes, L., Fulcher, B., Yücel, M., and Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional mri. NeuroImage, 171 :415–436.
- [Patrick, 2007] Patrick, N. (2007). Réseaux bayésiens. Eyrolles.
- [Penny et al., 2010] Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., and Leff, A. P. (2010). Comparing families of dynamic causal models. PLOS Computational Biology, 6(3) :1–14.
- [Penny et al., 2004] Penny, W. D., Stephan, K. E., Mechelli, A., and Friston, K. J. (2004). Comparing dynamic causal models. NeuroImage, 22 :1157–1172.
- [Pierigè et al., 2008] Pierigè, F., Serafini, S., Rossi, L., and Magnani, M. (2008). Cell-based drug delivery. Advanced Drug Delivery Reviews, 60 :286–295. Emerging Trends in Cell-Based Therapies.
- [Pilkington and McKELLAR, 1960] Pilkington, G. W. and McKELLAR, P. (1960). Inhibition as a concept in psychology. British Journal of Psychology, 51 :194.
- [Power et al., 2018] Power, J. D., Plitt, M., Gotts, S. J., Kundu, P., Voon, V., Bandettini, P. A., and Martin, A. (2018). Ridding fmri data of motion-related influences : Removal of signals with distinct spatial and physical bases in multiecho data. Proceedings of the National Academy of Sciences, 115 :E2105–E2114. doi : 10.1073/pnas.1720985115.
- [Ridgway, 2010] Ridgway, J. P. (2010). Cardiovascular magnetic resonance physics for clinicians : part i. Journal of Cardiovascular Magnetic Resonance, 12 :71.
- [Rosen et al., 1998] Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998). Event-related functional mri : Past, present, and future. Proceedings of the National Academy of Sciences, 95(3) :773–780.
- [Rosso et al., 2017] Rosso, I. M., Crowley, D. J., Silveri, M. M., Rauch, S. L., and Jensen, J. E. (2017). Hippocampus glutamate and n-acetyl aspartate markers of excitotoxic neuronal compromise in posttraumatic stress disorder. Neuropsychopharmacology, 42 :1698–1705.
- [Schmitz et al., 2017] Schmitz, T. W., Correia, M. M., Ferreira, C. S., Prescott, A. P., and Anderson, M. C. (2017). Hippocampal gaba enables inhibitory control over unwanted thoughts. Nature Communications, 8.
- [Sears and Hewett, 2021] Sears, S. M. and Hewett, S. J. (2021). Influence of glutamate and gaba transport on brain excitatory/inhibitory balance. Experimental Biology and Medicine, 246 :1069–1083.
- [Selten et al., 2018] Selten, M., van Bokhoven, H., and Kasri, N. N. (2018). Inhibitory control of the excitatory/inhibitory balance in psychiatric disorders. F1000Research, 7.
- [Seron, 2009] Seron, X. (2009). L'individualisation des fonctions exécutives : historique et repères. Revue de neuropsychologie, 1 :16.
- [Smith, 2012] Smith, K. (2012). Brain imaging : fmri 2.0. Nature, 484 :24–26.

- [Sohal and Rubenstein, 2019] Sohal, V. S. and Rubenstein, J. L. (2019). Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. Molecular Psychiatry, 24 :1248–1257.
- [Sporns, 2013] Sporns, O. (2013). Structure and function of complex brain networks. Dialogues in Clinical Neuroscience, 15 :247–262. PMID : 24174898.
- [Squire, 1992] Squire, L. R. (1992). Memory and the hippocampus : A synthesis from findings with rats, monkeys, and humans. Psychological Review, 99 :195–231.
- [Stephan et al., 2004] Stephan, K. E., Harrison, L. M., Penny, W. D., and Friston, K. J. (2004). Biophysical models of fmri responses. Current Opinion in Neurobiology, 14 :629–635.
- [Stephan et al., 2010] Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., and Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. NeuroImage, 49 :3099–3109.
- [Stephan et al., 2007] Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., and Friston, K. J. (2007). Comparing hemodynamic models with dcm. NeuroImage, 38 :387–401.
- [Uludağ et al., 2009] Uludağ, K., Müller-Bierl, B., and Uğurbil, K. (2009). An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. NeuroImage, 48 :150–165.
- [Viard et al., 2021] Viard, A., Eustache, F., and Segobin, S. (2021). History of magnetic resonance imaging : A trip down memory lane. Neuroscience, 474.
- [Vogel et al., 2015] Vogel, J., Haddadin, S., Donoghue, J., and van der Smagt, P. (2015). An assistive decision-and-control architecture for force-sensitive hand–arm systems driven by human–machine interfaces. The International Journal of Robotics Research, 34.
- [Wang et al., 2019] Wang, P., Kong, R., Kong, X., Liégeois, R., Orban, C., Deco, G., van den Heuvel, M. P., and Yeo, B. T. (2019). Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. Science Advances, 5(1) :eaat7854.
- [Wittenberg et al., 1970] Wittenberg, J. B., Wittenberg, B. A., Peisach, J., and Blumberg, W. E. (1970). On the state of the iron and the nature of the ligand in oxyhemoglobin. Proceedings of the National Academy of Sciences, 67 :1846–1853.
- [Wong and Wang, 2006] Wong, K. F. and Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. Journal of Neuroscience, 26 :1314–1328.
- [Worsley and Friston, 1995] Worsley, K. J. and Friston, K. J. (1995). Analysis of fmri time-series revisited—again. NeuroImage, 2 :173–181.
- [Yee et al., 2004] Yee, B. K., Hauser, J., Dolgov, V. V., Keist, R., Möhler, H., Rudolph, U., and Feldon, J. (2004). Gaba<sub>A</sub> receptors containing the alpha-5 subunit mediate the trace effect in aversive and appetitive conditioning and extinction of conditioned fear. European Journal of Neuroscience, 20 :1928–1936.
- [Zisis et al., 2021] Zisis, E., Keller, D., Lida, K., Arnaudon, A., Gevaert, M., Delemontex, T., Coste, B., Foni, A., Abdellah, M., Cali, C., Hess, K., Magistretti, P., Schürmann, F., and Markram, H. (2021). Digital reconstruction of the neuro-glia-vascular architecture. Cerebral cortex (New York, N.Y. : 1991), 31.



# Septième partie

## Annexes



# Chapitre 15

## Annexes de la partie I : Introduction

### 15.1 Etude du contrôle inhibiteur des intrusions mnésiques par contrastes

Nous présentons dans un premier temps les principaux résultats d'imagerie, sans entrer dans les détails des méthodes statistiques<sup>1</sup>. Nous précisons quand cela est possible les liens avec les données comportementales et avec les concentrations GABA dans les régions de contrôle ou les régions cibles.

#### 15.1.1 Dans les régions de contrôle

##### 15.1.1.1 Entre Think et No Think

Ce contraste permet d'étudier les variations d'activité des régions quand les participants essaient de garder l'image associée au mot en dehors de leur esprit (No Think), par rapport au cas où ils essaient de s'en rappeler (Think). Une augmentation de l'activité du cortex préfrontal dorsolatéral (DLPFC) et ventrolatéral (VLPFC) est mesurée dans la condition No Think par rapport à la condition Think ([Anderson et al., 2004], [Benoit and Anderson, 2012], [Levy and Anderson, 2012], [Benoit et al., 2015], [Anderson et al., 2016], [Schmitz et al., 2017]). C'est ce qu'illustre la figure 15.1.

Ces résultats sont mis en lien avec les mesures comportementales, i.e. les performances de rappel post-TNT. Plus l'augmentation d'activité du DLPFC par rapport à la condition Think est grande, et plus les items supprimés sont oubliés ([Anderson et al., 2004], [Benoit and Anderson, 2012]), comme le montre la figure 15.2.

[Schmitz et al., 2017] met en évidence que le niveau d'activité dans le DLPFC en No Think n'est pas corrélé à la concentration de GABA dans cette région.

---

1. Nous pourrions nous reporter aux différents articles pour cela. Comme analyses de premier niveau, certains utilisent un test statistique pour comparer les différences de moyennes entre les décours temporels de chaque condition (e.g. [Levy and Anderson, 2012]). D'autres utilisent un GLM avec un régresseur BOLD par condition, obtenu en convoluant le régresseur de tâche (créneaux) par la HRF (e.g. [Benoit et al., 2015]).



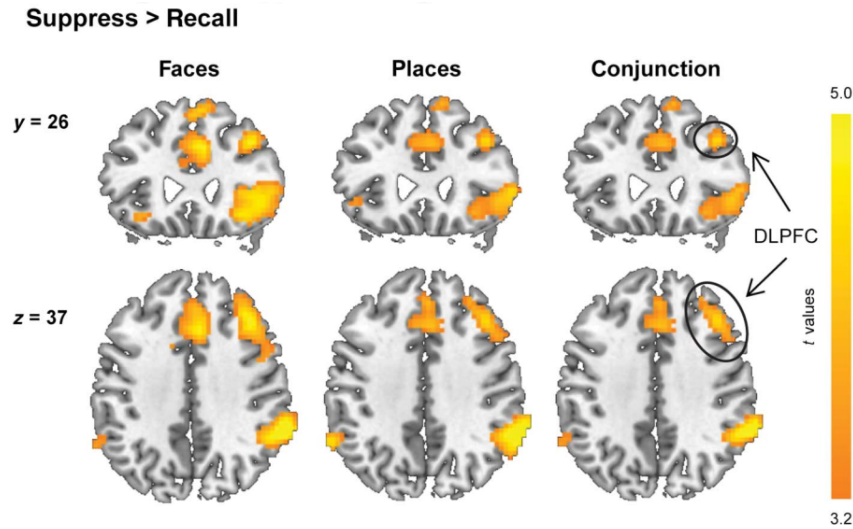


**Figure 15.1** – Contraste No Think > Think : régions statistiquement plus activées lors du contrôle inhibiteur de visages (gauche) ou de lieux (milieu) associés à un mot. Les régions activées à la fois par la suppression de ces deux types de stimuli sont représentées à droite. Elles sont principalement latéralisées à droite et incluent le cortex préfrontal dorsolatéral (DLPFC), ventrolatéral (VLPFC), le cortex cingulaire antérieur (ACC) et l'aire motrice pré-supplémentaire (pre-SMA) ([Benoit et al., 2015]).

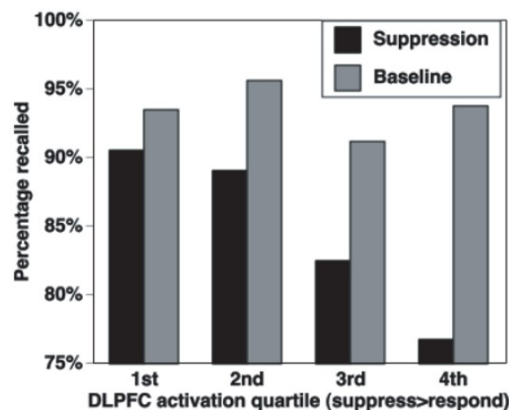


**Figure 15.2** – Plus l'activation du DLPFC est grande pendant la condition No Think par rapport à la condition Think (vers la droite sur l'axe des abscisses) et moins les images No Think sont rappelées après la tâche. Au contraire, le taux de rappel des images de Baseline n'est pas affecté par l'activation du DLPFC ([Anderson et al., 2004]).

#### 15.1.1.2 Entre No Think non intrusif et No Think intrusif

[Levy and Anderson, 2012] et [Benoit et al., 2015] demandent aux participants de rapporter après chaque item No Think s'ils ont eu ("brièvement" ou "souvent") ou non ("jamais")

une intrusion. Ceci permet de définir deux conditions expérimentales, No Think intrusive et No Think non intrusive, et de mesurer la fréquence des intrusions au cours de la tâche (i.e. pour chaque série de mots présentés).

[Benoit et al., 2015] montre que les participants présentent une activité du DLPFC droit plus grande quand ils ont une intrusion que quand ils n'en ont pas. Par ailleurs, plus cette augmentation d'activité est grande, et plus les participants réduisent la fréquence de leurs intrusions au cours de la tâche.

## 15.1.2 Dans les régions cibles

### 15.1.2.1 Entre Think et No Think

Une diminution de l'activité de l'hippocampe droit est mesurée dans la condition No Think par rapport à la condition Think ([Anderson et al., 2004], [Benoit and Anderson, 2012], [Levy and Anderson, 2012], [Schmitz et al., 2017]), i.e. pendant le contrôle inhibiteur d'une intrusion par rapport au moment de la récupération de l'image associée au mot.

[Anderson et al., 2016] explique néanmoins qu'un contraste négatif entre Think et No Think ne montre pas nécessairement que l'activité de l'hippocampe est abaissée spécifiquement pendant le contrôle inhibiteur des intrusions. L'hippocampe pourrait être activée pendant la récupération et non activée pendant la suppression. Cependant, les contrastes peuvent d'abord être mis en lien avec les mesures comportementales. Plus cette diminution est grande et plus les participants ont des performances de rappel post TNT faibles par rapport aux items de baseline ([Anderson et al., 2016], [Schmitz et al., 2017]). Le contraste entre No Think et baseline, en plus d'autres mesures de connectivité (fonctionnelle et effective) et de leur liens avec les données comportementales, que nous détaillons au chapitre 5, sont autant de points qui permettent de montrer que l'hippocampe est une région cible du contrôle inhibiteur des intrusions mnésiques.

[Schmitz et al., 2017] mesurent la concentration de GABA dans l'hippocampe au repos, par spectroscopie. Plus la concentration de GABA dans l'hippocampe est grande et plus l'activité BOLD de cette région est basse en No Think.

### 15.1.2.2 Think par rapport à baseline

En condition Think, l'activité de l'hippocampe augmente de manière statistiquement significative par rapport au signal mesuré hors période de stimulation. Comme attendu, l'activation de l'hippocampe est associée à la récupération de l'image associée au mot ([Levy and Anderson, 2012], [Benoit et al., 2015]).

### 15.1.2.3 No Think par rapport à baseline

En condition No Think au contraire, l'activité de l'hippocampe est réduite de manière statistiquement significative par rapport au signal mesuré hors période de stimulation ([Levy and Anderson, 2012], [Anderson et al., 2016]).

Nous montrons à la figure 15.3 le pourcentage de changement de signal par rapport à la baseline pour les conditions Think et No Think ([Schmitz et al., 2017]).

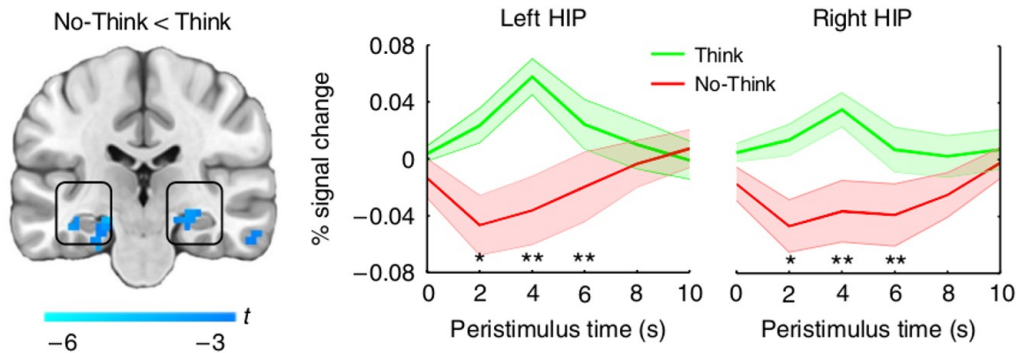


**Figure 15.3** – Variations d'activité du signal BOLD dans les hippocampes.

Gauche : Baisse de l'activité de l'hippocampe en No Think par rapport à la condition Think (analyse statistique de contraste).

Droite : En No Think (resp. Think), le signal BOLD de chaque hippocampe diminue (resp. augmente) au cours du temps par rapport à la baseline, avant de revenir au niveau de base ([Schmitz et al., 2017]).

#### 15.1.2.4 Entre No Think non intrusif, No Think intrusif et baseline

Le contraste entre No Think non intrusif et intrusif est négatif dans l'hippocampe droit. Autrement dit, l'hippocampe subit une baisse d'activité quand un participant a une intrusion en comparaison du cas il n'en a pas ([Levy and Anderson, 2012], [Benoit et al., 2015]).

Nous pouvons visualiser à la figure 15.4 le pourcentage de changement de signal par rapport à la baseline pour les conditions Think, No Think intrusive et non intrusive.

[Levy and Anderson, 2012] montrent que plus l'activité de l'hippocampe en No Think intrusif est abaissée par rapport à la Baseline et plus les participants ont "oublié" les images associées aux mots rouges, en comparaison des mots non présentés pendant la tâche. Au contraire, cette relation est absente quand on regarde les contrastes entre No Think non intrusif et Baseline.

La baisse d'activité de l'hippocampe plus grande en No Think intrusif semble être contre-intuitive : on aurait plutôt pensé qu'une intrusion puisse être associée à une augmentation d'activité de l'hippocampe par rapport à la condition No Think non intrusive, comme lors d'une récupération en Think. [Levy and Anderson, 2012] proposent que cette augmentation d'activité soit très brève et de fait non mesurable avec l'IRMf (cf paragraphe 15.1.4). Les mécanismes de contrôle inhibiteur se mettant en place pendant une durée plus longue, leurs effets sur l'hippocampe seraient par contre mesurables et mesurés. La condition No Think intrusive ferait également appel à des mécanismes réactifs supplémentaires, contrairement à la condition No Think non intrusive, ce que nous voyons au chapitre 5, section 5.5.2.

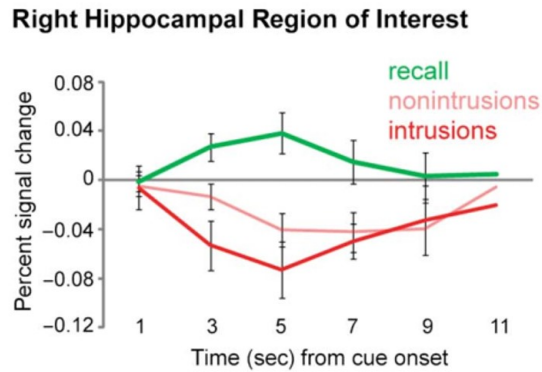


**Figure 15.4** – Variations d’activité du signal BOLD de l’hippocampe droit au cours du temps en fonction des conditions expérimentales Think, No Think intrusive et No Think non intrusive. Par rapport à la baseline, le signal augmente en Think et diminue en No Think (intrusif et non intrusif). Il diminue davantage en présence d’intrusions ([Benoit et al., 2015]).

### 15.1.3 Corrélations d’activité entre régions de contrôle et régions cibles

Le niveau d’activité BOLD du DLPFC en No Think est souvent corrélé négativement au niveau d’activité dans l’hippocampe ([Benoit and Anderson, 2012], [Anderson et al., 2016]). Plus l’activité du DLPFC augmente et plus l’activité de l’hippocampe diminue. Le DLPFC pourrait ainsi exercer un contrôle inhibiteur sur l’hippocampe, mais ces mesures de contrastes ne permettent pas d’étudier la communication entre ces régions.

### 15.1.4 Limites de l’utilisation des contrastes pour l’étude du contrôle inhibiteur

Pour étudier les mécanismes cérébraux du contrôle inhibiteur, la question n’est pas seulement de connaître les variations d’activité des régions de manière indépendante. Nous nous intéressons surtout aux influences réciproques. Pour montrer qu’une baisse d’activité dans l’hippocampe est liée à une augmentation d’activité dans les régions préfrontales, d’autres méthodes plus élaborées permettent d’analyser les signaux BOLD pour étudier la communication entre régions cérébrales. Nous les exposons au chapitre 5.



# Chapitre 16

## Annexes de la partie II : Choix de modélisation

### 16.1 Le modèle neural à une région isolée

On se propose de déterminer le temps nécessaire pour une région pour stabiliser son activité synaptique suite à la réception d'une stimulation.

#### 16.1.1 Evolution des activités synaptiques

Les équations différentielles régissant la dynamique neurale d'une région isolée sont :

$$\begin{cases} \dot{sn}(t) = -\beta^E sn(t) + \alpha^E T_{glu}(1 - sn(t))rn(t) = fn(sn, rn) \\ \dot{sg}(t) = -\beta^I sg(t) + \alpha^I T_{gaba}(1 - sg(t))rg(t) = fg(sg, rg) \end{cases} \quad (16.1)$$

avec les variables intermédiaires :

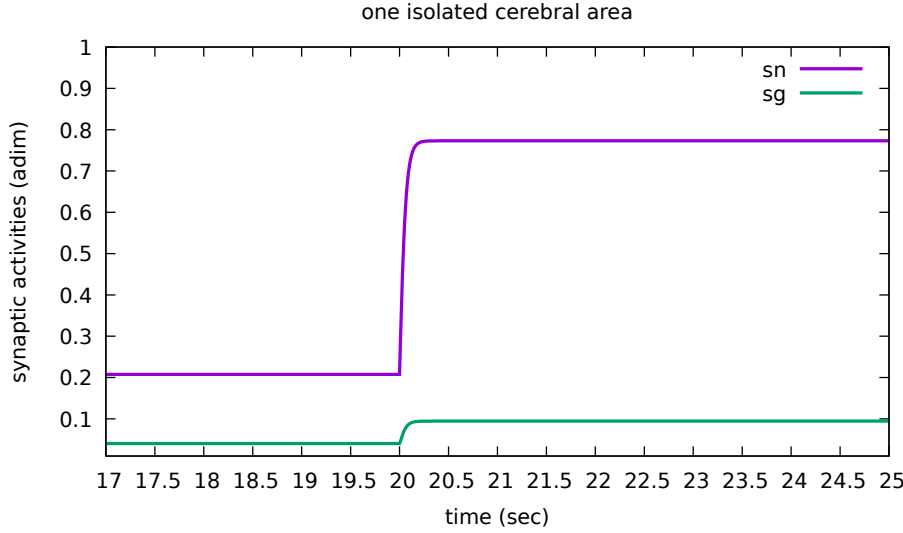
— les fréquences de décharge des potentiel d'action

$$\begin{cases} rn(t) = \frac{a_E xn(t) - b_E}{1 - e^{-d_E(a_E xn(t) - b_E)}} = hn(xn) \\ rg(t) = \frac{a_I xg(t) - b_I}{1 - e^{-d_I(a_I xg(t) - b_I)}} = hg(xg) \end{cases} \quad (16.2)$$

— les courants excitateurs et inhibiteurs

$$\begin{cases} xn(t) = W_+ J_{nmda} sn - J_{gaba} sg + W_E I_0 + I_{stim} U(t) = gn(sn, sg, U) \\ xg(t) = J_{nmda} sn - W_- sg + W_I I_0 = gg(sn, sg) \end{cases} \quad (16.3)$$

On observe les évolutions suivantes de l'activité synaptique en réponse à une stimulation :



On observe (à la précision visuelle près) un temps d'installation d'environ 0.4 sec.

Cette valeur va être confrontée à la valeur du temps caractéristique déterminé ci-dessous à l'aide de l'analyse linéaire.

### 16.1.2 Analyse linéaire au voisinage du Point Fixe.

Dans le registre des paramètres donnés, le système d'équations [16.1] possède un point fixe (PF) unique donné par la solution des équations algébriques :

$$\begin{cases} 0 = -\beta^E sn^* + \alpha^E T_{glu}(1 - sn^*)rn^* \\ 0 = -\beta^I sg^* + \alpha^I T_{gaba}(1 - sg^*)rg^* \end{cases}$$

La linéarisation du système d'équations différentielles [16.1] autour du point fixe  $\vec{s}^* = (sn^*, sg^*)$  et  $\vec{r}^* = (rn^*, rg^*)$  aboutit à un nouveau système d'équations différentielles, dit **système linéarisé tangent** :

$$\begin{pmatrix} \delta \dot{sn} \\ \delta \dot{sg} \end{pmatrix} = \underbrace{\overrightarrow{f(sn^*, rn^*)}}_{=0 \text{ par définition du PF}} + \left. \frac{\partial \overline{f}}{\partial s} \right|^* \begin{pmatrix} \delta sn \\ \delta sg \end{pmatrix}_{(t)} + \left. \frac{\partial \overline{f}}{\partial r} \right|^* \begin{pmatrix} \delta rn \\ \delta rg \end{pmatrix}_{(t)} \quad (16.4)$$

qui donne par dérivation partielle de  $f_i$  par rapport à  $sn, sg$  :

$$\begin{cases} \delta \dot{sn} = -\{\beta_E + \alpha^E T_{glu} rn^*\} \delta sn_{(t)} + \alpha^E T_{glu} (1 - sn^*) \delta rn_{(t)} \\ \delta \dot{sg} = -\{\beta^I + \alpha^I T_{gaba} rg^*\} \delta sg_{(t)} + \alpha^I T_{gaba} (1 - sg_{(t)}^*) \delta rg_{(t)} \end{cases} \quad (16.5)$$

Ce système linéarisé fournit d'une part une information sur la stabilité locale autour du PF, d'autre part une indication sur les temps caractéristiques du système de départ au voisinage du PF.

Du fait que les fréquences excitatrices ne dépendent que des courants excitateurs (et réciproquement pour les fréquences inhibitrices), on peut remplacer dans la première linéarisation  $\delta rn$  par les dérivations partielles  $\frac{\partial rn}{\partial xn} \frac{\partial xn}{\partial sn} \delta sn$  (et de même pour la seconde équation).

On peut alors résumer 16.5 par :

$$\begin{pmatrix} \delta \dot{sn} \\ \delta \dot{sg} \end{pmatrix}_{(t)} = \underbrace{\begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} \end{pmatrix}}_{\text{matrice } \mathcal{A}^*} \begin{pmatrix} \delta sn \\ \delta sg \end{pmatrix}_{(t)} \quad (16.6)$$

qui fait apparaître une **matrice** dite **d'avance** des états  $\mathcal{A}^*$  qui est la matrice jacobienne  $\frac{\partial f}{\partial s}$ , une fois ré-exprimées les  $\delta r$  en fonction des  $\delta s$ . Cette matrice est constante et dépend des valeurs des variables déterminées au PF.

Un théorème de Liapunov montre que les dynamiques du système dynamique de départ 16.1 et du système linéarisé tangent 16.6 sont identiques au voisinage du PF. On peut donc déduire les temps caractéristiques du système dynamique de départ à partir de l'étude des temps caractéristiques du système linéarisé tangent au voisinage du PF.

### 16.1.3 Détermination numérique des temps caractéristiques

La simulation numérique réalisée dans le logiciel ZOOM/Miniker permet de déterminer à chaque instant  $t$  de l'évolution du système la matrice d'avance  $\mathcal{A}_{(t)}$ , et également les valeurs propres de cette matrice d'avance du système linéarisé tangent, et donc notamment en fin d'évolution, quand le régime stationnaire est atteint (au PF), après application de la stimulation.

Lorsqu'on se trouve au voisinage du point fixe (PF), on observe que les deux valeurs propres sont deux réels négatifs, différent d'un ordre de grandeur ( rapport des deux  $\simeq 25$ ).

$$\begin{cases} \lambda_1 = -25.96 \\ \lambda_2 = -644.36 \end{cases} \quad (16.7)$$

et les deux temps caractéristiques sont donnés par l'inverse de leurs valeurs absolues :  $\tau_1 \simeq 0.039$  s et  $\tau_2 \simeq 0.0015$  s.

D'après la convention énoncée section 7.1.1.2.1, on aurait donc un temps d'installation  $t_{ins} \simeq 0.45$  s, ce qui concorde avec l'observation visuelle des courbes d'évolution.

### 16.1.4 Estimation des temps caractéristiques par voie analytique

Après dérivation, on obtient les éléments suivants pour la matrice d'avance :

$$\begin{cases} \mathcal{A}_{11} = -\beta^E - \alpha^E T_{glu} rn^* + \alpha^E T_{glu} (1 - sn^*) \frac{\partial hn}{\partial xn} \frac{\partial gn}{\partial sn} \\ \mathcal{A}_{12} = +\alpha^E T_{glu} (1 - sn^*) \frac{\partial hn}{\partial xn} \frac{\partial gn}{\partial sg} \\ \mathcal{A}_{21} = +\alpha^I T_{gaba} (1 - sg^*) \frac{\partial hg}{\partial xg} \frac{\partial gg}{\partial sn} \\ \mathcal{A}_{22} = -\beta^I - \alpha^I T_{gaba} rg^* + \alpha^I T_{gaba} (1 - sg^*) \frac{\partial hg}{\partial xg} \frac{\partial gg}{\partial sg} \end{cases} \quad (16.8)$$



Nous faisons deux approximations :

**1ère approximation :** en ce qui concerne les dérivées partielles  $\frac{\partial hn}{\partial xn}$  on observe qu'au point fixe, elles sont approchées à l'aide de la pente de l'asymptote oblique de la fonction, c'est-à-dire  $a_E$ . Cette hypothèse simplificatrice a été validée de façon numérique.

La dérivée partielle  $\frac{\partial gn}{\partial sn}$  est donnée par  $W_+ J_{nmda}$ .

Soit finalement après dérivations partielles :

$$\boxed{\mathcal{A}_{11} \simeq -\beta^E - \alpha^E T_{glu} r n^* + \alpha^E T_{glu} (1 - s n^*) a_E W_+ J_{nmda}} \quad (16.9)$$

**2ème approximation :** Les valeurs propres de la matrice  $\mathcal{A}^*$  se calculent aisément à l'aide des formules :

$$\lambda_{1,2} = \frac{(\mathcal{A}_{11} + \mathcal{A}_{22}) \pm \sqrt{(\mathcal{A}_{11} - \mathcal{A}_{22})^2 + 4\mathcal{A}_{12}\mathcal{A}_{21}}}{2} \quad (16.10)$$

On remarque que si  $|4\mathcal{A}_{12}\mathcal{A}_{21}| \ll (\mathcal{A}_{11} - \mathcal{A}_{22})^2$ , alors on est dans le cas d'une hypothèse de dominance diagonale de la matrice et on a  $\lambda_1 \simeq \mathcal{A}_{11}$  et  $\lambda_2 \simeq \mathcal{A}_{22}$ .

Lorsqu'on examine la matrice  $\mathcal{A}^*$  au PF après stimulation, on constate que cette approximation de dominance diagonale est vérifiée, et qu'on peut considérer par exemple  $\mathcal{A}_{11}$  comme du même ordre de grandeur que  $\lambda_1$  ; il en est de même pour l'estimation du temps caractéristique  $\tau_1$  qui en est l'inverse, changé de signe.

### 16.1.5 Paramètres du modèle influant sur ces temps

Dans le cadre de ces approximations confirmées par les simulations numériques, on peut donc estimer analytiquement quels paramètres "pilotent" le temps caractéristique de l'activité excitatrice , et donc du système à une région isolée en comparant les ordres de grandeur des termes :

On a trouvé, après application d'un stimulus  $I_{stim} = 0.1$  nA, et avec le jeu de paramètres de départ, qu'au PF :  $s n^* \simeq 0.773$  et  $r n^* \simeq 39.01$  Hz ; et  $sg^* \simeq 0.0945$  et  $rg^* \simeq 11.81$  Hz.

Le coefficient  $\mathcal{A}_{11}$  peut s'analyser quantitativement comme :

$$\mathcal{A}_{11}^* \simeq - \left\{ \underbrace{\beta^E}_{6.6} + \underbrace{\alpha^E T_{glu} r n^*}_{22.5} - \underbrace{\alpha^E T_{glu} (1 - s n^*) a_E W_+ J_{nmda}}_{8.5} \right\} \simeq - 20.6 \text{ s}^{-1} \quad (16.11)$$

On constate donc que  $\mathcal{A}_{11}$  et en suivant, la valeur propre  $\lambda_1$  sont pilotés principalement par le groupement  $\alpha^E T_{glu} r n^*$

Dans ce groupement, le paramètre  $\alpha^E$  (considéré dans la communauté neurobiologique comme invariant d'une région cérébrale à l'autre) vaut ici  $\alpha_E = 72.$ , et  $T_{glu} = 0.008$ , et leur produit vaut environ 0.6, proche de 1. C'est donc la fréquence des potentiels d'action  $r n^*$ ,

issue de la résolution au PF, qui impose l'ordre de grandeur de la dynamique du système neural.

Indirectement, on constate que c'est la valeur du paramètre  $J_{gaba}$  affecté à la région qui est important : c'est lui qui est fixé pour répondre à la contrainte sur la fréquence qui doit être de  $rn_0 = 3$  Hz à l'état de repos, et c'est donc lui aussi qui pilote la valeur de  $rn^*$  quand la région est stimulée.

On voit que, sous ce modèle, on ne peut s'attendre à de grands "effets retard" entre régions cérébrales, à partir de paramétrisations différentes d'une région à l'autre, si la variabilité des  $J_{gaba}$  (et des  $T_{glu}$ ) reste faible, c'est-à-dire limitée à un ordre de grandeur...



# Chapitre 17

## Annexes de la partie IV : Inversion

### 17.1 Première étape

#### 17.1.1 Loi Gamma de paramètres $\alpha$ et $\beta$

Soit  $X$  une variable aléatoire qui suit une loi  $\Gamma$ , de paramètres  $\alpha$  (shape parameter) et  $\beta$  (rate parameter).

$$X \sim \Gamma(\alpha, \beta) \quad (17.1)$$

La densité de probabilité s'écrit :

$$\forall x > 0, \alpha, \beta > 0, \Gamma(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{(\alpha-1)!} \quad (17.2)$$

L'espérance et la variance dépendent des paramètres  $\alpha$  et  $\beta$  :

$$\langle X \rangle = \frac{\alpha}{\beta} \quad (17.3)$$

$$Var(X) = \langle X^2 \rangle - \langle X \rangle^2 = \frac{\alpha}{\beta^2} \quad (17.4)$$

#### 17.1.2 Densité de probabilité d'un paramètre hémodynamique non transformé $sn$

Cette partie a été travaillée avec le cours de Benjamin Jourdin (<https://cermics.enpc.fr/~jourdain/probastat/poly.pdf>), paragraphe 3.1.2 (p 38) et p. 46.

Soient  $\tilde{Sn}$  et  $Sn$  deux variables aléatoires à valeurs respectivement dans  $\mathbb{R}$  et dans  $[0; 1]$ .  $Sn$  s'écrit en fonction de  $\tilde{Sn}$  :

$$Sn = \frac{1}{1 + \exp(-(\tilde{Sn} + \ln(\frac{s_0}{1-s_0})))} \quad (17.5)$$

$s_0$  est un paramètre de la distribution. Si  $\tilde{Sn} = 0$  alors  $Sn = s_0$ .

Nous écrirons  $sn$  et  $\tilde{sn}$  une réalisation de chacune de ces deux variables aléatoires.

Supposons que  $\tilde{Sn}$  suive une loi normale  $\mathcal{N}(\mu_\theta, \sigma_\theta^2)$ . Quelle est alors la distribution de probabilité que suit la nouvelle variable aléatoire  $Sn$  ?

Pour cela, on utilise la méthode de la fonction muette i.e. on se donne une fonction  $f : [0; 1] \longrightarrow \mathbb{R}$  bornée. On cherche à calculer l'espérance  $E(f(Sn))$  sous la forme  $\int_{[0;1]} f(sn)p(sn)dsn$ . On travaille à partir de ce que l'on connaît c'est-à-dire la loi de  $\tilde{Sn}$  :

$$E(f(Sn)) = E\left(f\left(\frac{1}{1 + \exp(-(\tilde{Sn} + \ln(\frac{s_{n0}}{1-s_{n0}})))}\right)\right) \quad (17.6)$$

Soit  $\Phi : \mathbb{R} \longrightarrow [0; 1]$ . Cette fonction donne l'expression de chaque réalisation  $sn$  de la nouvelle v.a.  $Sn$  en fonction d'une réalisation  $\tilde{sn}$  de la v.a.  $\tilde{Sn}$  dont on connaît déjà la densité de probabilité.

$$\Phi : \mathbb{R} \longrightarrow [0; 1] \quad (17.7)$$

$$\tilde{sn} \longrightarrow \frac{1}{1 + \exp(-(\tilde{sn} + \ln(\frac{s_0}{1-s_0})))} \equiv sn \quad (17.8)$$

$$(17.9)$$

La fonction  $g(\tilde{sn}) = f \circ \Phi(\tilde{sn}) = f\left(\frac{1}{1 + \exp(-(\tilde{sn} + \ln(\frac{s_{n0}}{1-s_{n0}})))}\right)$  est une fonction bornée sur  $\mathbb{R}$ .

$$E(g(\tilde{Sn})) = \int_{\mathbb{R}} g(\tilde{sn})p(\tilde{sn})d\tilde{sn} \quad (17.10)$$

$$= \int_{\mathbb{R}} g(\tilde{sn}) \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{(\tilde{sn} - \mu_\theta)^2}{2\sigma_\theta^2}\right) d\tilde{sn} \quad (17.11)$$

On réécrit l'égalité en explicitant  $g$  :

$$E\left(f\left(\frac{1}{1 + \exp(-(\tilde{Sn} + \ln(\frac{s_0}{1-s_0})))}\right)\right) = \int_{\mathbb{R}} f\left(\frac{1}{1 + \exp(-(\tilde{sn} + \ln(\frac{s_0}{1-s_0})))}\right) \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{(\tilde{sn} - \mu_\theta)^2}{2\sigma_\theta^2}\right) d\tilde{sn} \quad (17.12)$$

$$\equiv E(f(Sn)) \quad (17.13)$$

L'objectif est d'écrire cette intégrale sous la forme  $\int_{[0;1]} f(sn)p(sn)dsn$  en passant par le changement de variable  $\tilde{sn} = \Phi^{-1}(sn)$  où  $\Phi^{-1}$  est la bijection inverse de  $\Phi$  définie par :

$$\Phi^{-1} : [0; 1] \longrightarrow \mathbb{R} \quad (17.14)$$

$$sn \longrightarrow \tilde{sn} = \ln\left(\frac{sn}{1-sn}\right) - \ln\left(\frac{s_0}{1-s_0}\right) = \ln\left(\frac{sn(1-s_0)}{(1-sn)s_0}\right) \quad (17.15)$$

*Changement de variable :*

Nous écrivons d'abord  $d\tilde{sn}$  en fonction de  $dsn$  :

$$d\tilde{n} = d\left[\ln\left(\frac{sn(1-sn_0)}{(1-sn)sn_0}\right)\right] = \frac{d\Phi^{-1}}{dsn}dsn \quad (17.16)$$

Calculons  $\frac{d\Phi^{-1}}{dsn}$  :

$$\frac{d\Phi^{-1}}{dsn} = \frac{\partial}{\partial sn} \left[ \frac{sn(1-sn_0)}{(1-sn)sn_0} \right] \frac{1}{\frac{sn(1-sn_0)}{(1-sn)sn_0}} \quad (17.17)$$

$$= \frac{sn_0(1-sn_0)}{(1-sn)^2 sn_0} \frac{(1-sn)sn_0}{sn(1-sn_0)} \quad (17.18)$$

$$= \frac{1}{(1-sn)sn} \quad (17.19)$$

On en déduit

$$d\tilde{n} = \frac{1}{(1-sn)sn} dsn \quad (17.20)$$

L'intégrale (17.12) devient :

$$E(f(Sn)) = \int_{[0;1]} f(sn) \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{(\tilde{sn} - \mu_\theta)^2}{2\sigma_\theta^2}\right) \frac{1}{(1-sn)sn} dsn \quad (17.21)$$

$$= \int_{[0;1]} f(sn) \underbrace{\frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{[\ln\left(\frac{sn(1-s_0)}{(1-sn)s_0}\right) - \mu_\theta]^2}{2\sigma_\theta^2}\right)}_{p(sn)} \frac{1}{(1-sn)sn} dsn \quad (17.22)$$

La loi que suit la variable aléatoire  $Sn$  s'écrit :

$$\forall sn \in [0, 1], p(sn) = \frac{1}{(1-sn)sn\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{[\ln\left(\frac{sn(1-s_0)}{(1-sn)s_0}\right) - \mu_\theta]^2}{2\sigma_\theta^2}\right) \quad (17.23)$$

$s_0$  est la valeur de  $sn$  lorsque  $\tilde{sn} = 0$ .

Nous fixons  $s_0 = 0.5$ , milieu de l'intervalle des valeurs  $sn$  possibles (entre 0 et 1). Nous représentons à la figure 17.1, les lois que suivent  $\tilde{Sn}$  et  $Sn$  pour  $\mu_\theta = 0$  et  $\sigma_\theta^2 = 3$ .

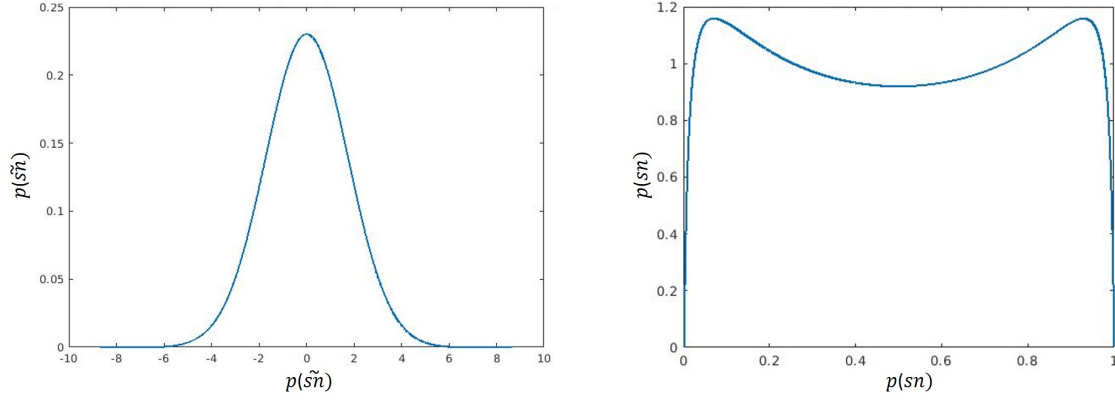


**Figure 17.1** – Gauche : Densité de probabilité que suit  $\tilde{S}n : \mathcal{N}(0, 3)$ , Droite : Densité de probabilité correspondante que suit  $S_n$ .

### 17.1.3 Densité de probabilité d’une variable hémodynamique non transformée

Cette partie a été travaillée avec le cours de Benjamin Jourdin (<https://cermics.enpc.fr/~jourdain/probastat/poly.pdf>), paragraphe 3.1.2 (p 38) et p. 46.

Soient  $\tilde{X}$  et  $X$  deux variables aléatoires à valeurs respectivement dans  $\mathbb{R}$  et dans  $\mathbb{R}^+$ .  $X$  s’écrit en fonction de  $\tilde{X}$  :

$$X = e^{\tilde{X}} \quad (17.24)$$

Nous écrirons  $x$  et  $\tilde{x}$  une réalisation de chacune de ces deux variables aléatoires.

Supposons que  $\tilde{X}$  suive une loi normale  $\mathcal{N}(\mu_0, \sigma_0^2)$ . Quelle est alors la distribution de probabilité que suit la nouvelle variable aléatoire  $X$  ?

Pour cela, on utilise la méthode de la fonction muette i.e. on se donne une fonction  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  bornée. On cherche à calculer l’espérance  $E(f(X))$  sous la forme  $\int_{\mathbb{R}^+} f(x)p(x)dx$ . On travaille à partir de ce que l’on connaît c’est-à-dire la loi de  $\tilde{X}$  :

$$E(f(X)) = E(f(e^{\tilde{X}})) \quad (17.25)$$

Soit  $\mathcal{C}$  une fonction qui donne l’expression de chaque réalisation  $x$  de la nouvelle v.a.  $X$  en fonction d’une réalisation  $\tilde{x}$  de la v.a.  $\tilde{X}$  dont on connaît déjà la densité de probabilité.

$$\mathcal{C} : \mathbb{R} \rightarrow \mathbb{R}^+ \quad (17.26)$$

$$\tilde{x} \rightarrow e^{\tilde{x}} \equiv x \quad (17.27)$$

$$(17.28)$$

La fonction  $g(\tilde{x}) = f \circ \mathcal{C}(\tilde{x}) = f(e^{\tilde{x}})$  est une fonction bornée sur  $\mathbb{R}$ .

$$E(g(\tilde{X})) = \int_{\mathbb{R}} g(\tilde{x}) p(\tilde{x}) d\tilde{x} \quad (17.29)$$

$$= \int_{\mathbb{R}} g(\tilde{x}) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\tilde{x} - \mu_0)^2}{2\sigma_0^2}\right) d\tilde{x} \quad (17.30)$$

On réécrit l'égalité en explicitant  $g$  :

$$E(f(e^{\tilde{X}})) = \int_{\mathbb{R}} f(e^{\tilde{x}}) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\tilde{x} - \mu_0)^2}{2\sigma_0^2}\right) d\tilde{x} \quad \equiv E(f(X)) \quad (17.31)$$

L'objectif est d'écrire cette intégrale sous la forme  $\int_{\mathbb{R}^+} f(x) p(x) dx$  en passant par le changement de variable  $\tilde{x} = \mathcal{C}^{-1}(x)$  où  $\mathcal{C}^{-1}$  est la bijection inverse de  $\mathcal{C}$  définie par :

$$\mathcal{C}^{-1} : \mathbb{R}^+ \longrightarrow \mathbb{R} \quad (17.32)$$

$$x \longrightarrow \tilde{x} = \ln(x) \quad (17.33)$$

*Changement de variable :*

Nous écrivons d'abord  $d\tilde{x}$  en fonction de  $dx$  :

$$d\tilde{x} = d[\ln(x)] = \frac{1}{x} dx \quad (17.34)$$

L'intégrale (17.31) devient :

$$E(f(X)) = \int_{\mathbb{R}^+} f(x) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\tilde{x} - \mu_0)^2}{2\sigma_0^2}\right) \frac{1}{x} dx \quad (17.35)$$

$$= \int_{\mathbb{R}^+} f(x) \underbrace{\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\ln(x) - \mu_0)^2}{2\sigma_0^2}\right) \frac{1}{x}}_{p(x)} dx \quad (17.36)$$

$$(17.37)$$

La loi que suit la variable aléatoire  $X$  s'écrit :

$$\forall x \in \mathbb{R}^+, p(x) = \frac{1}{x\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\ln(x) - \mu_0)^2}{2\sigma_0^2}\right) \quad (17.38)$$

Il s'agit de l'expression analytique de la loi dite log-normale.

Nous représentons à la figure 17.2, les lois que suivent  $\tilde{X}$  et  $X$  pour  $\mu_0 = 0$  et  $\sigma_0^2 = 10^{-4}$ .



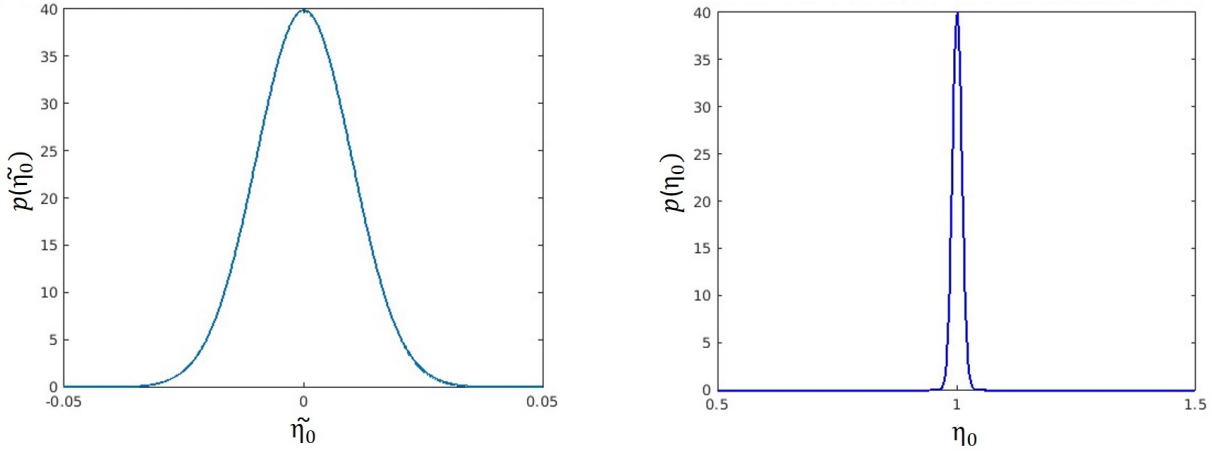


**Figure 17.2** – Gauche : Densité de probabilité a priori normale pour  $\tilde{X}$  :  $\mathcal{N}(0, 10^{-4})$ .  
Droite : Densité de probabilité lognormale correspondante pour  $X$ .

## 17.1.4 Calcul des jacobiennes du modèle différentiel hémodynamique transformé

### 17.1.4.1 Modèle hémodynamique transformé

Nous avons vu à la section 10.3.2, la nécessité de réécrire le modèle hémodynamique avec une transformation des paramètres hémodynamiques à inverser  $sn_{B_i}$  et  $sn_{stim_i}$ , ainsi que des variables d'état de flux  $f_i$ , volume  $v_i$  et quantité de déoxyhémoglobine  $q_i$ . Le modèle hémodynamique transformé s'écrit, pour chaque région  $i$  :

$$\begin{cases} sn_{B_i} = \Phi(s\tilde{n}_{B_i}) = \frac{1}{1 + \exp(-(s\tilde{n}_{B_i} + \ln(\frac{s_0}{1-s_0})))} \\ sn_{stim_i} = \Phi(sn\tilde{stim}_i) = \frac{1}{1 + \exp(-(sn\tilde{stim}_i + \ln(\frac{s_0}{1-s_0})))} \end{cases} \quad (17.39)$$

$$\begin{cases} \dot{z}_i(t) = \Phi(s\tilde{n}_{B_i})\delta_{u(t),0} + \Phi(sn\tilde{stim}_i)\delta_{u(t),1} - \kappa z_i(t) - \gamma(e^{\tilde{f}_i(t)} - 1) = \tilde{h}z_i(s\tilde{n}_{B_i}, sn\tilde{stim}_i, z_i, \tilde{f}_i) \\ \dot{\tilde{f}}_i(t) = z_i(t)e^{\tilde{f}_i(t)} = \tilde{h}f_i(z_i, \tilde{f}_i) \\ \dot{\tilde{v}}_i(t) = \frac{1}{\tau} \left( e^{\tilde{f}_i(t) - \tilde{v}_i(t)} - e^{(\frac{1}{\alpha} - 1)\tilde{v}_i(t)} \right) = \tilde{h}v_i(\tilde{f}_i, \tilde{v}_i) \\ \dot{\tilde{q}}_i(t) = \frac{1}{\tau\rho} e^{\tilde{f}_i(t) - \tilde{q}_i(t)} \left( 1 - (1 - \rho)e^{-\tilde{f}_i(t)} \right) - \frac{1}{\tau} e^{(\frac{1}{\alpha} - 1)\tilde{v}_i(t)} = \tilde{h}q_i(\tilde{f}_i, \tilde{q}_i, \tilde{v}_i) \end{cases} \quad (17.40)$$

$$\text{Avec } \begin{cases} u(t) = 0 \text{ en période de non stimulation} \\ u(t) = 1 \text{ en période de stimulation} \end{cases} \quad (17.41)$$

Le symbole  $\delta_{u(t),0}$  (resp.  $\delta_{u(t),1}$ ) est le delta de Kronecker qui est égal à 1 si  $u(t) = 0$  (resp.  $u(t) = 1$ ). Il permet d'appliquer deux forçages sur le signal de vasodilatation de chaque région en fonction des périodes de stimulation ou de non stimulation :  $sn_{B_i}^* = \Phi(sn_{B_i})$  en période de non stimulation et  $sn_{stim_i}^* = \Phi(sn_{stim_i})$  en période de stimulation.

Nous posons  $\vec{x} = (z, f, \tilde{v}, \tilde{q})$  le vecteur des variables d'état et  $\vec{\theta} = (sn_{B_i}, sn_{stim_i})$  le vecteur des paramètres hémodynamiques inconnus que nous souhaitons estimer.

$u(t)$  est l'input extérieur connu à l'instant  $t$ , qui permet d'appliquer des forçages  $sn_{B_i}$  ou  $sn_{stim_i}$  sur les signaux de vasodilatation  $z$  en fonction des périodes de repos ou de stimulation.

La fonction vectorielle  $\vec{H}$  nous permettra d'écrire le modèle hémodynamique à tout instant  $t$  :

$$\dot{\vec{x}}(t) = \vec{H}(\vec{x}(t), \vec{\theta}, u(t)) \quad (17.42)$$

#### 17.1.4.2 Calculs des dérivées partielles par rapport aux variables d'état

Pour chaque région  $i$ , nous donnons les éléments de matrice de la jacobienne  $\frac{\partial \vec{H}}{\partial x}$ , c'est-à-dire les dérivées partielles des 4 fonctions  $\tilde{h}z_i, \tilde{h}f_i, \tilde{h}v_i, \tilde{h}q_i$  par rapport à chacune des variables  $z_i, \tilde{f}_i, \tilde{q}_i, \tilde{v}_i$ , ainsi que  $z_j, \tilde{f}_j, \tilde{q}_j, \tilde{v}_j$  pour toute autre région  $j$ .

Puisque le modèle hemodynamique n'est pas couplé entre régions, nous pouvons d'abord écrire que :

$$\frac{\partial \tilde{h}z_i}{\partial z_{j \neq i}} = \frac{\partial \tilde{h}z_i}{\partial \tilde{f}_{j \neq i}} = \frac{\partial \tilde{h}z_i}{\partial \tilde{v}_{j \neq i}} = \frac{\partial \tilde{h}z_i}{\partial \tilde{q}_{j \neq i}} = 0 \quad (17.43)$$

et de même pour les fonctions  $\tilde{h}f_i, \tilde{h}v_i$  et  $\tilde{h}q_i$ .

Il nous reste à calculer les dérivées partielles des fonctions d'une région  $i$  ( $\tilde{h}z_i, \tilde{h}f_i, \tilde{h}v_i, \tilde{h}q_i$ ) par rapport aux variables d'état de cette même région ( $z_i, \tilde{f}_i, \tilde{v}_i, \tilde{q}_i$ ).

##### 17.1.4.2.1 Dérivées partielles par rapport à $z_i$

$$\frac{\partial \tilde{h}z_i}{\partial z_i} = -\kappa \quad (17.44)$$

$$\frac{\partial \tilde{h}f_i}{\partial z_i} = e^{-\tilde{f}_i} = \frac{1}{f_i} \quad (17.45)$$

$$\frac{\partial \tilde{h}v_i}{\partial z_i} = 0 \quad (17.46)$$

$$\frac{\partial \tilde{h}q_i}{\partial z_i} = 0 \quad (17.47)$$

#### 17.1.4.2.2 Dérivées partielles par rapport à $\tilde{f}_i$

$$\frac{\partial \tilde{h}z_i}{\partial \tilde{f}_i} = -\gamma e^{\tilde{f}_i} = -\gamma f_i \quad (17.48)$$

$$\frac{\partial \tilde{h}f_i}{\partial \tilde{f}_i} = -z_i e^{\tilde{f}_i} = -\frac{z_i}{f_i} \quad (17.49)$$

$$\frac{\partial \tilde{h}v_i}{\partial \tilde{f}_i} = \frac{1}{\tau} e^{\tilde{f}_i - \tilde{v}_i} = \frac{1}{\tau} \frac{f_i}{v_i} \quad (17.50)$$

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{f}_i} = \frac{1}{\tau \rho} \frac{f_i}{q_i} \left( \left( 1 - (1 - \rho)^{\frac{1}{f_i}} \right) + \frac{1}{f_i} \ln(1 - \rho)(1 - \rho)^{\frac{1}{f_i}} \right) * * \quad (17.51)$$

\*\*calcul détaillé :

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{f}_i} = \frac{\partial}{\partial \tilde{f}_i} \left( \frac{1}{\tau \rho} e^{\tilde{f}_i - \tilde{q}_i} \left( 1 - (1 - \rho)^{e^{-\tilde{f}_i}} \right) \right) \quad (17.52)$$

$$= \frac{1}{\tau \rho} \left( e^{\tilde{f}_i - \tilde{q}_i} \left( 1 - (1 - \rho)^{e^{-\tilde{f}_i}} \right) + e^{\tilde{f}_i - \tilde{q}_i} \frac{\partial}{\partial \tilde{f}_i} \left( 1 - (1 - \rho)^{e^{-\tilde{f}_i}} \right) \right) \quad (17.53)$$

Or nous avons :

$$\frac{\partial}{\partial \tilde{f}_i} \left( 1 - (1 - \rho)^{e^{-\tilde{f}_i}} \right) = -\frac{\partial}{\partial \tilde{f}_i} \left( (1 - \rho)^{e^{-\tilde{f}_i}} \right) \quad (17.54)$$

$$= -\frac{\partial}{\partial \tilde{f}_i} \left( \exp(e^{-\tilde{f}_i} \ln(1 - \rho)) \right) \quad (17.55)$$

$$= e^{-\tilde{f}_i} \ln(1 - \rho) (1 - \rho)^{e^{-\tilde{f}_i}} \quad (17.56)$$

$$= \frac{1}{f_i} \ln(1 - \rho) (1 - \rho)^{\frac{1}{f_i}} \quad (17.57)$$

Nous en déduisons :

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{f}_i} = \frac{1}{\tau \rho} \left( e^{\tilde{f}_i - \tilde{q}_i} \left( 1 - (1 - \rho)^{e^{-\tilde{f}_i}} \right) + e^{\tilde{f}_i - \tilde{q}_i} \frac{1}{f_i} \ln(1 - \rho) (1 - \rho)^{\frac{1}{f_i}} \right) \quad (17.58)$$

$$= \frac{1}{\tau \rho} \left( \frac{f_i}{q_i} \left( 1 - (1 - \rho)^{\frac{1}{f_i}} \right) + \frac{f_i}{q_i} \frac{1}{f_i} \ln(1 - \rho) (1 - \rho)^{\frac{1}{f_i}} \right) \quad (17.59)$$

$$= \frac{1}{\tau \rho} \frac{f_i}{q_i} \left( \left( 1 - (1 - \rho)^{\frac{1}{f_i}} \right) + \frac{1}{f_i} \ln(1 - \rho) (1 - \rho)^{\frac{1}{f_i}} \right) \quad (17.60)$$

#### 17.1.4.2.3 Dérivées partielles par rapport à $\tilde{v}_i$

$$\frac{\partial \tilde{h}z_i}{\partial \tilde{v}_i} = 0 \quad (17.61)$$

$$\frac{\partial \tilde{h}f_i}{\partial \tilde{v}_i} = 0 \quad (17.62)$$

$$\frac{\partial \tilde{h}v_i}{\partial \tilde{v}_i} = \frac{1}{\tau} \left( f_i \times (-e^{-\tilde{v}_i}) - \left( \frac{1}{\alpha} - 1 \right) e^{(\frac{1}{\alpha}-1)\tilde{v}_i} \right) = -\frac{1}{\tau} \left( \frac{f_i}{v_i} + \left( \frac{1}{\alpha} - 1 \right) v_i^{\frac{1}{\alpha}-1} \right) \quad (17.63)$$

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{v}_i} = -\frac{1}{\tau} \left( \frac{1}{\alpha} - 1 \right) e^{(\frac{1}{\alpha}-1)\tilde{v}_i} = \frac{1}{\tau} \left( 1 - \frac{1}{\alpha} \right) v_i^{\frac{1}{\alpha}-1} \quad (17.64)$$

#### 17.1.4.2.4 Dérivées partielles par rapport à $\tilde{q}_i$

$$\frac{\partial \tilde{h}z_i}{\partial \tilde{q}_i} = 0 \quad (17.65)$$

$$\frac{\partial \tilde{h}f_i}{\partial \tilde{q}_i} = 0 \quad (17.66)$$

$$\frac{\partial \tilde{h}v_i}{\partial \tilde{q}_i} = 0 \quad (17.67)$$

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{q}_i} = -\frac{1}{\tau \rho} e^{\tilde{f}_i - \tilde{q}_i} \left( 1 - (1 - \rho)^{\frac{1}{\tilde{f}_i}} \right) = -\frac{1}{\tau \rho} \frac{f_i}{q_i} \left( 1 - (1 - \rho)^{\frac{1}{f_i}} \right) \quad (17.68)$$

#### 17.1.4.3 Calculs des dérivées partielles par rapport aux paramètres à inverser

Pour chaque région  $i$ , nous donnons les éléments de matrice de la jacobienne  $\overline{\frac{\partial H}{\partial \theta}}$ , c'est-à-dire les dérivées partielles des 4 fonctions  $\tilde{h}z_i, \tilde{h}f_i, \tilde{h}v_i, \tilde{h}q_i$  par rapport à chacun des paramètres à inverser, ie  $\tilde{s}n_{B_i}$  et  $\tilde{s}n_{stim_i}$  pour la région  $i$  ainsi que  $\tilde{s}n_{B_j}$  et  $\tilde{s}n_{stim_j}$  pour toute autre région  $j$ .

Puisque le modèle hémodynamique n'est pas couplé entre régions, nous pouvons écrire directement :

$$\frac{\partial \tilde{h}z_i}{\partial \tilde{s}n_{B_j \neq i}} = \frac{\partial \tilde{h}z_i}{\partial \tilde{s}n_{stim_j \neq i}} = 0 \quad (17.69)$$

et de même pour les fonctions  $\tilde{h}f_i, \tilde{h}v_i, \tilde{h}q_i$ .

##### 17.1.4.3.1 Dérivées partielles par rapport à $\tilde{s}n_{B_i}$

Nous avons :

$$\frac{\partial \Phi(\tilde{s}n_{B_i})}{\partial \tilde{s}n_{B_i}} = \tilde{s}n_{B_i} (1 - \tilde{s}n_{B_i}) \quad (17.70)$$

Ce qui nous permet de déduire que :

$$\frac{\partial \tilde{h}z_i}{\partial \tilde{s}n_{B_i}} = \frac{\partial sn_{B_i}}{\partial \tilde{s}n_{B_i}} \delta_{u(t),0} = sn_{B_i}(1 - sn_{B_i}) \delta_{u(t),0} \quad (17.71)$$

$$\frac{\partial \tilde{h}f_i}{\partial \tilde{s}n_{B_i}} = 0 \quad (17.72)$$

$$\frac{\partial \tilde{h}v_i}{\partial \tilde{s}n_{B_i}} = 0 \quad (17.73)$$

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{s}n_{B_i}} = 0 \quad (17.74)$$

#### 17.1.4.3.2 Dérivées partielles par rapport à $\tilde{s}n_{stim_i}$

Nous avons :

$$\frac{\partial \Phi(sn_{stim_i})}{\partial \tilde{s}n_{stim_i}} = sn_{stim_i}(1 - sn_{stim_i}) \quad (17.75)$$

Nous déduisons :

$$\frac{\partial \tilde{h}z_i}{\partial \tilde{s}n_{stim_i}} = \frac{\partial sn_{stim_i}}{\partial \tilde{s}n_{stim_i}} \delta_{u(t),1} = sn_{stim_i}(1 - sn_{stim_i}) \delta_{u(t),1} \quad (17.76)$$

$$\frac{\partial \tilde{h}f_i}{\partial \tilde{s}n_{stim_i}} = 0 \quad (17.77)$$

$$\frac{\partial \tilde{h}v_i}{\partial \tilde{s}n_{stim_i}} = 0 \quad (17.78)$$

$$\frac{\partial \tilde{h}q_i}{\partial \tilde{s}n_{stim_i}} = 0 \quad (17.79)$$

#### 17.1.4.4 Jacobiennes à coder dans la toolbox VBA

A la section 10.3.3.1, nous avons échantillonné le temps en pas de durée  $\Delta t$  et défini une fonction  $\vec{e}$  qui permet d'obtenir  $\vec{x}(t_{m-1} + \Delta t)$  en fonction de  $\vec{x}(t_{m-1})$  :

$$\vec{x}(t_{m-1} + \Delta t) \approx \vec{x}(t_{m-1}) + \Delta t \vec{H}(\vec{x}(t_{m-1}), \vec{\theta}, u(t_{m-1})) \equiv \vec{e}(\vec{x}(t_{m-1}), \vec{\theta}, u(t_{m-1})) \quad (17.80)$$

C'est cette fonction  $\vec{e}$  que nous devons coder dans VBA et pour laquelle nous devons aussi fournir les jacobiennes par rapport aux variables d'état ( $\overline{\frac{\partial \vec{e}}{\partial x}}$ ) et par rapport aux paramètres à inverser ( $\overline{\frac{\partial \vec{e}}{\partial \theta}}$ ).

Nous pouvons calculer les jacobiennes de  $\vec{e}$  à partir de celles de  $\vec{H}$  :

$$\overline{\frac{\partial \vec{e}}{\partial x}} = \mathbb{I} + \Delta t \overline{\frac{\partial \vec{H}}{\partial x}} \quad (17.81)$$

$$\overline{\frac{\partial \vec{e}}{\partial \theta}} = \mathbb{I} + \Delta t \overline{\frac{\partial \vec{H}}{\partial \theta}} \quad (17.82)$$

## 17.1.5 Calcul des jacobiennes du modèle algébrique BOLD transformé

### 17.1.5.1 Modèle BOLD transformé

Nous avons vu à la section 10.3.2, la nécessité de réécrire le modèle BOLD avec une transformation des variables d'état de volume  $v_i$  et quantité de déoxyhémoglobine  $q_i$ . Le modèle BOLD transformé s'écrit, pour chaque région  $i$  :

$$y_i = V_0 \left( k_1(1 - e^{\tilde{q}_i}) + k_2(1 - e^{\tilde{q}_i - \tilde{v}_i}) + k_3(1 - e^{\tilde{v}_i}) \right) \equiv \tilde{B}_i(\tilde{v}_i, \tilde{q}_i) \quad (17.83)$$

La fonction vectorielle  $\vec{B}$  nous permet d'écrire le modèle BOLD transformé pour toutes les régions à tout instant  $t$  :

$$\boxed{\vec{y}(t) = \vec{B}(\vec{x}(t))} \quad (17.84)$$

### 17.1.5.2 Calculs des dérivées partielles par rapport aux variables d'état

Pour chaque région  $i$ , nous donnons les éléments de matrice de la jacobienne  $\overline{\frac{\partial \tilde{B}}{\partial x}}$ , c'est-à-dire les dérivées partielles de la fonction  $\tilde{B}_i$  par rapport à chacune des variables  $z_i, \tilde{f}_i, \tilde{q}_i, \tilde{v}_i$ , ainsi que  $z_j, \tilde{f}_j, \tilde{q}_j, \tilde{v}_j$  pour toute autre région  $j$ .

Le modèle hémodynamique étant découpé entre régions, nous avons :

$$\frac{\partial \tilde{B}_i}{\partial z_{j \neq i}} = \frac{\partial \tilde{B}_i}{\partial \tilde{f}_{j \neq i}} = \frac{\partial \tilde{B}_i}{\partial \tilde{v}_{j \neq i}} = \frac{\partial \tilde{B}_i}{\partial \tilde{q}_{j \neq i}} = 0 \quad (17.85)$$

De plus, la fonction  $\tilde{B}_i$  ne dépend pas de  $z_i$  ni de  $\tilde{f}_i$  donc :

$$\frac{\partial \tilde{B}_i}{\partial z_i} = \frac{\partial \tilde{B}_i}{\partial \tilde{f}_i} = 0 \quad (17.86)$$

Nous calculons les dérivées partielles par rapport à  $\tilde{v}_i$  et  $\tilde{q}_i$  :

$$\frac{\partial \tilde{B}_i}{\partial \tilde{v}_i} = V_0 (k_2 e^{\tilde{q}_i - \tilde{v}_i} - k_3 e^{\tilde{v}_i}) = V_0 \left( k_2 \frac{q_i}{v_i} - k_3 v_i \right) \quad (17.87)$$

$$\frac{\partial \tilde{B}_i}{\partial \tilde{q}_i} = -V_0 (k_1 e^{\tilde{q}_i} + k_2 e^{\tilde{q}_i - \tilde{v}_i}) = -V_0 \left( k_1 q_i + k_2 \frac{q_i}{v_i} \right) \quad (17.88)$$

### 17.1.5.3 Jacobienne à coder dans la toolbox VBA

A la section 10.3.3.2, nous avons discrétisé le temps et introduit une nouvelle notation pour la fonction  $\vec{B}$ , de manière à reprendre celle de VBA.  $\vec{g}$  qui donne le signal BOLD à l'instant  $t_m$  en fonction des variables d'état hémodynamique à  $t_m$  :

$$\vec{y}(t_m) = \vec{B}(\vec{x}(t_m)) \equiv \vec{g}(\vec{x}(t_m)) \quad (17.89)$$

Nous devons fournir à VBA :

$$\frac{\overline{\partial \vec{g}}}{\partial x} = \frac{\overline{\partial \vec{B}}}{\partial x} \quad (17.90)$$

Nous n'inversons pas de paramètres  $\vec{\Phi}$  dont dépend directement la fonction  $\vec{g}$ . VBA propose cette possibilité. Il faudrait alors calculer en plus la jacobienne  $\frac{\overline{\partial \vec{g}}}{\partial \vec{\Phi}}$ .

### 17.1.6 Sorties d'inversion des $sn$ sur simulation

A la section 10.4, nous avons simulé des données BOLD pour  $N = 15$  régions, sur 940  $TR$ . Pour cela, nous avons fixé : une condition initiale  $\vec{x}_{0_{simu}}$ , des paramètres  $\vec{\theta}_{simu}$  (deux, trois ou quatre  $\tilde{sn}$  par région, tirés aléatoirement dans la gaussienne a priori  $p(\vec{\theta})$ ) et l'inverse de la variance du bruit ( $\sigma_{simu}$ ). Nous avons inversé ces simulations avec la procédure d'inversion présentée à la section 10.3, en distinguant deux cas : la condition initiale  $\vec{x}_{0_{simu}}$  est contenue ou non dans l'a priori  $p(\vec{x}_0)$ .

#### 17.1.6.1 Cas où la condition initiale est contenue dans l'a-priori

Nous présentons le cas de l'inversion de 2 paramètres ( $\tilde{sn}_B$  et  $\tilde{sn}_{NT}$ ), 3 paramètres ( $\tilde{sn}_B$ ,  $\tilde{sn}_T$  et  $\tilde{sn}_{NT}$ ) et 4 paramètres ( $\tilde{sn}_B$ ,  $\tilde{sn}_T$ ,  $\tilde{sn}_{NT,int}$  et  $\tilde{sn}_{NT,Nint}$ ), comme exposé à la section 10.3.6.

Pour un niveau de bruit fixé (A :  $\sigma_{simu} = \infty$ , B :  $\sigma_{simu} = 10^4$ , C :  $\sigma_{simu} = 10^3$  et D :  $\sigma_{simu} = 10^2$ ), nous avons représenté pour chaque région la valeur moyenne estimée pour un paramètre hémodynamique transformé, en fonction de sa valeur fixée pour la simulation. Plus  $\sigma$  (inverse de la variance du bruit) diminue plus le niveau de bruit de la simulation est grand.

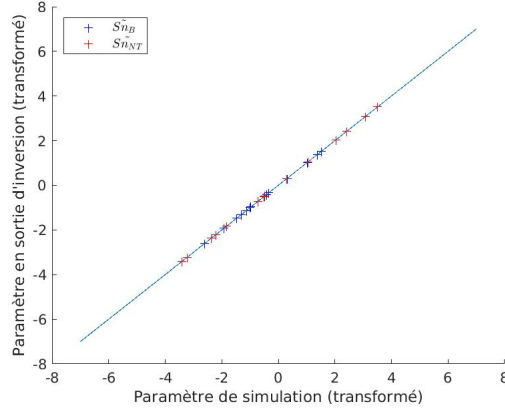
Nous vérifions que nous retrouvons exactement les paramètres de simulation en sortie d'inversion lorsque le signal n'est pas bruité ( $\sigma_{simu} = \infty$ ). Cela valide la première étape de la procédure d'inversion, sans bruit. Comme attendu, plus la simulation est bruitée et moins bien nous récupérons les paramètres de simulation en sortie d'inversion.

Au chapitre 7, nous avons vu que sous l'hypothèse de notre modèle l'amplitude maximale du signal BOLD d'une région vaut 0.04<sup>1</sup>. A partir du cas B, nous travaillons donc avec des signaux dont l'écart-type du bruit est du même ordre de grandeur ( $10^{-2}$ ) ou d'un ordre de grandeur supérieur à l'amplitude maximale du signal. Les signaux simulés sont donc très bruités, notamment dans le cas D.

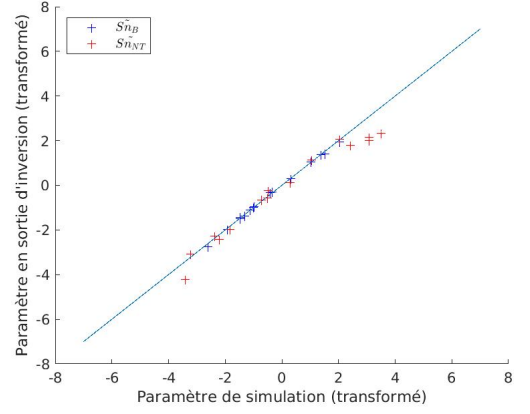
1. différences entre les valeurs de BOLD obtenues pour  $sn = 1$  et pour  $sn = 0$

### 17.1.6.1.1 2 paramètres transformés

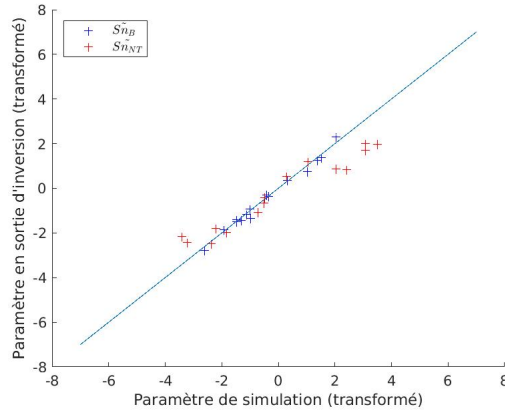
$$\sigma_{simu} = \infty$$



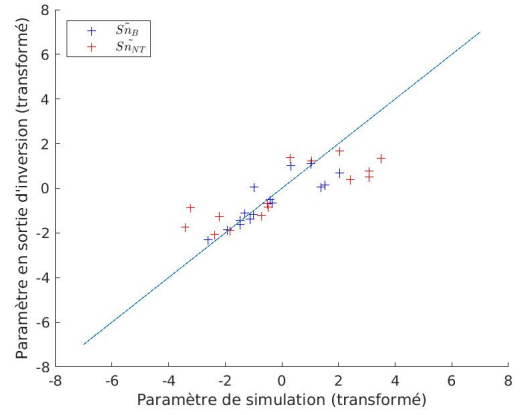
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$



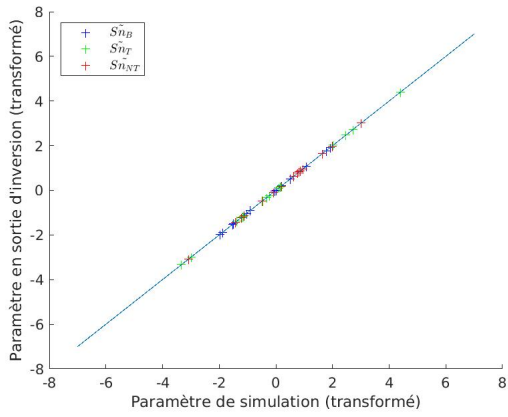
$$\sigma_{simu} = 10^2$$



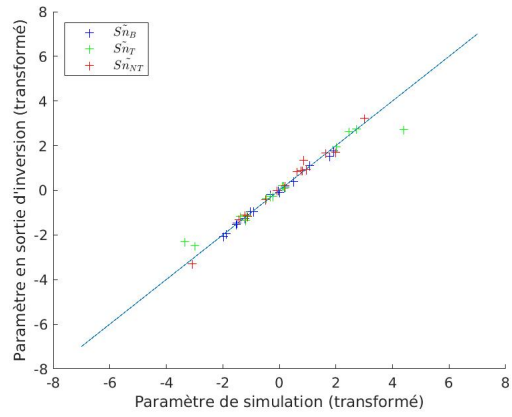


### 17.1.6.1.2 3 paramètres transformés

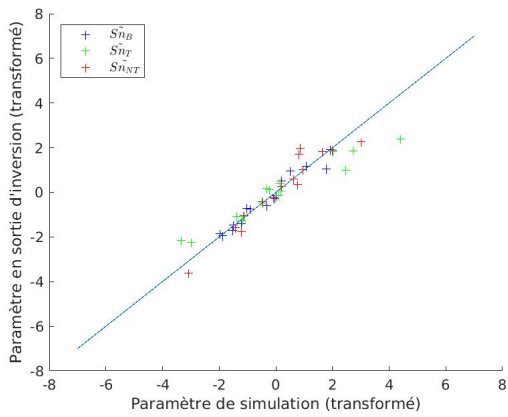
$$\sigma_{simu} = \infty$$



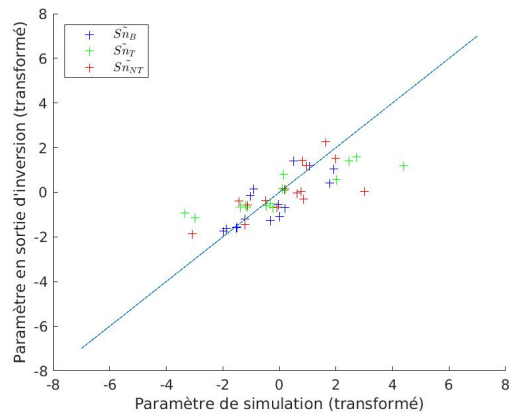
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$

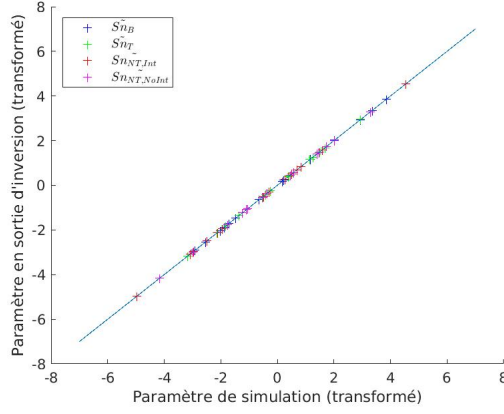


$$\sigma_{simu} = 10^2$$

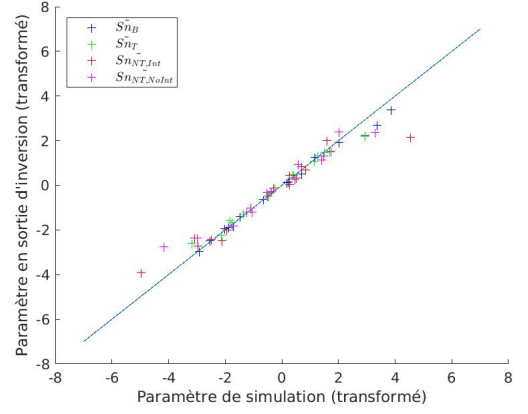


### 17.1.6.1.3 4 paramètres transformés

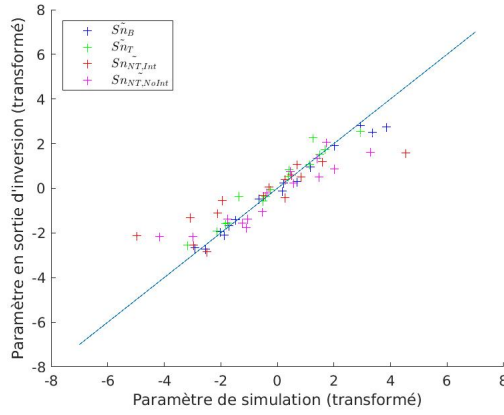
$$\sigma_{simu} = \infty$$



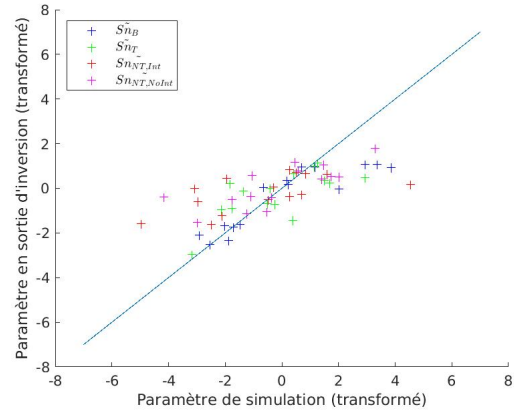
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$

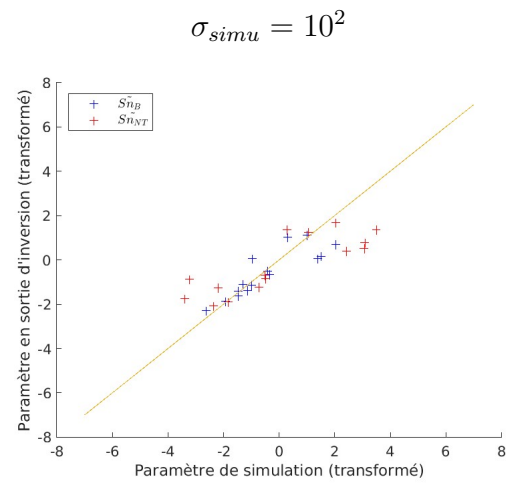
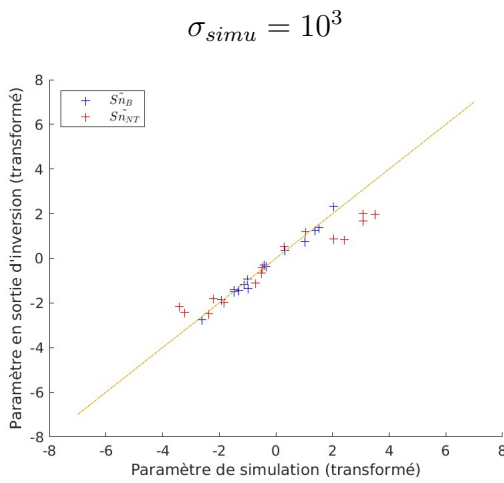
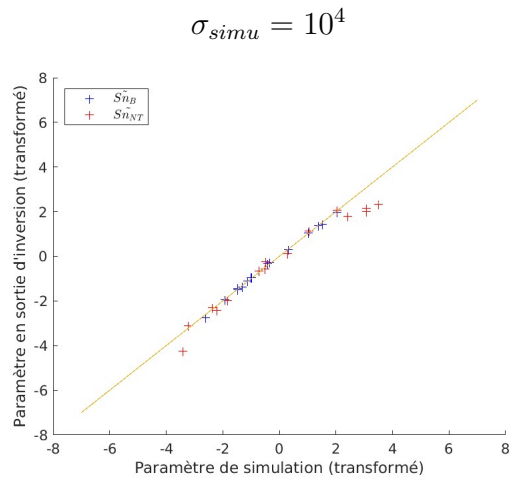
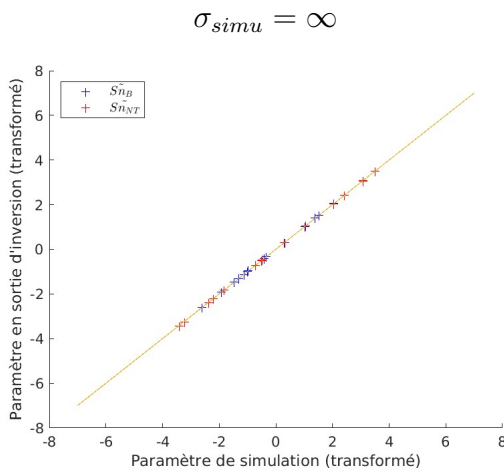


$$\sigma_{simu} = 10^2$$



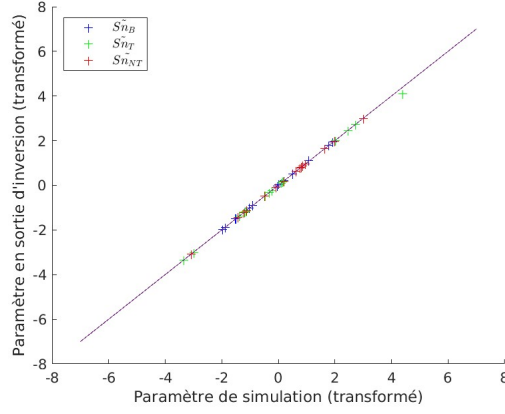
## 17.1.6.2 Cas où la condition initiale est éloignée de l'a-priori

### 17.1.6.2.1 2 paramètres transformés

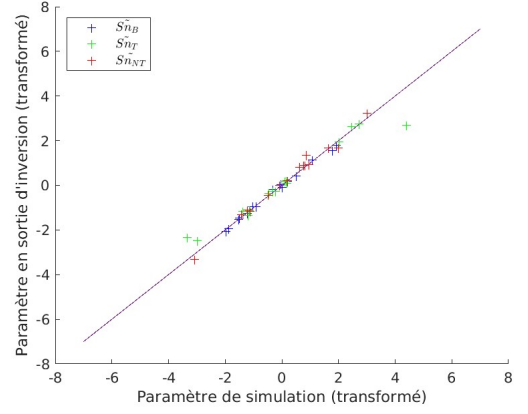


### 17.1.6.2.2 3 paramètres transformés

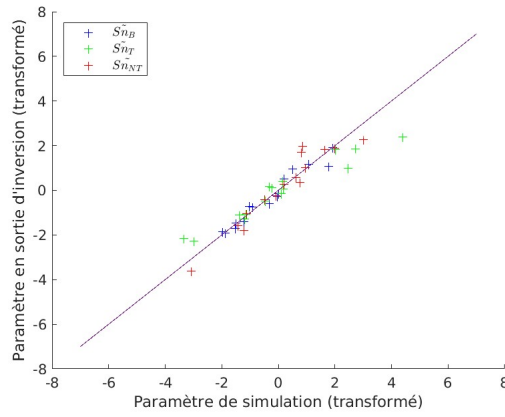
$$\sigma_{simu} = \infty$$



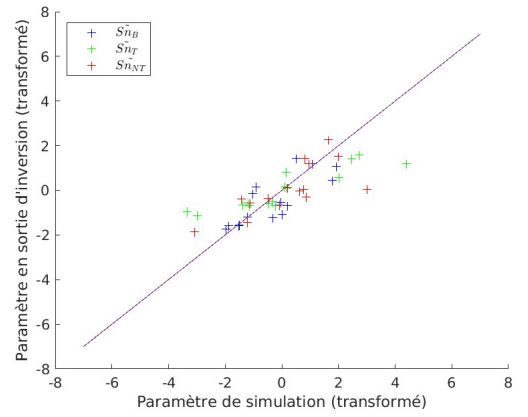
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$

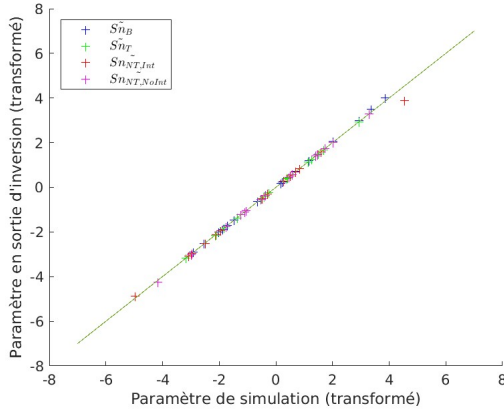


$$\sigma_{simu} = 10^2$$

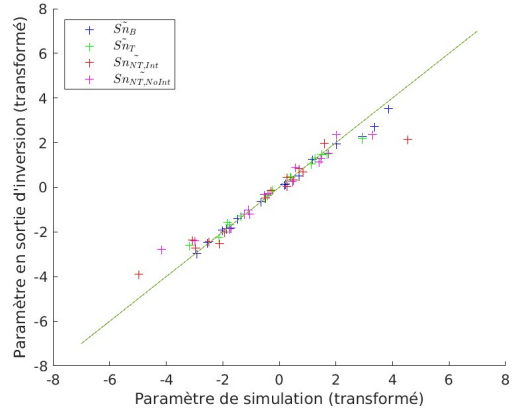


### 17.1.6.2.3 4 paramètres transformés

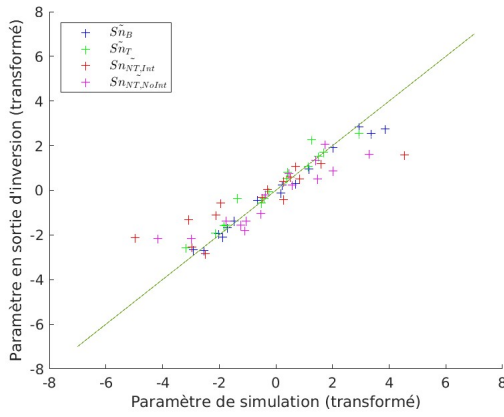
$$\sigma_{simu} = \infty$$



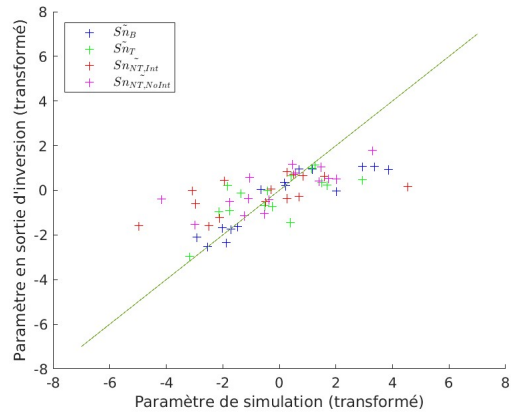
$$\sigma_{simu} = 10^4$$



$$\sigma_{simu} = 10^3$$



$$\sigma_{simu} = 10^2$$



## 17.2 Deuxième étape de la procédure d'inversion

### 17.2.1 Exploration des points fixes pour un pool inhibiteur

Nous étudions le système composé d'un pool inhibiteur isolé, pour lequel nous cherchons à préciser l'état stationnaire (notamment le nombre de points fixes et leur stabilité). Le modèle dynamique s'écrit :

$$\begin{cases} sg(t) = fg(sg(t), rg(t)) = -\beta^I sg(t) + \alpha^I T_{gaba} rg(t)(1 - sg(t)) \\ rg(t) = hg(xg(t)) = \frac{a_I xg(t) - b_I}{1 - e^{-d_I(a_I xg(t) - b_I)}} \\ xg(t) = wg(sg(t), z_I) = -W_- sg(t) + z_I \end{cases} \quad (17.91)$$

Avec :

$$\beta^I = 180 \text{ s}^{-1}, \alpha^I = 530 \text{ s}^{-1} \text{ mM}^{-1}, T_{gaba} = 3 \cdot 10^{-3} \text{ mM s.}^2.$$

$$a_I = 615 \text{ nC}^{-1}, b_I = 177 \text{ Hz}, d_I = 0.087 \text{ s}.$$

$$W_- = 1 \text{ nA}$$

#### 17.2.1.1 Recherche de point(s) fixe(s)

Nous pouvons réécrire le modèle sous la forme :

$$sg(t) = Fg(sg(t), z_I) \quad (17.92)$$

Pour une valeur fixée du forçage  $z_I$ , il est possible de tracer la fonction  $Fg$  pour tout  $sg \in [0, 1]$ . C'est ce que nous représentons à la figure 17.3, pour différentes valeurs fixées du forçage échantillonnées entre  $-2$  et  $2$ .

Pour  $z_I$  fixé, un point fixe  $sg^*$  vérifie que  $Fg(sg^*, z_I) = 0$  (intersection d'une courbe avec  $y = 0$ ). Le nombre de points fixes est donné par le nombre d'antécédents  $sg^*$  qui annulent la fonction  $Fg$ , i.e. par le nombre d'intersections avec l'axe  $y = 0$ . Il est toujours possible de rechercher les 0 de la fonction par dichotomie par exemple.

#### 17.2.1.2 Existence d'un unique point fixe

Nous remarquons sur la figure 17.3 que pour  $z_I$  fixé la fonction ne s'annule qu'une seule fois sur  $[0, 1]$ . Pour chaque valeur de forçage  $z_I$ , il n'existe donc qu'un seul point fixe  $sg^*$ . Pour des valeurs de  $z_I$  entre  $-2$  et  $2$ , nous avons recherché  $sg^*$  par dichotomie. Nous représentons à la figure 17.4 les valeurs du point fixe  $sg^*$  en fonction de  $z_I$ . Nous avons fait ressortir les points fixes illustrés à la figure 17.3 pour les valeurs de  $z_I$  correspondantes (abscisses des étoiles bleues).

---

2. médiane approchée de l'intervalle que donnent [Naskar et al., 2021] chez le sujet sains :  $T_{gaba} \in [1.5; 4.2] \text{ mM ms}$

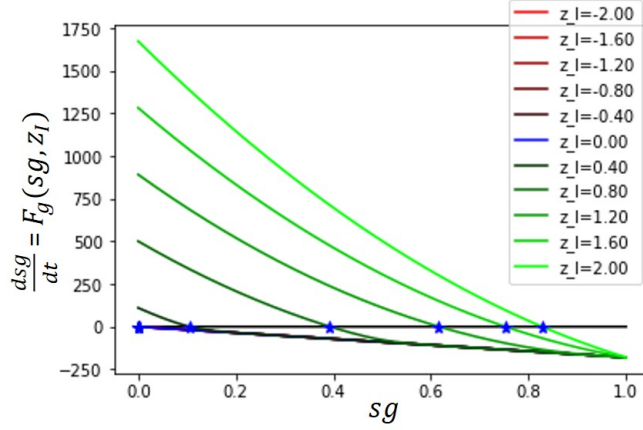


**Figure 17.3** – Diagramme de phase  $\dot{sg}$  en fonction de  $sg$ , pour un autocouplage fixé  $W_- = 1$  et différentes valeurs de forçage  $z_I$ .

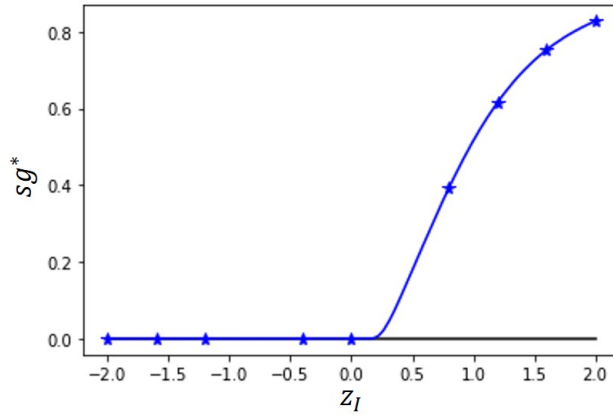


**Figure 17.4** – Point fixe  $sg^*$  pour un autocouplage fixé  $W_- = 1$ , en fonction de  $z_I$

### 17.2.1.3 Etude de la stabilité

Pour  $z_I$  fixé, la stabilité du point fixe est donnée par le signe de  $\frac{\partial Fg}{\partial sg}(sg^*, z_I)$ , autrement dit par le signe de la pente de la fonction  $Fg$  évaluée au point fixe (étoile bleue sur la figure 17.3).

Si  $\frac{\partial Fg}{\partial sg}(sg^*, z_I) < 0$  (i.e. pente négative) alors  $sg^*$  est un point fixe stable.

En effet, si  $sg < sg^*$  (à gauche de l'étoile bleue), alors  $Fg(sg, z_I) = \dot{sg} > 0$ . Si nous faisons tourner la dynamique en prenant une condition initiale telle que  $sg(0) < sg^*$ ,  $sg(t)$  augmenterait et nous nous rapprocherions du point fixe (i.e. de l'étoile bleue).

Inversement, si  $sg > sg^*$  (à droite de l'étoile bleue), alors  $Fg(sg, z_I) = \dot{sg} < 0$ . Si nous faisons tourner la dynamique en prenant une condition initiale telle que  $sg(0) > sg^*$ ,  $sg(t)$  diminuerait et nous nous rapprocherions du point fixe. Le point fixe est donc stable.

Si  $\frac{\partial Fg}{\partial sg}(sg^*, z_I) > 0$  (i.e. pente positive) alors  $sg^*$  est un point fixe instable.

Un raisonnement similaire à celui qui vient d'être présenté nous permet de comprendre qu'en faisant tourner la dynamique pour une condition initiale à gauche ( $sg(0) < sg^*$ ) ou à droite ( $sg(0) > sg^*$ ) du point fixe, nous ne pourrions que nous éloigner du point fixe. Celui-ci est

donc instable.

D'après la figure 17.3, pour  $z_I$  fixé, il existe un point fixe stable.

## 17.2.2 Propriétés des fonctions $hn$ et $hg$ qui donnent les fréquences de potentiels d'action

Soit la fonction  $H$  telle que

$$\forall x \in \mathbb{R} \setminus \left\{ \frac{b}{a} \right\}, \quad H(x) = \frac{ax - b}{1 - e^{-d(ax-b)}} \quad (17.93)$$

### 17.2.2.1 Etude du signe

D'une part, nous avons :

$$ax - b \geq 0 \iff x \geq \frac{b}{a} \quad (17.94)$$

D'autre part,

$$1 - e^{-d(ax-b)} \geq 0 \iff 1 \geq e^{-d(ax-b)} \quad (17.95)$$

$$\iff 0 \geq -d(ax - b) \quad (17.96)$$

$$\iff \frac{b}{a} \leq x \quad (17.97)$$

Nous pouvons alors construire le tableau de signe suivant :

| $x$                | $-\infty$ | $0$                   | $\frac{b}{a}$ | $+\infty$ |
|--------------------|-----------|-----------------------|---------------|-----------|
| $ax - b$           | $-$       | $-$                   |               | $+$       |
| $1 - e^{-d(ax-b)}$ | $-$       | $-$                   |               | $+$       |
| $H(x)$             | $+$       | $\frac{-b}{1-e^{db}}$ | $+$           | $+$       |

Nous en déduisons que la fonction  $H$  est strictement positive sur  $\mathbb{R} \setminus \left\{ \frac{b}{a} \right\}$

### 17.2.2.2 Etude des variations

Calculons la dérivée de  $H$  par rapport à  $x$  :

$$\forall x \in \mathbb{R} \setminus \left\{ \frac{b}{a} \right\}, \quad \frac{\partial H}{\partial x} = \frac{a - ae^{-d(ax-b)} - (ax - b)(dae^{-d(ax-b)})}{(1 - e^{-d(ax-b)})^2} \quad (17.98)$$

$$(17.99)$$

Nous posons

$$\forall x \in \mathbb{R} \setminus \left\{ \frac{b}{a} \right\}, \quad \tilde{X} = ax - b \quad (17.100)$$



Le signe de  $\frac{\partial H}{\partial x}$  est le même que celui du numérateur qui peut se réécrire :

$$a - ae^{-d\tilde{X}} - da\tilde{X}e^{-d\tilde{X}} = \underbrace{-e^{-d\tilde{X}}a}_{<0} \underbrace{(-e^{d\tilde{X}} + 1 + d\tilde{X})}_{G(\tilde{X})} \quad (17.101)$$

Nous définissons la fonction  $G$  telle que :

$$\forall \tilde{X} \in \mathbb{R}^*, \quad G(\tilde{X}) = -e^{d\tilde{X}} + 1 + d\tilde{X} \quad (17.102)$$

Pour connaître le signe de  $\frac{\partial H}{\partial x}$ , nous avons besoin de connaître celui de  $G(\tilde{X})$ . Pour connaître le signe de  $G(\tilde{X})$ , nous allons étudier les variations de la fonction  $G$ .

Etude des variations de  $G$  pour en déduire son signe

Nous pouvons calculer sa dérivée par rapport à  $\tilde{X}$  :

$$\forall \tilde{X} \in \mathbb{R}^*, \quad \frac{\partial G}{\partial \tilde{X}} = -de^{d\tilde{X}} + d = \underbrace{d}_{>0} (1 - e^{d\tilde{X}}) \quad (17.103)$$

Or,

$$1 - e^{d\tilde{X}} \geq 0 \iff 1 \geq e^{d\tilde{X}} \quad (17.104)$$

$$\iff 0 \geq \tilde{X} \quad (17.105)$$

Nous pouvons alors construire le tableau de variation suivant :

| $x$                                     | $-\infty$ | $0$ | $+\infty$ |
|-----------------------------------------|-----------|-----|-----------|
| $\frac{\partial G}{\partial \tilde{X}}$ | +         |     | -         |
| $G(\tilde{X})$                          |           |     |           |

Si  $G$  était définie sur  $\mathbb{R}$ ,  $G(0) = 0$ .  $G$  aurait été négative sur  $\mathbb{R}$ .

D'après les variations de  $G$ , la fonction est strictement négative sur  $\mathbb{R}^*$  :  $\forall \tilde{X} \in \mathbb{R}^* G(\tilde{X}) < 0$ .

Deduction du signe de  $\frac{\partial H}{\partial x}$  et variations de  $H$

D'après l'équation 17.101, nous pouvons alors en déduire que  $\forall x \in \mathbb{R} \setminus \left\{\frac{b}{a}\right\} \frac{\partial H}{\partial x} > 0$ . Donc  $H$  est une fonction strictement positive de  $x$ .

Voici son tableau de variations :

| $x$                             | $-\infty$ | $0$ | $\frac{b}{a}$ | $+\infty$ |
|---------------------------------|-----------|-----|---------------|-----------|
| $\frac{\partial H}{\partial x}$ | +         |     |               | +         |
| $H(x)$                          |           |     |               |           |

### 17.2.2.3 Fréquences de potentiels d'action pour les pools excitateur et inhibiteur

- Pour un pool excitateur, la fréquence de potentiels d'action  $rn$  s'écrit en fonction du courant  $xn$  :

$$rn = hn(xn) = \frac{a_E xn - b_E}{1 - e^{-d_E(a_E xn - b_E)}}, \text{ avec } a_E = 310nC^{-1}, b_E = 125Hz, d_E = 0.16s \quad (17.106)$$

- Pour le pool inhibiteur, la fréquence de potentiels d'action  $rg$  s'écrit en fonction du courant  $xg$  :

$$rg = hg(xg) = \frac{a_I xg - b_I}{1 - e^{-d_I(a_I xg - b_I)}}, \text{ avec } a_I = 615nC^{-1}, b_I = 177Hz, d_I = 0.087s \quad (17.107)$$

Nous représentons ces deux fonctions dans les figures 17.5 et 17.6.

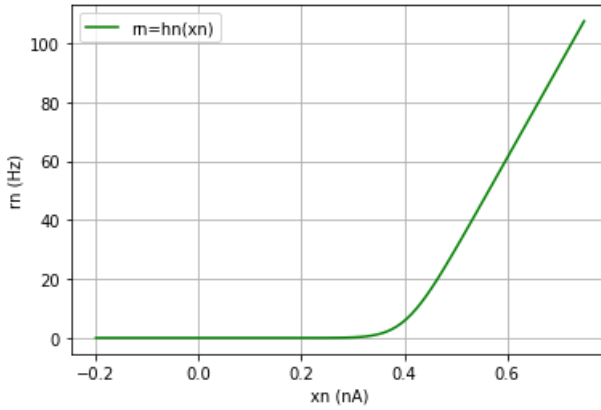


**Figure 17.5** – Fréquence de potentiels d'action  $rn$  en fonction du courant  $xn$  sur le pool excitateur d'une région

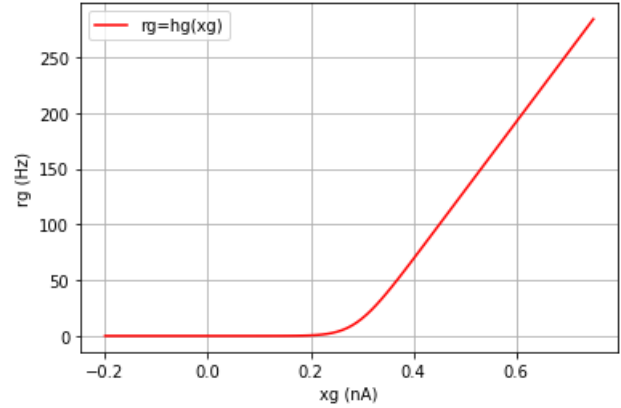


**Figure 17.6** – Fréquence de potentiels d'action  $rg$  en fonction du courant  $xg$  sur le pool inhibiteur d'une région



# Chapitre 18

## Annexes de la partie V : Discussion

### 18.1 Balance E/I : une région isolée

#### 18.1.1 Analyse de sensibilité : perturbation des paramètres dans les courants

L'objectif de cette partie est de mener une analyse de sensibilité à la perturbation des paramètres de couplage ( $J_{nmda}$ ,  $J_{gaba}$ ,  $W_+$ ,  $W_-$ ) et de forçage ( $x_E$ ,  $x_I$ ) qui se trouvent dans l'expression des courants du modèle de [Naskar et al., 2021].

##### 18.1.1.1 Formulation mathématique du modèle pour une région isolée

Pour une région isolée, le modèle dynamique s'écrit :

$$\begin{cases} \dot{sn}(t) = -\beta^E sn(t) + \alpha^E T_{glu}(1 - sn(t))rn(t) \equiv fn(sn(t), rn(t)) \\ \dot{sg}(t) = -\beta^I sg(t) + \alpha^I T_{gaba}(1 - sg(t))rg(t) \equiv fg(sg(t), rg(t)) \end{cases} \quad (18.1)$$

$$\begin{cases} rn(t) = \frac{a_E xn(t) - b_E}{1 - e^{-d_E(a_E xn(t) - b_E)}} \equiv hn(xn(t)) \\ rg(t) = \frac{a_I xg(t) - b_I}{1 - e^{-d_I(a_I xg(t) - b_I)}} \equiv hg(xg(t)) \end{cases} \quad (18.2)$$

$$\begin{cases} xn(t) = W_+ J_{nmda} sn(t) - J_{gaba} sg(t) + x_E \equiv wn(sn(t), sg(t), x_E, J_{gaba}, W_+, J_{nmda}) \\ xg(t) = J_{nmda} sn(t) - W_- sg(t) + x_I \equiv wg(sn(t), sg(t), x_I, J_{nmda}, W_-) \end{cases} \quad (18.3)$$

Nous écrivons les courants comme fonctions explicites des paramètres que nous allons perturber.

Au régime stationnaire , le modèle mathématique d'intérêt pour une région peut s'écrire :

$$\begin{cases} 0 = fn(sn^*, rn^*) \\ 0 = fg(sg^*, rg^*) \end{cases} \quad (18.4)$$

$$\begin{cases} rn^* = hn(xn^*) \\ rg^* = hg(xg^*) \end{cases} \quad (18.5)$$

$$\begin{cases} xn^* = wn(sn^*, sg^*, x_E, J_{gaba}, W_+, J_{nmda}) \\ xg^* = wg(sn^*, sg^*, x_I, J_{nmda}, W_-) \end{cases} \quad (18.6)$$

Nous définissons :

$$\vec{s} = \begin{pmatrix} sn^* \\ sg^* \end{pmatrix}, \quad \vec{r} = \begin{pmatrix} rn^* \\ rg^* \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} xn^* \\ xg^* \end{pmatrix} \quad (18.7)$$

Au point fixe, nous pouvons écrire :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{s}, x_E, x_I, J_{nmda}, J_{gaba}, W_+, W_-) \end{cases} \quad (18.8)$$

### 18.1.1.2 Analyse de sensibilité

Nous perturbons les forçages  $x_E$ ,  $x_I$  et les poids synaptiques  $J_{nmda}$ ,  $J_{gaba}$ ,  $W_+$  et  $W_-$ . Le nouveau modèle perturbé s'écrit :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\vec{s}, x_E + \delta x_E, x_I + \delta x_I, J_{nmda} + \delta J_{nmda}, J_{gaba} + \delta J_{gaba}, W_+ + \delta W_+, W_- + \delta W_-) \end{cases} \quad (18.9)$$

Nous linéarisons autour du point fixe :

$$\begin{cases} \frac{\partial \vec{f}}{\partial s} \delta \vec{s} + \frac{\partial \vec{f}}{\partial r} \delta \vec{r} = \vec{0} \\ \delta \vec{r} = \frac{\partial \vec{h}}{\partial x} \delta \vec{x} \\ \delta \vec{x} = \frac{\partial \vec{w}}{\partial s} \delta \vec{s} + \frac{\partial \vec{w}}{\partial x_E} \delta x_E + \frac{\partial \vec{w}}{\partial x_I} \delta x_I + \frac{\partial \vec{w}}{\partial J_{nmda}} \delta J_{nmda} + \frac{\partial \vec{w}}{\partial J_{gaba}} \delta J_{gaba} + \frac{\partial \vec{w}}{\partial W_+} \delta W_+ + \frac{\partial \vec{w}}{\partial W_-} \delta W_- \end{cases} \quad (18.10)$$

Par éliminations successives nous obtenons :

$$\overline{\overline{\frac{\partial f}{\partial s}}} \vec{\delta s} + \overline{\overline{\frac{\partial f}{\partial r} \frac{\partial h}{\partial x}}} \left( \overline{\overline{\frac{\partial w}{\partial s}}} \vec{\delta s} + \frac{\vec{\partial w}}{\partial x_E} \delta x_E + \frac{\vec{\partial w}}{\partial x_I} \delta x_I + \frac{\vec{\partial w}}{\partial J_{nm da}} \delta J_{nm da} + \frac{\vec{\partial w}}{\partial J_{gaba}} \delta J_{gaba} + \frac{\vec{\partial w}}{\partial W_+} \delta W_+ + \frac{\vec{\partial w}}{\partial W_-} \delta W_- \right) = \vec{0} \quad (18.11)$$

$$\begin{aligned} \left( \overline{\overline{\frac{\partial f}{\partial s}}} + \overline{\overline{\frac{\partial f}{\partial r} \frac{\partial h}{\partial x} \frac{\partial w}{\partial s}}} \right) \vec{\delta s} = & - \overline{\overline{\frac{\partial f}{\partial r} \frac{\partial h}{\partial x}}} \left( \frac{\vec{\partial w}}{\partial x_E} \delta x_E + \frac{\vec{\partial w}}{\partial x_I} \delta x_I + \frac{\vec{\partial w}}{\partial J_{nm da}} \delta J_{nm da} \right. \\ & \left. + \frac{\vec{\partial w}}{\partial J_{gaba}} \delta J_{gaba} + \frac{\vec{\partial w}}{\partial W_+} \delta W_+ + \frac{\vec{\partial w}}{\partial W_-} \delta W_- \right) \end{aligned} \quad (18.12)$$

Nous posons pour la suite

$$\overline{\overline{S}} = - \overline{\overline{\frac{\partial f}{\partial s}}}^{-1} \overline{\overline{\frac{\partial f}{\partial r} \frac{\partial h}{\partial x}}} \quad (18.13)$$

La variation infinitésimale  $\vec{\delta s}$  s'écrit alors :

$$\vec{\delta s} = \left( \overline{\overline{I}} - \overline{\overline{S}} \overline{\overline{\frac{\partial w}{\partial s}}} \right)^{-1} \overline{\overline{S}} \left( \frac{\vec{\partial w}}{\partial x_E} \delta x_E + \frac{\vec{\partial w}}{\partial x_I} \delta x_I + \frac{\vec{\partial w}}{\partial J_{nm da}} \delta J_{nm da} + \frac{\vec{\partial w}}{\partial J_{gaba}} \delta J_{gaba} + \frac{\vec{\partial w}}{\partial W_+} \delta W_+ + \frac{\vec{\partial w}}{\partial W_-} \delta W_- \right) \quad (18.14)$$

Or

$$\left\{ \begin{aligned} \frac{\vec{\partial w}}{\partial J_{gaba}} &= \begin{bmatrix} -sg^* \\ 0 \end{bmatrix} = -sg^* \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -sg^* \frac{\vec{\partial w}}{\partial x_E} \\ \frac{\vec{\partial w}}{\partial W_+} &= \begin{bmatrix} J_{nm da} sn^* \\ 0 \end{bmatrix} = J_{nm da} sn^* \frac{\vec{\partial w}}{\partial x_E} \\ \frac{\vec{\partial w}}{\partial W_-} &= \begin{bmatrix} 0 \\ -sg^* \end{bmatrix} = -sg^* \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -sg^* \frac{\vec{\partial w}}{\partial x_I} \\ \frac{\vec{\partial w}}{\partial J_{nm da}} &= \begin{bmatrix} W_+ sn^* \\ sn^* \end{bmatrix} = W_+ sn^* \begin{bmatrix} 1 \\ 0 \end{bmatrix} + sn^* \begin{bmatrix} 0 \\ 1 \end{bmatrix} = W_+ sn^* \frac{\vec{\partial w}}{\partial x_E} + sn^* \frac{\vec{\partial w}}{\partial x_I} \end{aligned} \right. \quad (18.15)$$

Nous en déduisons

$$\begin{aligned} \vec{\delta s} = \left( \overline{\overline{I}} - \overline{\overline{S}} \overline{\overline{\frac{\partial w}{\partial s}}} \right)^{-1} \overline{\overline{S}} \left( \frac{\vec{\partial w}}{\partial x_E} \delta x_E + \frac{\vec{\partial w}}{\partial x_I} \delta x_I + (W_+ sn^* \frac{\vec{\partial w}}{\partial x_E} + sn^* \frac{\vec{\partial w}}{\partial x_I}) \delta J_{nm da} \right. \\ \left. - sg^* \frac{\vec{\partial w}}{\partial x_E} \delta J_{gaba} + J_{nm da} sn^* \frac{\vec{\partial w}}{\partial x_E} \delta W_+ - sg^* \frac{\vec{\partial w}}{\partial x_I} \delta W_- \right) \end{aligned} \quad (18.16)$$

En réorganisant les termes :

$$\begin{aligned}\vec{\delta s} &= \left(\bar{\mathbb{I}} - \bar{S} \frac{\partial \bar{w}}{\partial s}\right)^{-1} \bar{S} \frac{\partial \bar{w}}{\partial x_E} (\delta x_E + W_+ sn^* \delta J_{nmda} + J_{nmda} sn^* \delta W_+ - sg^* \delta J_{gaba}) \\ &\quad + \left(\bar{\mathbb{I}} - \bar{S} \frac{\partial \bar{w}}{\partial s}\right)^{-1} \bar{S} \frac{\partial \bar{w}}{\partial x_I} (\delta x_I + sn^* \delta J_{nmda} - sg^* \delta W_-)\end{aligned}\quad (18.17)$$

Or dans l'article du chapitre 9, nous avons calculé :

$$\begin{cases} \left(\bar{\mathbb{I}} - \bar{S} \frac{\partial \bar{w}}{\partial s}\right)^{-1} \bar{S} \frac{\partial \bar{w}}{\partial x_E} = \frac{\vec{\delta s}}{\delta x_E} \equiv \begin{bmatrix} \mathcal{A}_{sn, x_E} \\ \mathcal{A}_{sg, x_E} \end{bmatrix} \\ \left(\bar{\mathbb{I}} - \bar{S} \frac{\partial \bar{w}}{\partial s}\right)^{-1} \bar{S} \frac{\partial \bar{w}}{\partial x_I} = \frac{\vec{\delta s}}{\delta x_I} \equiv \begin{bmatrix} \mathcal{A}_{sn, x_I} \\ \mathcal{A}_{sg, x_I} \end{bmatrix} \end{cases}\quad (18.18)$$

Nous pouvons en particulier en déduire l'expression de  $\delta sn^*$  :

$$\begin{aligned}\delta sn^* &= \mathcal{A}_{sn, x_E} (\delta x_E + W_+ sn^* \delta J_{nmda} + J_{nmda} sn^* \delta W_+ - sg^* \delta J_{gaba}) \\ &\quad + \mathcal{A}_{sn, x_I} (\delta x_I + sn^* \delta J_{nmda} - sg^* \delta W_-)\end{aligned}\quad (18.19)$$

En regroupant les contributions de  $J_{nmda}$  :

$$\boxed{\delta sn^* = \mathcal{A}_{sn, x_E} (\delta x_E + J_{nmda} sn^* \delta W_+ - sg^* \delta J_{gaba}) + \mathcal{A}_{sn, x_I} (\delta x_I - sg^* \delta W_-) + (\mathcal{A}_{sn, x_E} W_+ + \mathcal{A}_{sn, x_I}) sn^* \delta J_{nmda}}\quad (18.20)$$

### Etude du signe des contributions :

Nous perturbons les paramètres  $\pi_i$  un par un (i.e.  $\delta\pi_i > 0$  et  $\forall j \neq i, \delta\pi_j = 0$ ) pour connaître leur effet respectif sur le signe de  $\delta sn^*$ .

Nous avons vu que  $\mathcal{A}_{sn, x_E} > 0$  et  $\mathcal{A}_{sn, x_I} < 0$ . Tous les paramètres sont positifs,  $sn^* > 0$  et  $sg^* > 0$ . Nous pouvons alors construire le tableau suivant :

| Paramètre perturbé | $\delta sn^*$ | effet sur $sn^*$ dans une région |
|--------------------|---------------|----------------------------------|
| $x_E$              | +             | excitateur                       |
| $W_+$              | +             | excitateur                       |
| $J_{gaba}$         | -             | inhibiteur                       |
| $x_I$              | -             | inhibiteur                       |
| $W_-$              | +             | excitateur                       |
| $J_{nmda}$         | ?             | ?                                |

Cas d'une perturbation de  $J_{nmda}$  :

$$\delta sn^* = \underbrace{(\mathcal{A}_{sn, x_E} W_+ + \mathcal{A}_{sn, x_I})}_{>0} \underbrace{sn^*}_{>0} \delta J_{nmda}\quad (18.21)$$

Le signe de  $\delta sn^*$  dépend du signe de la somme  $\mathcal{A}_{sn,x_E} W_+ + \mathcal{A}_{sn,x_I}$ .

Remarque :

$J_{nmda}$  est un paramètre dont la perturbation produit à la fois une perturbation du couplage auto-excitateur et du couplage des excitateurs vers les inhibiteurs. Si nous perturbons uniquement le couplage des excitateurs vers les inhibiteurs nous aurions  $\delta sn^* = \mathcal{A}_{sn,x_I} sn^* < 0$ . La perturbation du couplage des excitateurs vers les inhibiteurs aurait alors un effet inhibiteur sur l'activité  $sn^*$  dans une région isolée.

## 18.1.2 Variation simultanée de deux paramètres

### 18.1.2.1 Construction analytique de la variation d'activité totale

Au point fixe, la valeur  $sn^*$  dépend des paramètres du modèle. On note  $sn^* = f(\pi_i, \pi_j)$ . Nous fixons les paramètres  $\pi_i$  et  $\pi_j$  et partons de la valeur de point fixe correspondante :  $sn^* = f(\widehat{\pi}_i, \widehat{\pi}_j)$ .

Nous faisons ensuite varier simultanément  $\pi_i$  entre  $\widehat{\pi}_i$  et  $\widehat{\pi}_i + \Delta\widehat{\pi}_i$  ; et  $\pi_j$  entre  $\widehat{\pi}_j$  et  $\widehat{\pi}_j + \Delta\widehat{\pi}_j$ .  $\Delta sn^*$  est la variation totale de  $sn^*$  due à la variation simultanée de ces deux paramètres. Par définition, nous avons :

$$\Delta sn^* \equiv f(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j) \quad (18.22)$$

Nous pouvons également construire  $\Delta sn^*$  en intégrant les sensibilités de  $sn^*$  par rapport à  $\pi_i$  et  $\pi_j$ , de deux manières différentes :

#### 18.1.2.1.1 Première procédure

1) Calcul de la variation de  $sn^*$  due à la variation de  $\pi_i$  entre  $\widehat{\pi}_i$  et  $\widehat{\pi}_i + \Delta\widehat{\pi}_i$ , évaluée en  $\widehat{\pi}_j$ . Nous la notons  $\Delta sn_{/\pi_i}(\widehat{\pi}_j)$ .

$$\Delta sn_{/\pi_i}(\widehat{\pi}_j) = \int_{\widehat{\pi}_i}^{\widehat{\pi}_i + \Delta\widehat{\pi}_i} S_{\pi_i}(\pi_i, \widehat{\pi}_j) d\pi_i \quad (18.23)$$

$$(18.24)$$

D'autre part, par définition  $\Delta sn_{/\pi_i}(\widehat{\pi}_j)$  est égal à la différence entre les points fixes dont on a fait varier  $\pi_i$ , évalués en  $\widehat{\pi}_j$  fixé :

$$\Delta sn_{/\pi_i}(\widehat{\pi}_j) \equiv f(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j) \quad (18.25)$$

2) Calcul de la variation de  $sn^*$  due à la variation de  $\pi_j$  entre  $\widehat{\pi}_j$  et  $\widehat{\pi}_j + \Delta\widehat{\pi}_j$ , évaluée en  $\widehat{\pi}_i + \Delta\widehat{\pi}_i$ .

$$\Delta sn_{/\pi_j}(\widehat{\pi}_i + \Delta\widehat{\pi}_i) = \int_{\widehat{\pi}_j}^{\widehat{\pi}_j + \Delta\widehat{\pi}_j} S_{\pi_j}(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \pi_j) d\pi_j \quad (18.26)$$

$$(18.27)$$

Or par définition :

$$\Delta sn_{/\pi_j}(\widehat{\pi}_i + \Delta\widehat{\pi}_i) \equiv f(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j) \quad (18.28)$$



3) Nous en déduisons :

$$\Delta sn \equiv f(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j) \quad (18.29)$$

$$= \Delta sn_{/\pi_i}(\widehat{\pi}_j) + \Delta sn_{/\pi_j}(\widehat{\pi}_i + \Delta\widehat{\pi}_i) \quad (18.30)$$

#### 18.1.2.1.2 Deuxième procédure :

1) Calcul de la variation de  $sn^*$  due à la variation de  $\pi_j$  entre  $\widehat{\pi}_j$  et  $\widehat{\pi}_j + \Delta\widehat{\pi}_j$ , évaluée en  $\widehat{\pi}_i$ .

$$\Delta sn_{/\pi_j}(\widehat{\pi}_i) = \int_{\widehat{\pi}_j}^{\widehat{\pi}_j + \Delta\widehat{\pi}_j} S_{\pi_j}(\widehat{\pi}_i, \pi_j) d\pi_j \quad (18.31)$$

$$(18.32)$$

D'autre part, par définition  $\Delta sn_{/\pi_j}(\widehat{\pi}_i)$  est égal à la différence entre les points fixes évalués en  $\widehat{\pi}_i$  fixé, et dont on a fait varier  $\pi_j$  :

$$\Delta sn_{/\pi_j}(\widehat{\pi}_i) \equiv f(\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j) \quad (18.33)$$

2) Calcul de la variation de  $sn^*$  due à la variation de  $\pi_i$  entre  $\widehat{\pi}_i$  et  $\widehat{\pi}_i + \Delta\widehat{\pi}_i$ , évaluée en  $\widehat{\pi}_j + \Delta\widehat{\pi}_j$ . Nous la notons  $\Delta sn_{/\pi_i}(\widehat{\pi}_j + \Delta\widehat{\pi}_j)$ .

$$\Delta sn_{/\pi_i}(\widehat{\pi}_j + \Delta\widehat{\pi}_j) = \int_{\widehat{\pi}_i}^{\widehat{\pi}_i + \Delta\widehat{\pi}_i} S_{\pi_i}(\pi_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) d\pi_i \quad (18.34)$$

$$(18.35)$$

Par définition :

$$\Delta sn_{/\pi_i}(\widehat{\pi}_j + \Delta\widehat{\pi}_j) \equiv f(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) \quad (18.36)$$

3) Nous en déduisons également :

$$\Delta sn \equiv f(\widehat{\pi}_i + \Delta\widehat{\pi}_i, \widehat{\pi}_j + \Delta\widehat{\pi}_j) - f(\widehat{\pi}_i, \widehat{\pi}_j) \quad (18.37)$$

$$= \Delta sn_{/\pi_j}(\widehat{\pi}_i) + \Delta sn_{/\pi_i}(\widehat{\pi}_j + \Delta\widehat{\pi}_j) \quad (18.38)$$

#### 18.1.2.1.3 Résumé

Nous partons du point fixe  $sn^* = f(\widehat{\pi}_i, \widehat{\pi}_j)$ . La variation totale  $\Delta sn^*$  due à la variation simultanée de  $\pi_i$  et  $\pi_j$ , respectivement entre  $[\widehat{\pi}_i; \widehat{\pi}_i + \Delta\widehat{\pi}_i]$  et  $[\widehat{\pi}_j; \widehat{\pi}_j + \Delta\widehat{\pi}_j]$ , se décompose en deux étapes successives correspondant à la variation successive de chaque paramètre. Elle peut s'écrire de deux manières<sup>1</sup> :

$$\begin{cases} \Delta sn = \Delta sn_{\pi_i}(\widehat{\pi}_j) + \Delta sn_{\pi_j}(\widehat{\pi}_i + \Delta\widehat{\pi}_i) \\ \quad = \Delta sn_{\pi_j}(\widehat{\pi}_i) + \Delta sn_{\pi_i}(\widehat{\pi}_j + \Delta\widehat{\pi}_j) \end{cases} \quad (18.39)$$

---

1. On retrouve la formule des accroissements finis au premier ordre d'une fonction de deux variables.

### 18.1.2.2 Limites de la balance E/I pour penser l'effet de la variation de plusieurs paramètres

La variation totale de l'activité excitatrice  $sn^*$  de la région s'écrit :

$$\Delta sn = \Delta sn_{/\pi_E}(\widehat{\pi}_I) + \Delta sn_{/\pi_I}(\widehat{\pi}_E + \Delta\widehat{\pi}_E) \quad (18.40)$$

$$= \Delta sn_{/\pi_I}(\widehat{\pi}_E) + \Delta sn_{/\pi_E}(\widehat{\pi}_I + \Delta\widehat{\pi}_I) \quad (18.41)$$

Pour que  $\Delta sn = 0$ , il faut vérifier la première condition :

$$\Delta sn_{/\pi_E}(\widehat{\pi}_I) = -\Delta sn_{/\pi_I}(\widehat{\pi}_E + \Delta\widehat{\pi}_E) \quad (18.42)$$

$$\int_{\widehat{\pi}_E}^{\widehat{\pi}_E + \Delta\widehat{\pi}_E} S_{\pi_E}(\pi_E, \widehat{\pi}_I) d\pi_E = - \int_{\widehat{\pi}_I}^{\widehat{\pi}_I + \Delta\widehat{\pi}_I} S_{\pi_I}(\widehat{\pi}_E + \Delta\widehat{\pi}_E, \pi_I) d\pi_I \quad (18.43)$$

$$\text{en posant } \Delta\widehat{\pi}_I \equiv \Delta\widehat{\pi}_E : \int_{\widehat{\pi}_E}^{\widehat{\pi}_E + \Delta\widehat{\pi}_E} S_{\pi_E}(\pi_E, \widehat{\pi}_I) d\pi_E = - \int_{\widehat{\pi}_I}^{\widehat{\pi}_I + \Delta\widehat{\pi}_E} S_{\pi_I}(\widehat{\pi}_E + \Delta\widehat{\pi}_E, \pi_I) d\pi_I \quad (18.44)$$

Ainsi que la deuxième :

$$\Delta sn_{/\pi_I}(\widehat{\pi}_E) = -\Delta sn_{/\pi_E}(\widehat{\pi}_I + \Delta\widehat{\pi}_I) \quad (18.45)$$

$$\int_{\widehat{\pi}_I}^{\widehat{\pi}_I + \Delta\widehat{\pi}_I} S_{\pi_I}(\widehat{\pi}_E, \pi_I) d\pi_I = - \int_{\widehat{\pi}_E}^{\widehat{\pi}_E + \Delta\widehat{\pi}_E} S_{\pi_E}(\pi_E, \widehat{\pi}_I + \Delta\widehat{\pi}_I) d\pi_E \quad (18.46)$$

$$\text{en posant } \Delta\widehat{\pi}_I \equiv \Delta\widehat{\pi}_E : \int_{\widehat{\pi}_I}^{\widehat{\pi}_I + \Delta\widehat{\pi}_E} S_{\pi_I}(\widehat{\pi}_E, \pi_I) d\pi_I = - \int_{\widehat{\pi}_E}^{\widehat{\pi}_E + \Delta\widehat{\pi}_E} S_{\pi_E}(\pi_E, \widehat{\pi}_I + \Delta\widehat{\pi}_I) d\pi_E \quad (18.47)$$

Ces deux conditions sont vérifiées dans le cas où les sensibilités sont des constantes sur tout l'intervalle d'intégration ; et de plus identiques en valeurs absolues mais de signe opposé.

$$\forall \pi_E \in [\widehat{\pi}_E; \widehat{\pi}_E + \Delta\widehat{\pi}_E], \forall \pi_I \in [\widehat{\pi}_I; \widehat{\pi}_I + \Delta\widehat{\pi}_I] \quad \begin{cases} S_{\pi_E}(\pi_E, \pi_I) = \mathcal{A} \\ S_{\pi_I}(\pi_E, \pi_I) = -\mathcal{A} \end{cases} \quad (18.48)$$

Pour la première condition, on a d'une part :

$$\int_{\widehat{\pi}_E}^{\widehat{\pi}_E + \Delta\widehat{\pi}_E} S_{\pi_E}(\pi_E, \widehat{\pi}_I) d\pi_E = \mathcal{A} \int_{\widehat{\pi}_E}^{\widehat{\pi}_E + \Delta\widehat{\pi}_E} d\pi_E = \mathcal{A} \Delta\widehat{\pi}_E \quad (18.49)$$

Et d'autre part :

$$- \int_{\widehat{\pi}_I}^{\widehat{\pi}_I + \Delta\widehat{\pi}_E} S_{\pi_I}(\widehat{\pi}_E + \Delta\widehat{\pi}_E, \pi_I) d\pi_I = -(-\mathcal{A}) \int_{\widehat{\pi}_I}^{\widehat{\pi}_I + \Delta\widehat{\pi}_E} d\pi_I = \mathcal{A} \Delta\widehat{\pi}_E \quad (18.50)$$

La première condition est donc bien vérifiée. La deuxième condition se vérifie de la même manière.

## 18.2 Deux régions : Effet d'une augmentation du forçage dans la région de contrôle

### 18.2.1 Expressions analytiques des sensibilités par rapport à une perturbation du forçage de la région de contrôle

#### 18.2.1.1 Avec rétroaction

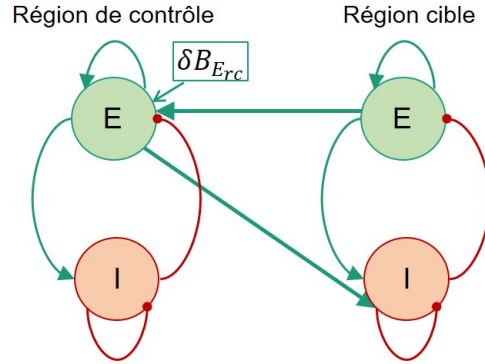


**Figure 18.1** – Schéma d'une connectivité E-I entre une région de contrôle à gauche et une région cible à droite. Les flèches vertes (resp. rouges) représentent des coefficients de couplages positifs (resp. négatifs). Nous avons représenté la perturbation du forçage  $B_{Erc}$  sur le pool excitateur de la région de contrôle.

Pour une perturbation du forçage  $B_{Erc}$  dans la région de contrôle, nous avons écrit  $\delta sn_{rc}/B_{Erc}$  et  $\delta sn_c/B_{Erc}$  dans le système couplé en EI :

$$\begin{cases} g_{E-I} = \mathcal{A}_{sn_{rc}, x_{Erc}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{Ic}} \kappa_{c-rc} \\ \delta sn_{rc}/B_{Erc} = \frac{\mathcal{A}_{sn_{rc}, x_{Erc}}}{1 - g_{E-I}} \delta B_{Erc} \\ \delta sn_c/B_{Erc} = \frac{\mathcal{A}_{sn_c, x_{Ic}} \kappa_{c-rc} \mathcal{A}_{sn_{rc}, x_{Erc}}}{1 - g_{E-I}} \delta B_{Erc} = \mathcal{A}_{sn_c, x_{Ic}} \kappa_{c-rc} \delta sn_{rc} \end{cases} \quad (18.51)$$

Dans l'article du chapitre 9, nous avons tracé  $\delta sn_{rc}/B_{Erc}$  et  $\delta sn_c/B_{Erc}$  ainsi que le gain pour différentes valeurs de  $B_{Erc}$ , en particulier pour une connectivité EI.

L'intégration de ces perturbations entre  $\widehat{B_{Erc}} = 0.382$  et  $\widehat{B_{Erc}} + I_{stim}$  nous donne  $\Delta sn_{rc}/B_{Erc}$  et  $\Delta sn_c/B_{Erc}$ .

$\Delta sn_i/B_{Erc}$  peut par ailleurs être trouvée numériquement par des simulations. Nous pouvons faire tourner la dynamique pour trouver  $sn_{stim,i}^*$  en période de stimulation et  $sn_{B,i}^*$  au repos, puis nous en faisons la différence ( $\Delta sn_i/B_{Erc} = sn_{stim,i}^* - sn_{B,i}^*$ ). L'intégration des sensibilités dont on connaît le signe permet de savoir prédire le signe des variations  $\Delta sn_i/B_{Erc}$ , ainsi que leur sens de variation quand la stimulation  $I_{stim}$  augmente (cf sections 18.2.2.1.1 et 18.2.2.2.1).

### 18.2.1.2 Sans la rétroaction de la cible vers la région de contrôle

Nous pouvons étudier ce que donne sur  $sn_{rc}^*$  et  $sn_c^*$  la même perturbation du pool excitateur de la région de contrôle quand la rétroaction de la cible vers la région de contrôle n'existe pas.

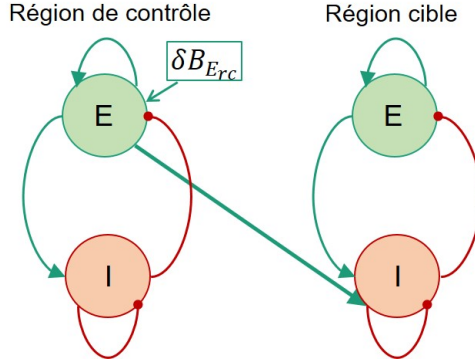


**Figure 18.2** – Schéma d'un couplage entre une région de contrôle à gauche et une région cible à droite, sans rétroaction de la cible vers la région de contrôle.

Si nous coupons la rétroaction de la cible vers la région de contrôle, les calculs de sensibilités à  $B_{Erc}$  menés au chapitre 9 avec la rétroaction restent valables, à condition de :

1. fixer le coefficient de couplage  $\kappa_{rc-c} = 0$ ,
2. prendre les valeurs de points fixes dans le système sans la rétroaction

Nous reprenons alors le système 18.51 et nous pouvons écrire un nouveau système découplé, qui donne  $\delta sn_{rc/B_{Erc}}^{\text{dec}}$  dans la région de contrôle et  $\delta sn_{c/B_{Erc}}^{\text{dec}}$  dans la cible pour une perturbation de  $B_{Erc}$  :

$$\begin{cases} g_{E-I} = 0 \\ \delta sn_{rc/B_{Erc}}^{\text{dec}} = \mathcal{A}_{sn_{rc}, x_{Erc}} \delta B_{Erc} \\ \delta sn_{c/B_{Erc}}^{\text{dec}} = \mathcal{A}_{sn_c, x_{Ic}} \kappa_{c-rc} \mathcal{A}_{sn_{rc}, x_{Erc}} \delta B_{Erc} = \mathcal{A}_{sn_c, x_{Ic}} \kappa_{c-rc} \delta sn_{rc/B_{Erc}}^{\text{dec}} \end{cases} \quad (18.52)$$

Ces expressions doivent être évaluées avec les valeurs de points fixes dans le système sans la rétroaction.

Il est intéressant de pouvoir lire l'expression de  $\delta sn_{c/B_{Erc}}^{\text{dec}}$  comme le produit entre un terme d'**émission** ( $\delta sn_{rc/B_{Erc}}^{\text{dec}}$ ), de **transmission** ( $\kappa_{c-rc}$ ) et de **réception** ( $\mathcal{A}_{sn_c, x_{Ic}}$ ). De manière causale, la perturbation de  $B_{Erc}$  a un effet sur la région de contrôle puis cet effet est transmis à la cible (que l'on peut comprendre à partir du schéma de la figure 18.2). L'expression traduit en perturbation la chaîne de causalité décrite dans l'équation (13.24) et qui traduit une représentation immédiate du contrôle inhibiteur.

Les sensibilités dans ce système découplé sont égales à  $\frac{\delta sn_{rc/B_{Erc}}^{\text{dec}}}{\delta B_{Erc}}$  et  $\frac{\delta sn_{c/B_{Erc}}^{\text{dec}}}{\delta B_{Erc}}$ .  $\Delta sn_{rc/B_{Erc}}^{\text{dec}}$  et  $\Delta sn_{c/B_{Erc}}^{\text{dec}}$  sont les variations d'activité obtenues pour une variation du forçage

$B_{Erc}$  entre  $\widehat{B_{Erc}} = 0.382$  et  $\widehat{B_{Erc}} + I_{stim}$ . Elles peuvent aussi être calculées en intégrant respectivement  $\delta sn_{rc/B_{Erc}}^{dec}$  et  $\delta sn_{c/B_{Erc}}^{dec}$  entre  $\widehat{B_{Erc}} = 0.382$  et  $\widehat{B_{Erc}} + I_{stim}$ .

$\Delta sn_{i/B_{Erc}}^{dec}$  peut par ailleurs être trouvée numériquement par des simulations, en faisant tourner la dynamique pour trouver  $sn_{stim,i}^{dec}$  en période de stimulation et  $sn_{B,i}^{dec}$  au repos, puis en faisant la différence :  $\Delta sn_{i/B_{Erc}}^{dec} = sn_{stim,i}^{dec} - sn_{B,i}^{dec}$ . L'intégration des sensibilités dont on connaît le signe permet de savoir prédire le signe des variations  $\Delta sn_{i/B_{Erc}}^{dec}$ , ainsi que leur sens de variation quand la stimulation  $I_{stim}$  augmente (cf sections 18.2.2.1.2 et 18.2.2.2.2).

## 18.2.2 Variation du forçage sur le pool excitateur de la région de contrôle de $\Delta B_{Erc} = I_{stim}$

Pour une valeur de  $J_{gaba_c}$  fixé, nous calculons pour la région de contrôle et la région cible l'activité  $sn_i^*$ , la sensibilité de  $sn_i^*$  par rapport à  $B_{Erc}$  ainsi que la variation d'activité  $\Delta sn_{i/B_{Erc}}$  en fonction du forçage. Nous présentons d'abord les résultats pour la région cible, dans le système couplé en E-I puis pour le système découplé sans la rétroaction de la cible vers la région de contrôle. Nous passerons ensuite à la région de contrôle, en E-I puis découplée. La rétroaction ne change pas le sens des variations de  $sn_i^*$  quand  $I_{stim}$  augmente, mais change les amplitudes.

### 18.2.2.1 Effet sur la région cible

#### 18.2.2.1.1 Cas de la connectivité EI

Sur la figure 18.3 A, nous avons représenté l'activité  $sn_c^*$  de la cible, en fonction de  $I_{stim}$  (c'est-à-dire pour une valeur de forçage sur le pool excitateur de la région de contrôle égal à  $B_{Erc} = 0.382 + I_{stim}$ ), et pour une valeur de  $J_{gaba_c}$  fixée dans la région cible. Le couple  $(J_{gaba_{rc}}, J_{gaba_c})$  assure que  $rn_{rc} = rn_c = 3Hz$ . Il est approximativement égal à (1.758, 0.929) dans le cas couplé. Pour calculer la variation, nous pouvons faire une différence entre  $sn_c^*$  stimulé et non stimulé (cf flèche rouge sur le graphe A, reportée sur le graphe B).

Sur le graphe B, la variation d'activité de la cible dépend de manière non linéaire de  $I_{stim}$ . Dans le cas de cette architecture de connectivité EI, elle décroît quand  $I_{stim}$  augmente.

Sur le graphe C, la sensibilité  $S_{sn_c/B_{Erc}}(B_{Erc}, \widehat{J_{gaba_c}})$ , évaluée pour une valeur fixée du forçage  $B_{Erc} = \widehat{B_{Erc}} + I_{stim} = 0.382 + I_{stim}$  et de  $J_{gaba_c} = \widehat{J_{gaba_c}} \approx 0.929$  (obtenue avec les expressions analytiques section 18.2.1.1), est aussi égale à la pente de la tangente à la courbe  $sn_c^* = f_c(I_{stim}, J_{gaba_c})$  sur le graphe A. Nous vérifions bien que les pentes sont négatives et varient en fonction de  $I_{stim}$ . L'intégration de toutes les sensibilités entre 0 et  $I_{stim}$ , égale à l'opposé de l'aire hachurée, permet de calculer par une deuxième méthode la variation  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$ . Elle peut être traduite par une somme de  $\delta sn_c^*$  (cf petites doubles flèches rouges sur le graphe A).

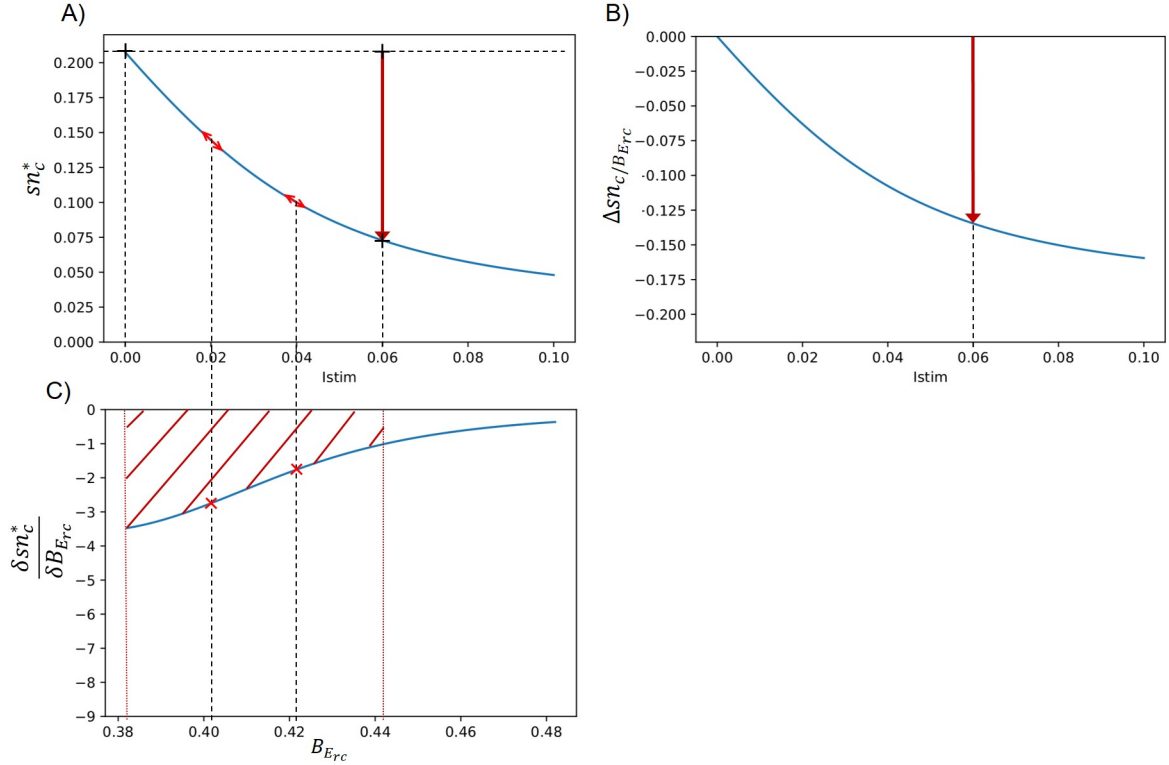


**Figure 18.3** – Etude du comportement d'une cible en fonction du forçage appliqué sur le pool excitateur de la région de contrôle dans le cas d'une connectivité E-I. A : Activité  $sn_c^*$  de la cible en fonction de  $I_{stim}$  sur le pool excitateur de la région de contrôle, pour  $J_{gaba_c}$  fixé. Le forçage sur le pool excitateur de la région de contrôle vaut  $B_{Erc} = 0.382 + I_{stim}$ . B : Variation d'activité de la cible calculée entre la stimulation ( $I_{stim}$ ) et le repos ( $I_{stim} = 0$  i.e.  $B_{Erc} = 0.382$ ), pour  $J_{gaba_c}$  fixé. C : Sensibilité de  $sn_c^*$  par rapport au forçage  $B_{Erc}$ , en fonction de  $B_{Erc}$ , pour  $J_{gaba_c}$  fixé.

#### 18.2.2.1.2 Sans la rétroaction de la cible vers la région de contrôle

Sur la figure 18.4 A, nous avons représenté l'activité  $sn_c^{*dec}$  et la variation d'activité de la cible  $\Delta sn_c^{dec}/B_{Erc}(J_{gaba_c})$ , en fonction de  $I_{stim}$ , pour une valeur de  $J_{gaba_c}$  fixée dans la région cible. Le couple  $(J_{gabar_c}, J_{gaba_c})$  assure que  $rn_{rc} = rn_c = 3Hz$ . Il est approximativement égal à (1.222, 0.929) dans le cas découplé.

Comme précédemment, la différence entre  $sn_c^{*dec}$  stimulé et non stimulé permet de calculer la variation d'activité due à  $I_{stim}$  (cf flèche rouge sur A, reportée sur B).

Sur A, la pente de la tangente à la courbe  $sn_c^{*dec} = f_c^{dec}(I_{stim}, J_{gaba_c})$  est égale à la sensibilité de  $sn_c^{*dec}$  par rapport au forçage  $B_{Erc}$ , évaluée en  $B_{Erc} = 0.382 + I_{stim}$  et  $J_{gaba_c} = \widehat{J_{gaba_c}} \approx 0.929$ . La courbe de sensibilité du graphe C a été obtenue à partir des calculs de la section 18.2.1.2. L'absence de rétroaction modifie les amplitudes par rapport au cas couplé.

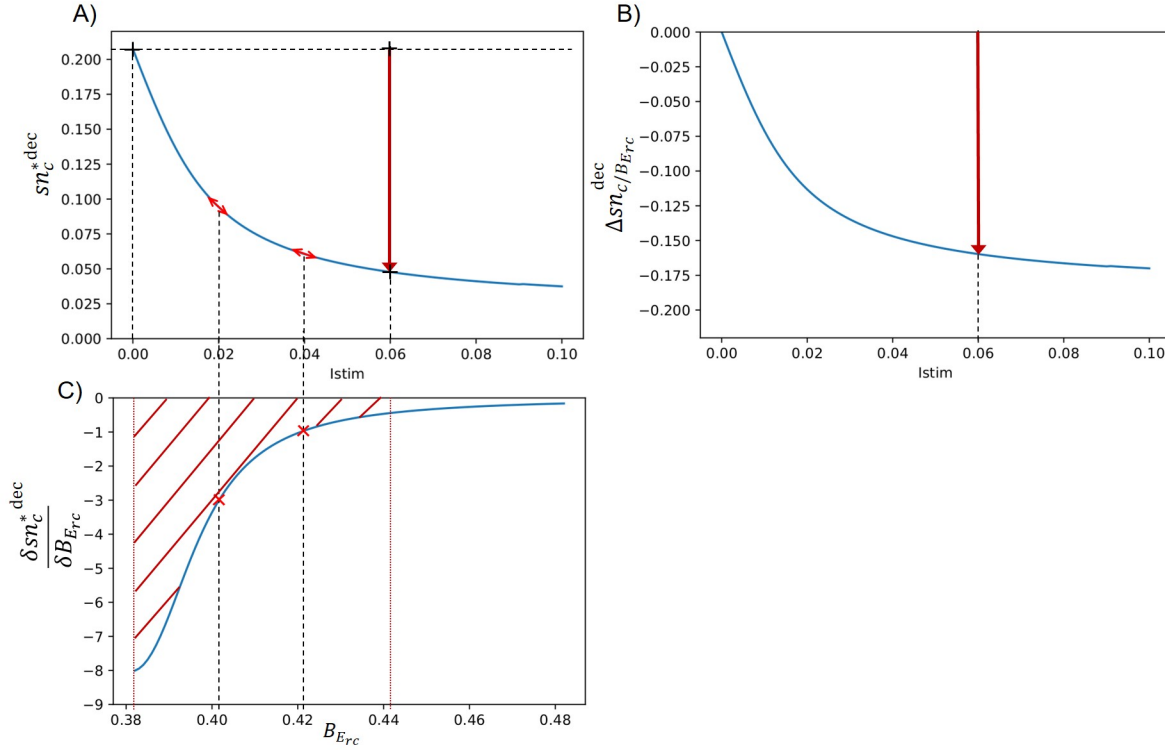


**Figure 18.4** – Etude du comportement d'une cible en fonction du forçage appliqué sur le pool excitateur de la région de contrôle dans le cas découplé (sans rétroaction de la cible vers la région de contrôle). La légende est identique à la figure 18.3.

#### 18.2.2.2 Effet sur la région de contrôle

Comme pour la région cible, nous présentons aux figures 18.5 et 18.6 l'activité, la variation et la sensibilité de l'activité de la région de contrôle dans les cas avec et sans rétroaction. Les explications détaillées permettant de passer entre les graphes dans chacune de ces deux figures sont identiques à celles des sections 18.2.2.1.1 et 18.2.2.1.2. Les pentes des courbes sur les graphes A sont positives et nous retrouvons bien une sensibilité positive sur les graphes B correspondant. L'intégrale de la sensibilité est directement égale à l'aire hachurée. La variation d'activité de la région de contrôle augmente quand  $I_{stim}$  augmente, avec des différences d'amplitude entre les cas EI et découplé.

### 18.2.2.2.1 Cas de la connectivité EI

Pour rappel, le couple  $(J_{gaba_{rc}}, J_{gaba_c})$  est approximativement égal à  $(1.758, 0.929)$  dans le cas couplé. Il assure que  $rn_{rc} = rn_c = 3Hz$ .

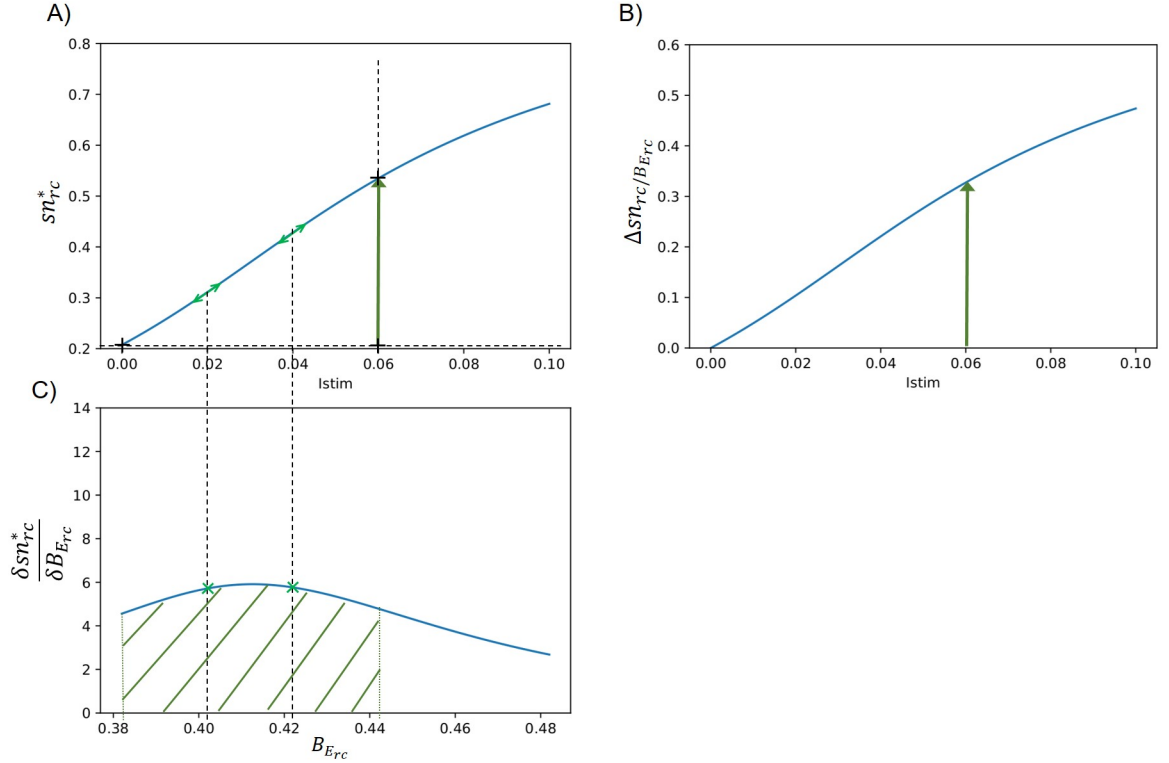


**Figure 18.5** – Etude du comportement d'une région de contrôle en fonction du forçage appliqué sur le pool excitateur de la région de contrôle dans le cas d'une connectivité E-I. A : Activité  $sn_{rc}^*$  de la région de contrôle en fonction de  $I_{stim}$  sur le pool excitateur de la région de contrôle, pour  $J_{gaba_c}$  fixé. Le forçage sur le pool excitateur de la région de contrôle vaut  $B_{erc} = 0.382 + I_{stim}$ . B : Variation d'activité de la région de contrôle calculée entre la stimulation ( $I_{stim}$ ) et le repos ( $I_{stim} = 0$  i.e.  $B_{erc} = 0.382$ ), pour  $J_{gaba_c}$  fixé. C : Sensibilité de  $sn_{rc}^*$  par rapport au forçage  $B_{erc}$ , en fonction de  $B_{erc}$ , pour  $J_{gaba_c}$  fixé.



### 18.2.2.2.2 Sans la rétroaction de la cible vers la région de contrôle

Le couple  $(J_{gaba_{rc}}, J_{gaba_c})$  assure que  $rn_{rc} = rn_c = 3Hz$ . Il est approximativement égal à  $(1.222, 0.929)$  dans le cas découplé.

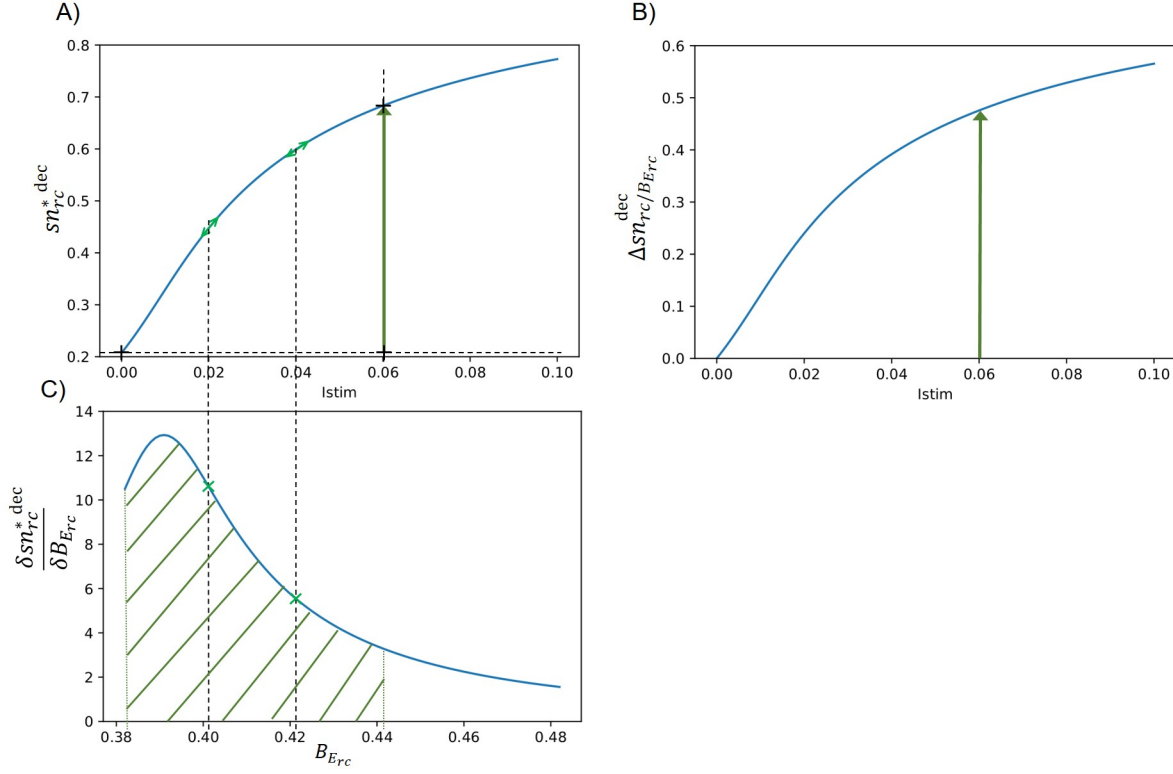


**Figure 18.6** – Etude du comportement d’une région de contrôle en fonction du forçage appliqué sur le pool excitateur de la région de contrôle dans le cas découplé (sans rétroaction de la cible vers la région de contrôle). La légende est identique à la figure 18.5.

### 18.2.2.3 Hypothèse de linéarité de la variation d'activité de la cible

Une hypothèse immédiate de l'effet d'une augmentation de la stimulation de la région de contrôle sur la baisse d'activité de la cible peut être formulée comme suit :

*"Pour multiplier par  $\alpha$  la baisse d'activité de la cible, nous avons besoin de multiplier par  $\alpha$  la stimulation de la région de contrôle."*

Supposons que  $J_{gaba_c}$  soit fixé et égal à  $\widehat{J_{gaba_c}}$ .  $I_{stim}$  cause une variation d'activité dans la cible égale à  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$  dans le cas d'une connectivité EI ( $\Delta sn_{c/B_{Erc}}^{dec}(\widehat{J_{gaba_c}})$  dans le cas sans rétroaction de la cible vers la région de contrôle). L'hypothèse immédiate peut se reformuler :

*"Si  $I_{stim}$  dans la région de contrôle est multiplié par  $\alpha$ , l'activité de la cible est aussi multipliée par  $\alpha$  et varie alors de  $\alpha \Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$  (ou  $\alpha \Delta sn_{c/B_{Erc}}^{dec}(\widehat{J_{gaba_c}})$  dans le cas sans rétroaction de la cible vers la région de contrôle)".*

Nous montrons qu'avec ou sans rétroaction, cette hypothèse est fausse.

#### 18.2.2.3.1 Conditions de vérification de la linéarité

La variation d'activité de la cible due à une variation du forçage  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + I_{stim}$  peut s'écrire comme l'intégrale de la sensibilité de  $sn_c^*$  par rapport à  $B_{Erc}$ . Dans le cas d'une connectivité EI :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + I_{stim}} S_{sn_{c/B_{Erc}}}(B_{Erc}, \widehat{J_{gaba_c}}) dB_{Erc} \quad (18.53)$$

Dans le cas sans rétroaction, nous intégrons la sensibilité découplée  $S_{sn_{c/B_{Erc}}}^{dec}(B_{Erc}, \widehat{J_{gaba_c}})$ , pour obtenir  $\Delta sn_{c/B_{Erc}}^{dec}(\widehat{J_{gaba_c}})$ . Nous poursuivons notre présentation pour le cas EI.

Pour que l'hypothèse de linéarité soit vraie, il faudrait que la sensibilité de  $sn_c^*$  par rapport à  $B_{Erc}$  soit constante sur tout l'intervalle d'intégration.

En effet, si  $\forall B_{Erc} S_{sn_{c/B_{Erc}}}(B_{Erc}, \widehat{J_{gaba_c}}) = \mathcal{A}(\widehat{J_{gaba_c}})$ , alors :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + I_{stim}} S_{sn_{c/B_{Erc}}}(B_{Erc}, \widehat{J_{gaba_c}}) dB_{Erc} \quad (18.54)$$

$$= \mathcal{A}(\widehat{J_{gaba_c}}) I_{stim} \quad (18.55)$$

Nous pourrions ainsi en déduire que :

$$\forall \alpha \in \mathcal{R}^+, \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \alpha I_{stim}} S_{sn_{c/B_{Erc}}}(B_{Erc}, \widehat{J_{gaba_c}}) dB_{Erc} = \mathcal{A}(\widehat{J_{gaba_c}}) \alpha I_{stim} \quad (18.56)$$

$$= \alpha \Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) \quad (18.57)$$

#### 18.2.2.3.2 Condition non validée

Or, nous voyons sur la figure 18.3 C) que pour une valeur fixée de  $J_{gaba_c}$ , la sensibilité de  $sn_c^*$  par rapport à  $B_{Erc}$  varie en fonction de  $B_{Erc}$ . La condition précédente n'est donc pas validée. Il en est de même pour le cas sans rétroaction (cf figure 18.4).

Dans les deux cas, nous pouvons observer que la variation d'activité de la cible ne dépend pas linéairement de  $I_{stim}$  (graphes B).

## 18.3 Deux régions couplées : perturbation des $J_{gaba}$

### 18.3.1 Formulation en $x_{inter}$ et $x_{intra}$

Au régime stationnaire, le modèle s'écrit :

$$\begin{cases} 0 = -\beta^E sn_i^* + \alpha^E T_{glu}(1 - sn_i^*)rn_i^* & \equiv fn_i(sn_i^*, rn_i^*) \\ 0 = -\beta^I sg_i^* + \alpha^I T_{gaba}(1 - sg_i^*)rg_i^* & \equiv fg_i(sn_i^*, rn_i^*) \end{cases} \quad (18.58)$$

avec

$$\begin{cases} rn_i^* = \frac{a_E xn_i^* - b_E}{1 - e^{-d_E(a_E xn_i^* - b_E)}} & \equiv hn_i(xn_i^*) \\ rg_i^* = \frac{a_I xg_i^* - b_I}{1 - e^{-d_I(a_I xg_i^* - b_I)}} & \equiv hg_i(xg_i^*) \end{cases} \quad (18.59)$$

où  $i \in \{1, 2\}$ .

Dans l'équation (18.59),  $xn_i^*$  et  $xg_i^*$  représentent le courant total en entrée des pools excitateur et inhibiteur de la région  $i$ .

Nous voulons exprimer les sensibilités par rapport à  $\overrightarrow{J_{gaba}}$  dans le système couplé en fonction des sensibilités des régions découplées par rapport aux forçages sur les pools excitateurs. Nous séparons ces courants totaux en deux composantes.

Le courant interne ou intra-région sera noté  $\overrightarrow{x_{intra}}$  et contient tous les couplages internes à une région (i.e. couplages entre pools et les auto-couplages excitateur ou inhibiteur).

$$\begin{cases} xn_{intra,i}^* = W_+ J_{nmda} sn_i^* - J_{gaba,i} sg_i^* & \equiv wn_{intra,i}(sn_i^*, sg_i^*) \\ xg_{intra,i}^* = J_{nmda} sn_i^* - W_- sg_i^* & \equiv wg_{intra,i}(sn_i^*, sg_i^*) \end{cases} \quad (18.60)$$

Le courant externe, noté  $\overrightarrow{x_{inter}}$  contient les couplages entre les régions et les forçages externes :

$$\begin{cases} xn_{inter,i}^* = k_{E_{ij}} \kappa_{ij} sn_j^* + B_{E_i} & \equiv wn_{inter,i}(sn_j^*) \quad j \neq i \\ xg_{inter,i}^* = (1 - k_{E_{ij}}) \kappa_{ij} sn_j^* + B_{I_i} & \equiv wg_{inter,i}(sn_j^*) \quad j \neq i \end{cases} \quad (18.61)$$

Le courant total s'écrit :

$$\begin{cases} xn_i^* = xn_{intra,i}^* + xn_{inter,i}^* = wn_i(xn_{intra,i}^*, xn_{inter,i}^*) \\ xg_i^* = xg_{intra,i}^* + xg_{inter,i}^* = wg_i(xg_{intra,i}^*, xg_{inter,i}^*) \end{cases} \quad (18.62)$$

### 18.3.2 Elimination des variables intermédiaires

Nous définissons :

$$\vec{s} = \begin{pmatrix} sn_1 \\ sn_2 \\ sg_1 \\ sg_2 \end{pmatrix}, \quad \vec{r} = \begin{pmatrix} rn_1 \\ rn_2 \\ rg_1 \\ rg_2 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} xn_1 \\ xn_2 \\ xg_1 \\ xg_2 \end{pmatrix}, \quad \overrightarrow{x_{intra}} = \begin{pmatrix} xn_{intra,1} \\ xn_{intra,2} \\ xg_{intra,1} \\ xg_{intra,2} \end{pmatrix}, \quad \overrightarrow{x_{inter}} = \begin{pmatrix} xn_{inter,1} \\ xn_{inter,2} \\ xg_{inter,1} \\ xg_{inter,2} \end{pmatrix} \quad (18.63)$$

ainsi que

$$\overrightarrow{J_{gaba}} = \begin{pmatrix} J_{gaba_1} \\ J_{gaba_2} \end{pmatrix}, \quad (18.64)$$

Le modèle au régime stationnaire se traduit par le système :

$$\begin{cases} \vec{f}(\vec{s}, \vec{r}) = \vec{0} \\ \vec{r} = \vec{h}(\vec{x}) \\ \vec{x} = \vec{w}(\overrightarrow{x_{intra}}, \overrightarrow{x_{inter}}) \\ \overrightarrow{x_{intra}} = \overrightarrow{w_{intra}}(\vec{s}, \overrightarrow{J_{gaba}}) \\ \overrightarrow{x_{inter}} = \overrightarrow{w_{inter}}(\vec{s}) \end{cases} \quad (18.65)$$

Nous éliminons les variables intermédiaires  $\vec{r}$ ,  $\vec{x}$  et  $\overrightarrow{x_{intra}}$  dans la première équation. L'équation 18.65 peut alors se réécrire sous la forme :

$$\begin{cases} \vec{F}(\vec{s}, \overrightarrow{x_{inter}}, \overrightarrow{J_{gaba}}) = \vec{0} \\ \overrightarrow{x_{inter}} = \overrightarrow{w_{inter}}(\vec{s}) \end{cases} \quad (18.66)$$

où

$$\begin{cases} Fn_i(sn_i^*, sg_i^*, xn_{inter,i}^*, J_{gaba_i}) = -\beta^E sn_i^* \\ \quad + \alpha^E T_{glu}(1 - sn_i^*) \frac{a_E(W_+ J_{nmda} sn_i^* - J_{gaba_i} sg_i^* + xn_{inter,i}^*) - b_E}{1 - e^{-d_E[a_E(W_+ J_{nmda} sn_i^* - J_{gaba_i} sg_i^* + xn_{inter,i}^*) - b_E]}} \\ Fg_i(sn_i^*, sg_i^*, xg_{inter,i}^*) = -\beta^I sg_i^* \\ \quad + \alpha^I T_{gaba}(1 - sg_i^*) \frac{a_I(J_{nmda} sn_i^* - W_- sg_i^* + xg_{inter,i}^*) - b_I}{1 - e^{-d_I[a_I(J_{nmda} sn_i^* - W_- sg_i^* + xg_{inter,i}^*) - b_I]}} \end{cases} \quad (18.67)$$

$\overrightarrow{x_{inter}}$ , donné dans l'équation (18.61), contient les variables de transfert.

### 18.3.3 Expression des sensibilités par rapport aux couplages $J_{gaba}$

#### 18.3.3.1 Modèle perturbé et expression canonique

Le nouveau modèle perturbé s'écrit :

$$\begin{cases} \vec{F}(\vec{s}, \vec{x}_{inter}, \vec{J}_{gaba} + \delta \vec{J}_{gaba}) = \vec{0} \\ \vec{x}_{inter} = \vec{w}_{inter}(\vec{s}) \end{cases} \quad (18.68)$$

Nous linéarisons autour du point fixe

$$\begin{cases} \frac{\partial \vec{F}}{\partial \vec{s}} \delta \vec{s} + \frac{\partial \vec{F}}{\partial \vec{x}_{inter}} \delta \vec{x}_{inter} + \frac{\partial \vec{F}}{\partial \vec{J}_{gaba}} \delta \vec{J}_{gaba} = \vec{0} \\ \delta \vec{x}_{inter} = \frac{\partial \vec{w}_{inter}}{\partial \vec{s}} \delta \vec{s} \end{cases} \quad (18.69)$$

Nous éliminons  $\vec{x}_{inter}$  en insérant la seconde équation dans la première :

$$\left( \frac{\partial \vec{F}}{\partial \vec{s}} + \frac{\partial \vec{F}}{\partial \vec{x}_{inter}} \frac{\partial \vec{w}_{inter}}{\partial \vec{s}} \right) \delta \vec{s} = - \frac{\partial \vec{F}}{\partial \vec{J}_{gaba}} \delta \vec{J}_{gaba} \quad (18.70)$$

L'équation 18.70 peut s'exprimer sous la forme canonique suivante :

$$\left( \bar{\mathbb{I}} - \left( - \frac{\partial \vec{F}}{\partial \vec{s}}^{-1} \frac{\partial \vec{F}}{\partial \vec{x}_{inter}} \frac{\partial \vec{w}_{inter}}{\partial \vec{s}} \right) \right) \delta \vec{s} = - \frac{\partial \vec{F}}{\partial \vec{s}}^{-1} \frac{\partial \vec{F}}{\partial \vec{J}_{gaba}} \delta \vec{J}_{gaba} \quad (18.71)$$

Dans cette expression,  $\frac{\partial \vec{w}_{inter}}{\partial \vec{s}}$  contient tous les couplages entre les deux régions :

$$\frac{\partial \vec{w}_{inter}}{\partial \vec{s}} = \begin{pmatrix} 0 & \frac{\partial w_{inter,1}}{\partial s_{n2}} & 0 & 0 \\ \frac{\partial w_{inter,2}}{\partial s_{n1}} & 0 & 0 & 0 \\ 0 & \frac{\partial w_{inter,1}}{\partial s_{n2}} & 0 & 0 \\ \frac{\partial w_{inter,2}}{\partial s_{n1}} & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & k_{E_{12}} \kappa_{12} & 0 & 0 \\ k_{E_{21}} \kappa_{21} & 0 & 0 & 0 \\ 0 & (1 - k_{E_{12}}) \kappa_{12} & 0 & 0 \\ (1 - k_{E_{21}}) \kappa_{21} & 0 & 0 & 0 \end{pmatrix} \quad (18.72)$$

Si nous annulons  $\frac{\partial \vec{w}_{inter}}{\partial \vec{s}} = \vec{0}$ , le second membre reste inchangé. Nous obtenons l'effet de la perturbation  $\vec{J}_{gaba}$  sur la variable d'observation  $\delta \vec{s}$  dans le système découplé<sup>2</sup> :

$$\delta \vec{s}^O = - \frac{\partial \vec{F}}{\partial \vec{s}}^{-1} \frac{\partial \vec{F}}{\partial \vec{J}_{gaba}} \delta \vec{J}_{gaba} \quad (18.73)$$

---

2. notation avec l'exposant  $O$  pour "open loop"

L'expression canonique se lit :

$$\boxed{(\bar{\mathbb{I}} - \bar{G})\vec{\delta s} = \vec{\delta s}^O} \quad (18.74)$$

où  $\bar{G}$  est la matrice de gains de rétroaction :

$$\bar{G} = -\frac{\overline{\partial F}}{\partial s}^{-1} \frac{\overline{\partial F}}{\partial x_{inter}} \frac{\overline{\partial w_{inter}}}{\partial s} \quad (18.75)$$

### 18.3.3.2 Sensibilités découplées par rapport à $J_{gaba_1}$ et $J_{gaba_2}$

1) Perturbation de  $J_{gaba_1}$ . En boucle ouverte, la région 2 n'est pas perturbée.

Nous notons  $\mathcal{A}_{sn_1, J_{gaba_1}}$  et  $\mathcal{A}_{sg_1, J_{gaba_1}}$  les sensibilités par rapport à  $J_{gaba_1}$  des pools exciteur et inhibiteur de la région 1 découplée de la région 2 :

$$\vec{\delta s}_{/J_{gaba_1}}^O = \begin{pmatrix} \delta sn_1 \\ \delta sn_2 \\ \delta sg_1 \\ \delta sg_2 \end{pmatrix}_{/J_{gaba_1}}^O = \begin{pmatrix} \mathcal{A}_{sn_1, J_{gaba_1}} \delta J_{gaba_1} \\ 0 \\ \mathcal{A}_{sg_1, J_{gaba_1}} \delta J_{gaba_1} \\ 0 \end{pmatrix} \quad (18.76)$$

Nous allons exprimer ces sensibilités en fonction de la sensibilité de la région 1 découplée par rapport à son forçage sur le pool exciteur.

2) Perturbation de  $J_{gaba_2}$ . En boucle ouverte, la région 1 n'est pas perturbée.

Nous notons  $\mathcal{A}_{sn_2, J_{gaba_2}}$  et  $\mathcal{A}_{sg_2, J_{gaba_2}}$  les sensibilités par rapport à  $J_{gaba_2}$  des pools exciteur et inhibiteur de la région 2 découplée de la région 1 :

$$\vec{\delta s}_{/J_{gaba_2}}^O = \begin{pmatrix} \delta sn_1 \\ \delta sn_2 \\ \delta sg_1 \\ \delta sg_2 \end{pmatrix}_{/J_{gaba_2}}^O = \begin{pmatrix} 0 \\ \mathcal{A}_{sn_2, J_{gaba_2}} \delta J_{gaba_2} \\ 0 \\ \mathcal{A}_{sg_2, J_{gaba_2}} \delta J_{gaba_2} \end{pmatrix} \quad (18.77)$$

Nous allons exprimer ces sensibilités en fonction de la sensibilité de la région 2 découplée par rapport à son forçage sur le pool exciteur.

### 18.3.3.3 En utilisant les sensibilités de régions découplées par rapport aux forçages sur les pools excitateurs

Pour écrire les sensibilités découplées respectivement par rapport à  $J_{gaba_1}$  et  $J_{gaba_2}$  en fonction de celles par rapport aux forçages sur les pools excitateurs, nous allons d'abord exprimer  $\overline{\overline{\frac{\partial F}{\partial J_{gaba}}}}$  en fonction de  $\overline{\overline{\frac{\partial F}{\partial x_{inter}}}}$  :

$$\overline{\overline{\frac{\partial F}{\partial J_{gaba}}}} = \begin{pmatrix} \frac{\partial F n_1}{\partial J_{gaba_1}} & \frac{\partial F n_1}{\partial J_{gaba_2}} \\ \frac{\partial F n_2}{\partial J_{gaba_1}} & \frac{\partial F n_2}{\partial J_{gaba_2}} \\ \frac{\partial F g_1}{\partial J_{gaba_1}} & \frac{\partial F g_1}{\partial J_{gaba_2}} \\ \frac{\partial F g_2}{\partial J_{gaba_1}} & \frac{\partial F g_2}{\partial J_{gaba_2}} \end{pmatrix} = \begin{pmatrix} \frac{\partial F n_1}{\partial J_{gaba_1}} & 0 \\ 0 & \frac{\partial F n_2}{\partial J_{gaba_2}} \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (18.78)$$

Chaque éléments  $\frac{\partial F n_i}{\partial J_{gaba_i}}$  peut s'écrire en fonction de  $\frac{\partial F n_i}{\partial x n_{inter,i}}$ . Pour cela, nous reformulons  $F n_i$  (cf équation (18.67)) en séparant le terme contenant  $J_{gaba_i}$

$$\begin{cases} \tilde{F} n_i(s n_i^*, x n_{gaba,i}^*, x n_{inter,i}^*) = -\beta^E s n_i^* \\ \quad + \alpha^E T_{glu} (1 - s n_i^*) \frac{a_E (W + J_{nmda} s n_i^* + x n_{gaba,i}^* + x n_{inter,i}^*) - b_E}{1 - e^{-d_E [a_E (W + J_{nmda} s n_i^* + x n_{gaba,i}^* + x n_{inter,i}^*) - b_E]}} \\ x n_{gaba,i}^* = w n_{gaba,i} (J_{gaba_i}, s g_i^*) = -J_{gaba_i} s g_i^* \end{cases} \quad (18.79)$$

Nous notons que

$$\frac{\partial F n_i}{\partial J_{gaba_i}} = \frac{\partial \tilde{F} n_i}{\partial x n_{gaba,i}} \frac{\partial w n_{gaba,i}}{\partial J_{gaba_i}} \quad (18.80)$$

Par ailleurs,

$$\frac{\partial \tilde{F} n_i}{\partial x n_{gaba,i}} = \frac{\partial \tilde{F} n_i}{\partial x n_{inter,i}} = \frac{\partial F n_i}{\partial x n_{inter,i}} \quad (18.81)$$

Nous en déduisons que

$$\frac{\partial F n_i}{\partial J_{gaba_i}} = \frac{\partial F n_i}{\partial x n_{inter,i}} (-s g_i^*) \quad (18.82)$$

La matrice  $\overline{\overline{\frac{\partial F}{\partial J_{gaba}}}}$  s'écrit alors

$$\overline{\overline{\frac{\partial F}{\partial J_{gaba}}}} = \begin{pmatrix} \frac{\partial F n_1}{\partial x n_{inter,1}} (-s g_1^*) & 0 \\ 0 & \frac{\partial F n_2}{\partial x n_{inter,2}} (-s g_2^*) \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (18.83)$$

Or

$$\overline{\overline{\frac{\partial F}{\partial x_{inter}}}} = \begin{pmatrix} \frac{\partial F n_1}{\partial x_{inter,1}} & \frac{\partial F n_1}{\partial x_{inter,2}} & \frac{\partial F n_1}{\partial x_{g_{inter,1}}} & \frac{\partial F n_1}{\partial x_{g_{inter,2}}} \\ \frac{\partial F n_2}{\partial x_{inter,1}} & \frac{\partial F n_2}{\partial x_{inter,2}} & \frac{\partial F n_2}{\partial x_{g_{inter,1}}} & \frac{\partial F n_2}{\partial x_{g_{inter,2}}} \\ \frac{\partial F g_1}{\partial x_{inter,1}} & \frac{\partial F g_1}{\partial x_{inter,2}} & \frac{\partial F g_1}{\partial x_{g_{inter,1}}} & \frac{\partial F g_1}{\partial x_{g_{inter,2}}} \\ \frac{\partial F g_2}{\partial x_{inter,1}} & \frac{\partial F g_2}{\partial x_{inter,2}} & \frac{\partial F g_2}{\partial x_{g_{inter,1}}} & \frac{\partial F g_2}{\partial x_{g_{inter,2}}} \end{pmatrix} = \begin{pmatrix} \frac{\partial F n_1}{\partial x_{inter,1}} & 0 & 0 & 0 \\ 0 & \frac{\partial F n_2}{\partial x_{inter,2}} & 0 & 0 \\ 0 & 0 & \frac{\partial F g_1}{\partial x_{g_{inter,1}}} & 0 \\ 0 & 0 & 0 & \frac{\partial F g_2}{\partial x_{g_{inter,2}}} \end{pmatrix} \quad (18.84)$$

Définissons  $\overline{\overline{sg}}$  une matrice contenant les valeurs  $sg_i^*$  dans le système couplé :

$$\overline{\overline{sg}} = \begin{pmatrix} sg_1^* & 0 \\ 0 & sg_2^* \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (18.85)$$

en conséquence, nous avons

$$\overline{\overline{\frac{\partial F}{\partial J_{gaba}}}} = \begin{pmatrix} \frac{\partial F n_1}{\partial x_{inter,1}}(-sg_1^*) & 0 \\ 0 & \frac{\partial F n_2}{\partial x_{inter,2}}(-sg_2^*) \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (18.86)$$

$$= \begin{pmatrix} \frac{\partial F n_1}{\partial x_{inter,1}} & 0 & 0 & 0 \\ 0 & \frac{\partial F n_2}{\partial x_{inter,2}} & 0 & 0 \\ 0 & 0 & \frac{\partial F g_1}{\partial x_{g_{inter,1}}} & 0 \\ 0 & 0 & 0 & \frac{\partial F g_2}{\partial x_{g_{inter,2}}} \end{pmatrix} \begin{pmatrix} sg_1^* & 0 \\ 0 & sg_2^* \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \overline{\overline{\frac{\partial F}{\partial x_{inter}}}}(-\overline{\overline{sg}}) \quad (18.87)$$

Nous remplaçons  $\overline{\overline{\frac{\partial F}{\partial J_{gaba}}}}$  dans l'expression canonique (18.71) :

$$\left( \overline{\overline{\mathbb{I}}} - \left( -\frac{\overline{\overline{\partial F}}}{\partial s} \frac{\overline{\overline{\partial F}}}{\partial x_{inter}} \frac{\overline{\overline{\partial w_{inter}}}}{\partial s} \right) \right) \overrightarrow{\delta s} = -\frac{\overline{\overline{\partial F}}}{\partial s} \frac{\overline{\overline{\partial F}}}{\partial x_{inter}} (-\overline{\overline{sg}}) \overrightarrow{\delta J_{gaba}} \quad (18.88)$$

Nous reprenons alors le travail analytique sur les sensibilités dans deux régions par rapport au forçage sur le pool exciteur. Celui-ci se trouve dans l'annexe de l'article (chapitre 9).

Nous y avons défini :

$$-\frac{\overline{\overline{\partial F}}}{\partial s} \frac{\overline{\overline{\partial F}}}{\partial x_{inter}} = \overline{\overline{S}} \quad (18.89)$$

$$= \left( \begin{array}{cc|cc} \mathcal{A}_{sn_1, x_{E_1}} & 0 & \mathcal{A}_{sn_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sn_2, x_{E_2}} & 0 & \mathcal{A}_{sn_2, x_{I_2}} \\ \hline \mathcal{A}_{sg_1, x_{E_1}} & 0 & \mathcal{A}_{sg_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sg_2, x_{E_2}} & 0 & \mathcal{A}_{sg_2, x_{I_2}} \end{array} \right) \quad (18.90)$$

$$= \left( \begin{array}{c|c} \overline{\overline{\mathcal{A}_{sn, x_E}}} & \overline{\overline{\mathcal{A}_{sn, x_I}}} \\ \hline \overline{\overline{\mathcal{A}_{sg, x_E}}} & \overline{\overline{\mathcal{A}_{sg, x_I}}} \end{array} \right) \quad (18.91)$$



Les sensibilités des régions découplées par rapport aux forçages sont évaluées aux valeurs de points fixes  $\vec{s}_1^* = (sn_1^*, sg_1^*)$  and  $\vec{s}_2^* = (sn_2^*, sg_2^*)$  dans le système couplé. Il est nécessaire de prendre en considération les forçages externes totaux  $xn_{inter,i}^*$  et  $xg_{inter,i}^*$ . Ceux-ci sont définis dans l'équation (18.61).

$$\begin{cases} \mathcal{A}_{sn_i, x_{E_i}} = \Phi n_E(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \\ \mathcal{A}_{sg_i, x_{E_i}} = \Phi g_E(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \\ \mathcal{A}_{sn_i, x_{I_i}} = \Phi n_I(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \\ \mathcal{A}_{sg_i, x_{I_i}} = \Phi g_I(\vec{s}^* = \vec{s}_i^*, x_E = xn_{inter,i}^*, x_I = xg_{inter,i}^*) \end{cases} \quad (18.92)$$

Les expressions analytiques des fonctions  $\Phi n_E$ ,  $\Phi g_E$ ,  $\Phi n_I$  and  $\Phi g_I$  sont données de manière explicite dans l'annexe de l'article.

Nous avons donc :

$$\vec{\delta s}^O = \left( \begin{array}{cc|cc} \mathcal{A}_{sn_1, x_{E_1}} & 0 & \mathcal{A}_{sn_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sn_2, x_{E_2}} & 0 & \mathcal{A}_{sn_2, x_{I_2}} \\ \hline \mathcal{A}_{sg_1, x_{E_1}} & 0 & \mathcal{A}_{sg_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sg_2, x_{E_2}} & 0 & \mathcal{A}_{sg_2, x_{I_2}} \end{array} \right) \left( \begin{array}{cc} -sg_1^* & 0 \\ 0 & -sg_2^* \\ \hline 0 & 0 \\ 0 & 0 \end{array} \right) \begin{pmatrix} \delta J_{gaba_1} \\ \delta J_{gaba_2} \end{pmatrix} \quad (18.93)$$

$$= \begin{pmatrix} \mathcal{A}_{sn_1, x_{E_1}}(-sg_1^*)\delta J_{gaba_1} \\ \mathcal{A}_{sn_2, x_{E_2}}(-sg_2^*)\delta J_{gaba_2} \\ \mathcal{A}_{sg_1, x_{E_1}}(-sg_1^*)\delta J_{gaba_1} \\ \mathcal{A}_{sg_2, x_{E_2}}(-sg_2^*)\delta J_{gaba_2} \end{pmatrix} \quad (18.94)$$

Nous retrouvons

$$\begin{cases} \mathcal{A}_{sn_1, J_{gaba_1}} = -\mathcal{A}_{sn_1, x_{E_1}} sg_1^* \\ \mathcal{A}_{sn_2, J_{gaba_2}} = -\mathcal{A}_{sn_2, x_{E_2}} sg_2^* \\ \mathcal{A}_{sg_1, J_{gaba_1}} = -\mathcal{A}_{sg_1, x_{E_1}} sg_1^* \\ \mathcal{A}_{sg_2, J_{gaba_2}} = -\mathcal{A}_{sg_2, x_{E_2}} sg_2^* \end{cases} \quad (18.95)$$

#### 18.3.3.4 Sensibilités dans le système couplé en fonction des sensibilités découplées

En introduisant la matrice  $\bar{\bar{S}}$  dans l'expression canonique (18.88) nous obtenons :

$$\left( \bar{\bar{I}} - \bar{\bar{S}} \frac{\partial w_{inter}}{\partial s} \right) \vec{\delta s} = \bar{\bar{S}}(-\bar{\bar{sg}}) \vec{\delta J_{gaba}} \quad (18.96)$$

La perturbation  $\vec{\delta s}$  dans le système couplé vaut :

$$\vec{\delta s} = \left( \bar{\mathbb{I}} - \bar{\overline{S}} \frac{\overline{\partial w_{inter}}}{\partial s} \right)^{-1} \bar{\overline{S}} (-\bar{s}\bar{g}) \overrightarrow{\delta J_{gaba}} = \left( \bar{\mathbb{I}} - \bar{\overline{G}} \right)^{-1} \bar{\overline{S}} (-\bar{s}\bar{g}) \overrightarrow{\delta J_{gaba}} \quad (18.97)$$

Pour calculer  $\vec{\delta s}$ , nous reprenons l'expression de la matrice de gain, égale au produit matriciel suivant :

$$\bar{\overline{G}} = \begin{pmatrix} \mathcal{A}_{sn_1, x_{E_1}} & 0 & \mathcal{A}_{sn_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sn_2, x_{E_2}} & 0 & \mathcal{A}_{sn_2, x_{I_2}} \\ \mathcal{A}_{sg_1, x_{E_1}} & 0 & \mathcal{A}_{sg_1, x_{I_1}} & 0 \\ 0 & \mathcal{A}_{sg_2, x_{E_2}} & 0 & \mathcal{A}_{sg_2, x_{I_2}} \end{pmatrix} \begin{pmatrix} 0 & k_{E_{12}} \kappa_{12} & 0 & 0 \\ k_{E_{21}} \kappa_{21} & 0 & 0 & 0 \\ 0 & (1 - k_{E_{12}}) \kappa_{12} & 0 & 0 \\ (1 - k_{E_{21}}) \kappa_{21} & 0 & 0 & 0 \end{pmatrix} \quad (18.98)$$

que nous écrivons dans l'article comme

$$\bar{\overline{G}} = \begin{bmatrix} 0 & G_{12} & 0 & 0 \\ G_{21} & 0 & 0 & 0 \\ 0 & G_{32} & 0 & 0 \\ G_{41} & 0 & 0 & 0 \end{bmatrix} \quad (18.99)$$

avec

$$\begin{cases} G_{12} = \mathcal{A}_{sn_1, x_{E_1}} k_{E_{12}} \kappa_{12} + \mathcal{A}_{sn_1, x_{I_1}} (1 - k_{E_{12}}) \kappa_{12} \\ G_{21} = \mathcal{A}_{sn_2, x_{E_2}} k_{E_{21}} \kappa_{21} + \mathcal{A}_{sn_2, x_{I_2}} (1 - k_{E_{21}}) \kappa_{21} \\ G_{32} = \mathcal{A}_{sg_1, x_{E_1}} k_{E_{12}} \kappa_{12} + \mathcal{A}_{sg_1, x_{I_1}} (1 - k_{E_{12}}) \kappa_{12} \\ G_{41} = \mathcal{A}_{sg_2, x_{E_2}} k_{E_{21}} \kappa_{21} + \mathcal{A}_{sg_2, x_{I_2}} (1 - k_{E_{21}}) \kappa_{21} \end{cases} \quad (18.100)$$

Nous définissons les blocs de la différence  $\bar{\mathbb{I}} - \bar{\overline{G}}$  :

$$\bar{\mathbb{I}} - \bar{\overline{G}} = \left( \begin{array}{cc|cc} 1 & -G_{12} & 0 & 0 \\ -G_{21} & 1 & 0 & 0 \\ 0 & -G_{32} & 1 & 0 \\ -G_{41} & 0 & 0 & 1 \end{array} \right) = \left( \begin{array}{c|c} \bar{\overline{B}}_1 & \bar{\overline{B}}_2 \\ \hline \bar{\overline{B}}_3 & \bar{\overline{B}}_4 \end{array} \right) \quad (18.101)$$

$(\bar{\mathbb{I}} - \bar{\overline{G}})^{-1} \bar{\overline{S}}$  a été calculé dans l'article :

$$(\bar{\mathbb{I}} - \bar{\overline{G}})^{-1} \bar{\overline{S}} = \left( \begin{array}{c|c} \bar{\overline{B}}_1^{-1} \bar{\overline{\mathcal{A}}}_{sn, x_E} & \bar{\overline{B}}_1^{-1} \bar{\overline{\mathcal{A}}}_{sn, x_I} \\ \hline -\bar{\overline{B}}_3 \bar{\overline{B}}_1^{-1} \bar{\overline{\mathcal{A}}}_{sn, x_E} + \bar{\overline{\mathcal{A}}}_{sg, x_E} & -\bar{\overline{B}}_3 \bar{\overline{B}}_1^{-1} \bar{\overline{\mathcal{A}}}_{sn, x_I} + \bar{\overline{\mathcal{A}}}_{sg, x_I} \end{array} \right) \quad (18.102)$$

Nous pouvons finalement calculer la perturbation  $\vec{\delta s}$ , dans le système couplé suite à une perturbation des paramètres  $\overrightarrow{J_{gaba}}$  :

$$\vec{\delta s} = (\bar{\mathbb{I}} - \bar{G})^{-1} \bar{S}(-\bar{sg}) \overrightarrow{\delta J_{gaba}} \quad (18.103)$$

$$= \left( \begin{array}{c|c} \overline{\overline{B_1}^{-1} \mathcal{A}_{sn,x_E}} & \overline{\overline{B_1}^{-1} \mathcal{A}_{sn,x_I}} \\ \hline \overline{\overline{-B_3 B_1}^{-1} \mathcal{A}_{sn,x_E} + \mathcal{A}_{sg,x_E}} & \overline{\overline{-B_3 B_1}^{-1} \mathcal{A}_{sn,x_I} + \mathcal{A}_{sg,x_I}} \end{array} \right) \begin{pmatrix} -sg_1^* & 0 \\ 0 & -sg_2^* \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \overrightarrow{\delta J_{gaba}} \quad (18.104)$$

$$= \begin{pmatrix} \delta sn_1 \\ \delta sn_2 \\ \delta sg_1 \\ \delta sg_2 \end{pmatrix} \quad (18.105)$$

$\vec{\delta s}$  est exprimé en fonction des sensibilités des régions découplées par rapport à leur forçage, qui dépendent elles-mêmes des sensibilités des pools découplés exciteur et inhibiteur de la région d'intérêt.

### 18.3.4 Sensibilités dans un système contrôle-cible

#### 18.3.4.1 Extraction des cas d'intérêt

Nous nous intéressons particulièrement à l'activité des pools excitateurs des régions de contrôle et cible :

$$\begin{pmatrix} \delta sn_{rc} \\ \delta sn_c \end{pmatrix} = \overline{\overline{B_1}^{-1} \mathcal{A}_{sn,x_E}} \begin{pmatrix} -sg_{rc}^* & 0 \\ 0 & -sg_c^* \end{pmatrix} \overrightarrow{\delta J_{gaba}} \quad (18.106)$$

D'après la définition 18.101, nous avons :

$$\overline{\overline{B_1}^{-1}} = \begin{pmatrix} 1 & -G_{12} \\ -G_{21} & 1 \end{pmatrix}^{-1} = \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} 1 & G_{12} \\ G_{21} & 1 \end{pmatrix} \quad (18.107)$$

Nous retenons :

$$\begin{pmatrix} \delta sn_{rc} \\ \delta sn_c \end{pmatrix} = \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} 1 & G_{12} \\ G_{21} & 1 \end{pmatrix} \begin{pmatrix} \mathcal{A}_{sn_{rc},x_{E_{rc}}} & 0 \\ 0 & \mathcal{A}_{sn_c,x_{E_c}} \end{pmatrix} \begin{pmatrix} -sg_{rc}^* & 0 \\ 0 & -sg_c^* \end{pmatrix} \begin{pmatrix} \delta J_{gaba_{rc}} \\ \delta J_{gaba_c} \end{pmatrix} \quad (18.108)$$

$$= \frac{1}{1 - G_{12}G_{21}} \begin{pmatrix} \mathcal{A}_{sn_{rc},x_{E_{rc}}}(-sg_{rc}^*)\delta J_{gaba_{rc}} + G_{12}\mathcal{A}_{sn_c,x_{E_c}}(-sg_c^*)\delta J_{gaba_c} \\ G_{21}\mathcal{A}_{sn_{rc},x_{E_{rc}}}(-sg_{rc}^*)\delta J_{gaba_{rc}} + \mathcal{A}_{sn_c,x_{E_c}}(-sg_c^*)\delta J_{gaba_c} \end{pmatrix} \quad (18.109)$$

Avec

$$\begin{cases} G_{12} = \mathcal{A}_{sn_{rc}, x_{E_{rc}}} k_{E_{rc}-c} \kappa_{rc-c} + \mathcal{A}_{sn_{rc}, x_{I_{rc}}} (1 - k_{E_{rc}-c}) \kappa_{rc-c} \\ G_{21} = \mathcal{A}_{sn_c, x_{E_c}} k_{E_c-rc} \kappa_{c-rc} + \mathcal{A}_{sn_c, x_{I_c}} (1 - k_{E_c-rc}) \kappa_{c-rc} \end{cases} \quad (18.110)$$

Le gain de rétroaction  $g = G_{12}G_{21}$  est strictement identique à celui obtenu pour une perturbation du forçage excitateur  $B_{E_{rc}}$  (cf article du chapitre 9).

### 18.3.4.2 Ecriture générique de la connectivité

Nous reprenons la convention d'écriture générique de connectivité présente dans l'article pour écrire :

$$\boxed{\begin{cases} g_{\square-\square} = \mathcal{A}_{sn_{rc}, x_{\square_{rc}}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{\square_c}} \kappa_{c-rc} \\ \delta sn_{rc} = \frac{1}{1 - g_{\square-\square}} \left( \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^*) \delta J_{gaba_{rc}} + \mathcal{A}_{sn_{rc}, x_{\square_{rc}}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{E_c}} (-sg_c^*) \delta J_{gaba_c} \right) \\ \delta sn_c = \frac{1}{1 - g_{\square-\square}} \left( \mathcal{A}_{sn_c, x_{\square_c}} \kappa_{c-rc} \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^*) \delta J_{gaba_{rc}} + \mathcal{A}_{sn_c, x_{E_c}} (-sg_c^*) \delta J_{gaba_c} \right) \end{cases}} \quad (18.111)$$

### 18.3.4.3 Perturbation de $J_{gaba_c}$ uniquement

#### 18.3.4.3.1 Dans le système couplé

Dans le cas où  $\delta J_{gaba_{rc}} = 0$ , le système (18.111) s'écrit :

$$\begin{cases} g_{\square-\square} = \mathcal{A}_{sn_{rc}, x_{\square_{rc}}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{\square_c}} \kappa_{c-rc} \\ \delta sn_{rc}/J_{gaba_c} = \frac{1}{1 - g_{\square-\square}} \left( \mathcal{A}_{sn_{rc}, x_{\square_{rc}}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{E_c}} (-sg_c^*) \delta J_{gaba_c} \right) \\ \delta sn_c/J_{gaba_c} = \frac{1}{1 - g_{\square-\square}} \left( \mathcal{A}_{sn_c, x_{E_c}} (-sg_c^*) \delta J_{gaba_c} \right) \end{cases} \quad (18.112)$$

Supposons  $1 - g_{\square-\square} > 0$  (même raisonnement que l'article du chapitre 9). Nous étudions le signe de  $\delta sn_{rc}/J_{gaba_c}$  et  $\delta sn_c/J_{gaba_c}$  en fonction de la connectivité.

— Pour la région cible :

$\mathcal{A}_{sn_c, x_{E_c}} > 0$  et  $sg_c^* > 0$  donc  $\delta sn_c/J_{gaba_c} < 0$ .

Pour n'importe quel type de connectivité, la perturbation de  $\delta J_{gaba_c}$  conduit à un abaissement de l'activité excitatrice  $sn_c^*$ .

— Pour la région de contrôle :

Le signe de  $\delta sn_{rc}/J_{gaba_c}$  dépend par contre de  $\mathcal{A}_{sn_{rc}, x_{\square_{rc}}}$  et donc du type de connectivité :

| $\square - \square$ | $\mathcal{A}_{sn_{rc}, x_{\square_{rc}}}$ | $\delta sn_{rc} / J_{gaba_c}$ |
|---------------------|-------------------------------------------|-------------------------------|
| $E - E$             | $> 0$                                     | $< 0$                         |
| $I - E$             | $< 0$                                     | $> 0$                         |
| $E - I$             | $> 0$                                     | $< 0$                         |
| $I - I$             | $< 0$                                     | $> 0$                         |

Le signe de  $\delta sn_{rc} / J_{gaba_c}$  dépend uniquement du pool par lequel la région de contrôle est connectée, et ne dépend pas du mode de connexion sur la région cible. Quand la région de contrôle est connectée par son pool excitateur, la perturbation de  $\delta J_{gaba_c}$  conduit à un abaissement de l'activité excitatrice  $sn_{rc}^*$ . Au contraire, quand la région de contrôle est connectée par son pool inhibiteur, la perturbation de  $\delta J_{gaba_c}$  conduit à une augmentation de l'activité excitatrice  $sn_{rc}^*$ .

Nous allons tracer les sensibilités à une perturbation de  $J_{gaba_c}$  dans le cas d'une architecture de connectivité EI (représentée à la figure 18.7) et dans le cas découplé où il n'y a pas de rétroaction de la cible vers la région de contrôle (cf figure 18.8).

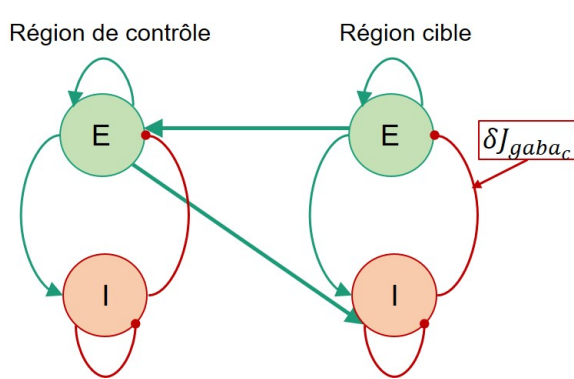


**Figure 18.7** – Perturbation de  $\delta J_{gaba_c}$  dans le cas de la connectivité EI

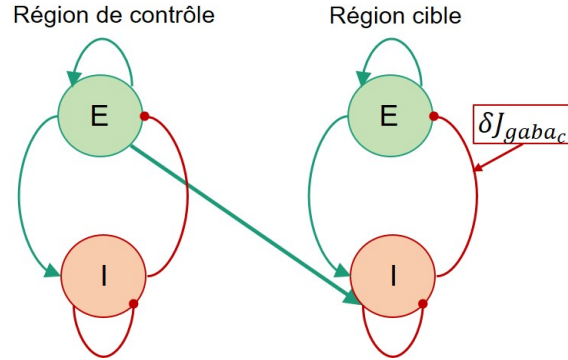


**Figure 18.8** – Perturbation de  $\delta J_{gaba_c}$  dans le cas sans rétroaction de la cible vers la région de contrôle

### 18.3.4.3.2 Cas de la connectivité EI

Dans le cas d'une connectivité EI entre la région de contrôle et la cible, le système (18.112) devient

$$\begin{cases} g_{E-I} = \mathcal{A}_{sn_{rc}, x_{Erc}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{Ic}} \kappa_{c-rc} \\ \delta sn_{rc}/J_{gaba_c} = \frac{1}{1 - g_{E-I}} \mathcal{A}_{sn_{rc}, x_{Erc}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{Ec}} (-sg_c^*) \delta J_{gaba_c} \\ \delta sn_c/J_{gaba_c} = \frac{1}{1 - g_{E-I}} \mathcal{A}_{sn_c, x_{Ec}} (-sg_c^*) \delta J_{gaba_c} \end{cases} \quad (18.113)$$

Comme attendu,  $\delta sn_{rc}/J_{gaba_c} \neq 0$ , la perturbation de  $J_{gaba_c}$  a un effet sur l'activité de la région de contrôle, et pas uniquement sur celle de la cible (cf figure 18.7).

Nous avons représenté sur la figure 18.9 l'activité  $sn_{rc}^*$  de la région de contrôle et  $sn_c^*$  de la cible au régime stationnaire, en fonction de  $J_{gaba_c}$ , pour deux valeurs différentes du forçage sur le pool excitateur de la région de contrôle :  $\widehat{B_{Erc}} = 0.382 + I_{stim}$ .

Dans les deux régions, l'activité excitatrice de la région au régime stationnaire n'est pas constante quand  $J_{gaba_c}$  varie, ce qui est cohérent.

La sensibilité  $\frac{\delta sn_{rc}/J_{gaba_c}}{\delta J_{gaba_c}}$ , évaluée pour une valeur fixée du forçage  $B_{Erc} = \widehat{B_{Erc}}$  et de  $J_{gaba_c} = \widehat{J_{gaba_c}}$ , est donnée par la pente de la tangente à la courbe  $sn_{rc}^* = f_{rc}(\widehat{B_{Erc}}, J_{gaba_c})$  au voisinage de  $J_{gaba_c} = \widehat{J_{gaba_c}}$ .

Il en est de même pour la sensibilité de la cible à  $J_{gaba_c}$ , i.e.  $\frac{\delta sn_c/J_{gaba_c}}{\delta J_{gaba_c}}$ , en considérant la courbe  $sn_c^* = f_c(\widehat{B_{Erc}}, J_{gaba_c})$  sur la figure 18.9 à droite.

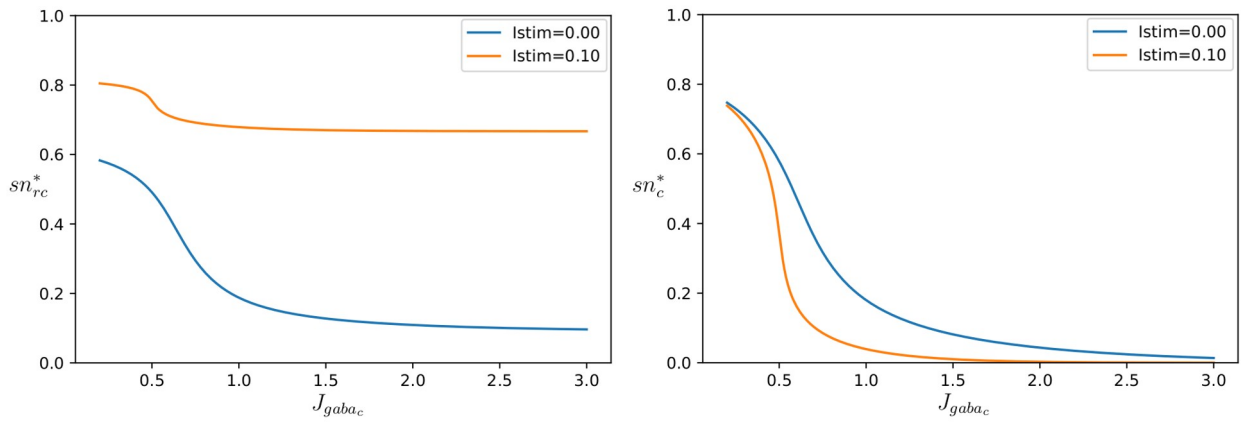


**Figure 18.9** – Activité de la région de contrôle  $sn_{rc}^*$  (à gauche) et de la cible  $sn_c^*$  (à droite) connectée en EI, en fonction de  $J_{gaba_c}$  dans la cible, pour deux valeurs du forçage sur le pool excitateur de la région de contrôle. Avec cette boucle de rétroaction, la variation de  $J_{gaba_c}$  a un effet sur l'activité des deux régions.

### 18.3.4.3.3 Sans la rétroaction de la cible vers la région de contrôle

Si nous coupons la rétroaction de la cible vers la région de contrôle, les calculs de sensibilités à  $J_{gaba_c}$  restent valables. Il alors est nécessaire de :

1. fixer le coefficient de couplage  $\kappa_{rc-c} = 0$ ,
2. prendre les valeurs du point fixe dans le système sans la rétroaction, noté  $\vec{s}^{\text{dec}} = (sn_{rc}^{*\text{dec}}, sg_{rc}^{*\text{dec}}, sn_c^{*\text{dec}}, sg_c^{*\text{dec}})$ .

Avec  $\kappa_{rc-c} = 0$ , le système (18.113) devient :

$$\begin{cases} g_{E-I} = 0 \\ \delta sn_{rc}^{\text{dec}} / J_{gaba_c} = 0 \\ \delta sn_c^{\text{dec}} / J_{gaba_c} = \mathcal{A}_{sn_c, x_{E_c}}(-sg_c^{*\text{dec}}) \delta J_{gaba_c} \end{cases} \quad (18.114)$$

Comme attendu, la perturbation de  $J_{gaba_c}$  n'a pas d'effet sur l'activité de la région de contrôle ( $\delta sn_{rc}^{\text{dec}} / J_{gaba_c} = 0$ ), mais uniquement sur celle de la cible ( $\delta sn_c^{\text{dec}} / J_{gaba_c} \neq 0$ ). Nous pouvons nous reporter à la figure 18.8 pour visualiser la boucle de rétroaction.

Nous avons représenté sur la figure 18.10 l'activité  $sn_{rc}^{*\text{dec}}$  de la région de contrôle et  $sn_c^{*\text{dec}}$  de la cible dans le système sans rétroaction, en fonction de  $J_{gaba_c}$ , pour deux valeurs différentes du forçage sur le pool excitateur de la région de contrôle :  $\widehat{B}_{E_{rc}} = 0.382 + I_{stim}$ . L'activité excitatrice de la région de contrôle au régime stationnaire est constante quand  $J_{gaba_c}$  varie, contrairement à la région cible, ce qui est cohérent. Les sensibilités  $\frac{\delta sn_{rc}^{\text{dec}} / J_{gaba_c}}{\delta J_{gaba_c}}$  et  $\frac{\delta sn_c^{\text{dec}} / J_{gaba_c}}{\delta J_{gaba_c}}$ , évaluées pour une valeur fixée du forçage  $B_{E_{rc}} = \widehat{B}_{E_{rc}}$  et de  $J_{gaba_c} = \widehat{J}_{gaba_c}$ , sont respectivement égales aux pentes des tangentes aux courbes  $sn_{rc}^{*\text{dec}} = f_{rc}^{\text{dec}}(\widehat{B}_{E_{rc}}, J_{gaba_c})$  et  $sn_c^{*\text{dec}} = f_c^{\text{dec}}(\widehat{B}_{E_{rc}}, J_{gaba_c})$ , au voisinage de  $J_{gaba_c} = \widehat{J}_{gaba_c}$ .

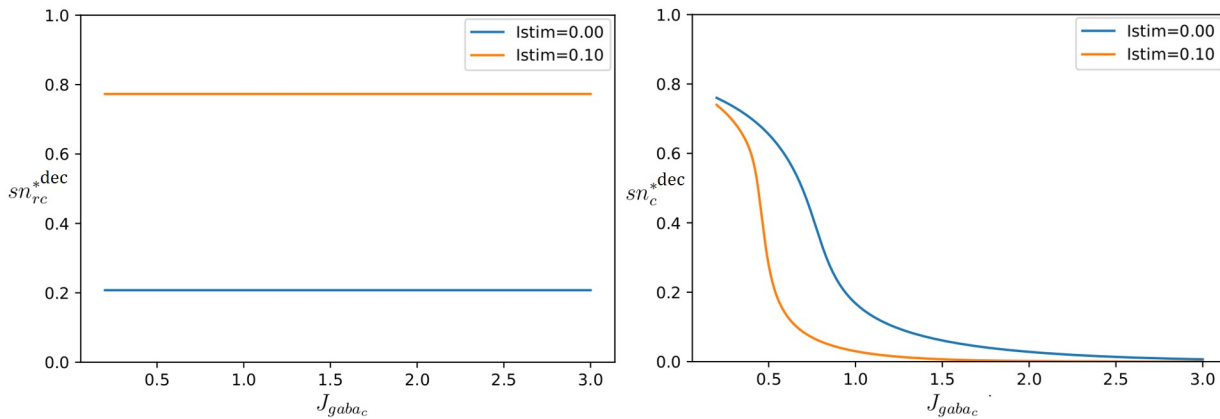


**Figure 18.10** – Activité de la région de contrôle  $sn_{rc}^{*\text{dec}}$  (à gauche) et de la cible  $sn_c^{*\text{dec}}$  (à droite) sans rétroaction de la cible vers la région de contrôle. Avec cette boucle de rétroaction, la variation de  $J_{gaba_c}$  n'a pas d'effet sur l'activité de la région de contrôle.

### 18.3.4.4 Perturbation de $J_{gaba_{rc}}$ uniquement

#### 18.3.4.4.1 Dans le système couplé

Dans le cas où  $\delta J_{gaba_c} = 0$ , le système (18.111) devient :

$$\begin{cases} g_{\square-\square} = \mathcal{A}_{sn_{rc}, x_{\square_{rc}}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{\square_c}} \kappa_{c-rc} \\ \delta sn_{rc}/J_{gaba_{rc}} = \frac{1}{1 - g_{\square-\square}} \left( \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^*) \delta J_{gaba_{rc}} \right) \\ \delta sn_c/J_{gaba_{rc}} = \frac{1}{1 - g_{\square-\square}} \left( \mathcal{A}_{sn_c, x_{\square_c}} \kappa_{c-rc} \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^*) \delta J_{gaba_{rc}} \right) \end{cases} \quad (18.115)$$

Les résultats analytiques sont symétriques à ceux de la perturbation de  $J_{gaba_c}$ , en inversant la région de contrôle et la cible.

Le gain de retroaction est strictement identique à celui obtenu pour une perturbation du forçage  $B_{E_{rc}}$  dans la région de contrôle. En reprenant, les résultats analytiques de l'article du chapitre 9 au paragraphe 2.5, nous pouvons écrire une égalité entre les sensibilités par rapport à  $J_{gaba_{rc}}$  et par rapport au forçage  $B_{E_{rc}}$  :

$$\frac{\delta sn_{rc}/J_{gaba_{rc}}}{\delta J_{gaba_{rc}}} = -sg_{rc}^* \frac{\delta sn_{rc}/B_{E_{rc}}}{\delta B_{E_{rc}}} \quad (18.116)$$

$$\frac{\delta sn_c/J_{gaba_{rc}}}{\delta J_{gaba_{rc}}} = -sg_{rc}^* \frac{\delta sn_c/B_{E_{rc}}}{\delta B_{E_{rc}}} \quad (18.117)$$

Nous allons voir ce que deviennent les sensibilités dans les deux cas représentés sur les figures 18.11 et 18.12.

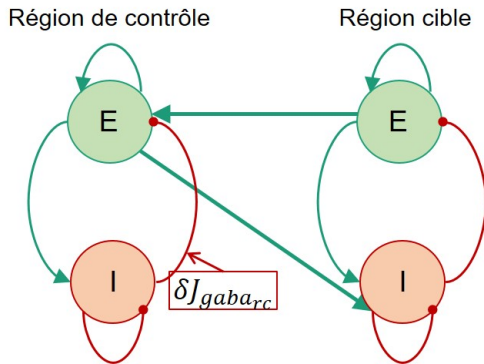


**Figure 18.11** – Perturbation de  $\delta J_{gaba_{rc}}$  dans le cas de la connectivité EI.

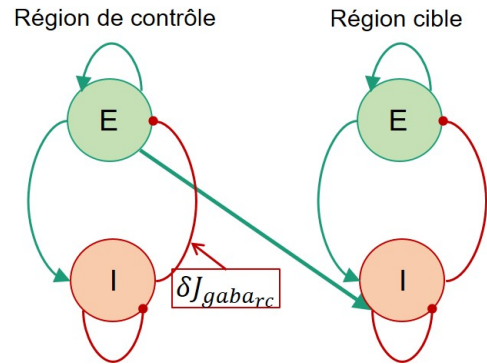


**Figure 18.12** – Perturbation de  $\delta J_{gaba_{rc}}$  dans le cas sans rétroaction de la cible vers la région de contrôle.



#### 18.3.4.4.2 Cas de la connectivité EI

Dans le cas d'une connectivité EI entre la région de contrôle et la cible, le système (18.115) devient :

$$\begin{cases} g_{E-I} = \mathcal{A}_{sn_{rc}, x_{E_{rc}}} \kappa_{rc-c} \mathcal{A}_{sn_c, x_{I_c}} \kappa_{c-rc} \\ \delta sn_{rc}/J_{gaba_{rc}} = \frac{1}{1 - g_{E-I}} \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^*) \delta J_{gaba_{rc}} \\ \delta sn_c/J_{gaba_{rc}} = \frac{1}{1 - g_{E-I}} \mathcal{A}_{sn_c, x_{I_c}} \kappa_{c-rc} \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^*) \delta J_{gaba_{rc}} \end{cases} \quad (18.118)$$

La perturbation de  $J_{gaba_{rc}}$  affecte la région de contrôle et la région cible, ce qui était attendu étant donné la boucle de rétroaction (cf figure 18.11).

#### 18.3.4.4.3 Sans la rétroaction de la cible vers la région de contrôle

Si nous coupons la rétroaction de la cible vers la région de contrôle, les calculs de sensibilités à  $J_{gaba_{rc}}$  restent valables. Il alors est nécessaire de :

1. fixer le coefficient de couplage  $\kappa_{rc-c} = 0$ ,
2. prendre les valeurs du point fixe dans le système sans la rétroaction.

Avec  $\kappa_{rc-c} = 0$ , le système (18.118) devient :

$$\begin{cases} g_{E-I} = 0 \\ \delta sn_{rc}^{\text{dec}}/J_{gaba_{rc}} = \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^{*\text{dec}}) \delta J_{gaba_{rc}} \\ \delta sn_c^{\text{dec}}/J_{gaba_{rc}} = \mathcal{A}_{sn_c, x_{I_c}} \kappa_{c-rc} \mathcal{A}_{sn_{rc}, x_{E_{rc}}} (-sg_{rc}^{*\text{dec}}) \delta J_{gaba_{rc}} = \mathcal{A}_{sn_c, x_{I_c}} \kappa_{c-rc} \delta sn_{rc}^{\text{dec}}/J_{gaba_{rc}} \end{cases} \quad (18.119)$$

De manière intéressante, il est possible de lire l'expression de  $\delta sn_c^{\text{dec}}/J_{gaba_{rc}}$  comme le produit entre un terme d'**émission** ( $\delta sn_{rc}^{\text{dec}}/J_{gaba_{rc}}$ ), de **transmission** ( $\kappa_{c-rc}$ ) et de **réception** ( $\mathcal{A}_{sn_c, x_{I_c}}$ ). De manière causale, la perturbation de  $J_{gaba_{rc}}$  a un effet sur la région de contrôle puis cet effet est transmis à la cible (cf figure 18.12).

## 18.4 Deux régions : Effet d'une augmentation de $I_{stim}$ et $J_{gaba_c}$ sur l'activité $sn_c^*$ de la cible

### 18.4.1 En perturbation

Au régime stationnaire, nous perturbons simultanément  $B_{Erc}$  de  $\delta B_{Erc}$  et  $J_{gaba_c}$  de  $\delta J_{gaba_c}$ .

La linéarisation du modèle perturbé autour du point fixe nous permet d'écrire la perturbation  $\delta sn_c$  comme une somme de deux variations dues respectivement à  $\delta B_{Erc}$  et à  $\delta J_{gaba_c}$ .

$$\delta sn_c = \delta sn_{c/B_{Erc}} + \delta sn_{c/J_{gaba_c}} \quad (18.120)$$

$$= S_{sn_{c/B_{Erc}}}(B_{Erc}, J_{gaba_c})\delta B_{Erc} + S_{sn_{c/J_{gaba_c}}}(B_{Erc}, J_{gaba_c})\delta J_{gaba_c} \quad (18.121)$$

$S_{sn_{c/B_{Erc}}}(B_{Erc}, J_{gaba_c})$  et  $S_{sn_{c/J_{gaba_c}}}(B_{Erc}, J_{gaba_c})$  désignent respectivement les sensibilités de  $sn_c^*$  par rapport à  $B_{Erc}$  et à  $J_{gaba_c}$ . Elles dépendent toutes les deux de  $B_{Erc}$  et  $J_{gaba_c}$ .

Leurs expressions analytiques respectives se trouvent dans les sections 18.2 et 18.3.

### 18.4.2 En variation

Au point fixe, la valeur de  $sn_c^*$  dépend de  $B_{Erc}$  et  $J_{gaba_c}$ . Nous notons  $sn_c^* = f_c(B_{Erc}, J_{gaba_c})$ .

Nous partons de  $sn_c^*$  évalué en  $B_{Erc} = \widehat{B_{Erc}}$  et  $J_{gaba_c} = \widehat{J_{gaba_c}}$ , i.e.  $sn_c^* = f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}})$ . Nous faisons varier simultanément  $B_{Erc}$  de  $\Delta \widehat{B_{Erc}}$  et  $J_{gaba_c}$  de  $\Delta \widehat{J_{gaba_c}}$ .  $\Delta sn_c$  désigne la variation totale de  $sn_c$  dans la cible due à la variation simultanée du forçage  $B_{Erc}$  et du couplage  $J_{gaba_c}$ . Par définition, nous avons donc :

$$\Delta sn_c \equiv f_c(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.122)$$

Nous pouvons également construire  $\Delta sn_c$  en intégrant les sensibilités par rapport à  $B_{Erc}$  et  $J_{gaba_c}$ , de deux manières différentes :

#### 18.4.2.1 Première procédure :

1. Nous calculons la variation de l'activité de la cible  $sn_c^*$  due à la stimulation (i.e. à la variation du forçage  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}$ ), pour  $\widehat{J_{gaba_c}}$  fixé :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}} S_{sn_{c/B_{Erc}}}(B_{Erc}, \widehat{J_{gaba_c}}) dB_{Erc} \quad (18.123)$$

D'autre part,  $\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}})$  est égal par définition à la différence entre points fixes quand on a fait varier  $B_{Erc}$ , évalués en  $\widehat{J_{gaba_c}}$  fixé :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) \equiv f_c(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}, \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.124)$$

2. Nous calculons la variation de l'activité de la cible  $sn_c^*$  due à la variation de  $J_{gaba_c}$  entre  $\widehat{J_{gaba_c}}$  et  $\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$ , évaluée en période de stimulation i.e.  $B_{Erc} = \widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$  :

$$\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}) = \int_{\widehat{J_{gaba_c}}}^{\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}} S_{sn_c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, J_{gaba_c}) dJ_{gaba_c} \quad (18.125)$$

$$\equiv f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.126)$$

3. Nous sommions pour retrouver la variation totale d'activité due à la variation des deux paramètres :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) + \Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}) \quad (18.127)$$

$$= f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) + \quad (18.128)$$

$$f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.129)$$

$$= f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.130)$$

$$= \Delta sn_c \quad (18.131)$$

#### 18.4.2.2 Deuxième procédure :

1. Nous calculons la variation de l'activité de la cible  $sn_c^*$  due à la variation de  $J_{gaba_c}$  entre  $\widehat{J_{gaba_c}}$  et  $\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$ , évaluée hors stimulation i.e.  $B_{Erc} = \widehat{B_{Erc}}$  :

$$\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}}) = \int_{\widehat{J_{gaba_c}}}^{\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}} S_{sn_c/J_{gaba_c}}(\widehat{B_{Erc}}, J_{gaba_c}) dJ_{gaba_c} \quad (18.132)$$

$$\equiv f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.133)$$

2. Nous calculons la variation de l'activité de la cible  $sn_c^*$  due à la stimulation (i.e. variation du forçage  $B_{Erc}$  entre  $\widehat{B_{Erc}}$  et  $\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}$ ), pour  $J_{gaba_c} = \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$  fixé :

$$\Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) = \int_{\widehat{B_{Erc}}}^{\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}} S_{sn_c/B_{Erc}}(B_{Erc}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) dB_{Erc} \quad (18.134)$$

$$\equiv f_c(\widehat{B_{Erc}} + \Delta\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) \quad (18.135)$$

3. Nous sommes pour retrouver la variation totale d'activité due à la variation des deux paramètres :

$$\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}}) + \Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) \quad (18.136)$$

$$= f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.137)$$

$$+ f_c(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) \quad (18.138)$$

$$= f_c(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}, \widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) - f_c(\widehat{B_{Erc}}, \widehat{J_{gaba_c}}) \quad (18.139)$$

$$= \Delta sn_c \quad (18.140)$$

### 18.4.2.3 Conclusion

Nous avons montré que :

$$\begin{cases} \Delta sn_c = \Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}}) + \Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}) \\ \quad = \Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}}) + \Delta sn_{c/B_{Erc}}(\widehat{J_{gaba_c}} + \Delta \widehat{J_{gaba_c}}) \end{cases} \quad (18.141)$$

Le signe de  $\Delta sn_c^*$  dépend de l'architecture de connectivité. En effet, il est possible d'évaluer séparément les signes des deux contributions à  $\Delta sn_c^*$  :

1.  $\Delta sn_{c/J_{gaba_c}} < 0$  pour n'importe quel type de connectivité (cf annexe 18.3). Plus  $\Delta J_{gaba_c}$  augmente et plus  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}}) < 0$  ou  $\Delta sn_{c/J_{gaba_c}}(\widehat{B_{Erc}} + \Delta \widehat{B_{Erc}}) < 0$  devient grande en valeur absolue.
2.  $\Delta sn_{c/B_{Erc}} < 0$  pour une connectivité E-I et I-I.  $\Delta sn_{c/B_{Erc}} > 0$  dans le cas E-E et I-E (cf article chapitre 9)

Dans le cas E-E et I-E, si  $\Delta sn_c^*$  est positif, l'augmentation de  $J_{gaba_c}$  produit une moindre augmentation de l'activité de la cible. Plus  $J_{gaba_c}$  augmente et plus  $\Delta sn_c^* > 0$  diminue en valeur absolue. A priori un changement de signe peut se produire.

#### 18.4.2.4 Une autre méthode pour construire $\Delta sn_{c/B_{E_{rc}}}$ en fonction de $J_{gaba_c}$

Pour toute variation fixée du forçage  $B_{E_{rc}}$  entre  $\widehat{B_{E_{rc}}}$  et  $\widehat{B_{E_{rc}}} + \Delta\widehat{B_{E_{rc}}}$ , du système 18.141 nous déduisons :

$$\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) = \Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}}) + \Delta sn_{c/J_{gaba_c}}(\widehat{B_{E_{rc}}} + \Delta\widehat{B_{E_{rc}}}) - \Delta sn_{c/J_{gaba_c}}(\widehat{B_{E_{rc}}}) \quad (18.142)$$

Nous pouvons alors utiliser les variations par rapport à  $J_{gaba_c}$  (i.e. la fonction  $\Delta sn_{c/J_{gaba_c}}$ ) pour reconstruire  $\Delta sn_{c/B_{E_{rc}}}$  en fonction de  $J_{gaba_c}$ .

Pour cela, nous calculons une seule fois  $\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}})$ . Nous faisons varier  $\Delta\widehat{J_{gaba_c}}$ . Pour chaque valeur de  $\Delta\widehat{J_{gaba_c}}$ , les variations par rapport à  $J_{gaba_c}$  peuvent être calculées en intégrant les sensibilités par rapport à  $J_{gaba_c}$  (pour les deux valeurs du forçage  $\widehat{B_{E_{rc}}}$  et  $\widehat{B_{E_{rc}}} + \Delta\widehat{B_{E_{rc}}}$ ). Nous pouvons alors tracer la fonction  $\Delta sn_{c/B_{E_{rc}}}$  pour différentes valeurs de  $J_{gaba_c}$ <sup>3</sup>. Dans le cas de la connectivité EI par exemple, nous retrouvons bien la courbe en "U" de l'article au chapitre 9.

Cette méthode pour construire  $\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}})$  en fonction de  $J_{gaba_c}$  vient s'ajouter aux deux premières qui ont été présentées à la section 18.2. Pour obtenir chaque point de la courbe qui donne  $\Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}})$  en fonction de  $J_{gaba_c}$ , nous avons vu que nous pouvons procéder par :

- soustraction de points fixes  $sn_c^*$  en période de stimulation et au repos, pour une valeur de  $J_{gaba_c}$  fixée
- intégration des sensibilités de  $sn_c^*$  par rapport à  $B_{E_{rc}}$  entre  $\widehat{B_{E_{rc}}}$  et  $\widehat{B_{E_{rc}}} + \Delta\widehat{B_{E_{rc}}}$ , pour une valeur de  $J_{gaba_c}$  fixée.

---

3. De manière équivalente, il serait possible d'utiliser les variations par rapport à  $B_{E_{rc}}$  (i.e. la fonction  $\Delta sn_{c/B_{E_{rc}}}$ ) pour reconstruire  $\Delta sn_{c/J_{gaba_c}}$  en fonction de  $B_{E_{rc}}$ . Pour toute variation fixée de  $J_{gaba_c}$  (entre  $\widehat{J_{gaba_c}}$  et  $\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}$ ) :

$$\Delta sn_{c/J_{gaba_c}}(\widehat{B_{E_{rc}}} + \Delta\widehat{B_{E_{rc}}}) = \Delta sn_{c/J_{gaba_c}}(\widehat{B_{E_{rc}}}) + \Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}} + \Delta\widehat{J_{gaba_c}}) - \Delta sn_{c/B_{E_{rc}}}(\widehat{J_{gaba_c}}) \quad (18.143)$$



## Etude de la balance Excitation/Inhibition de régions cérébrales impliquées dans une tâche de contrôle inhibiteur : modélisation de données obtenues en Imagerie par Résonance Magnétique fonctionnelle et inversion

En psychologie, le contrôle inhibiteur est un mécanisme cognitif qui permet de stopper une réponse motrice, émotionnelle ou cognitive non adaptée pour la réalisation d'un but désiré. Au niveau cérébral, le contrôle inhibiteur est associé au fonctionnement en réseau de régions cérébrales, qui peut être mesuré à partir du signal BOLD en IRMf. Des régions de contrôle préfrontales abaissent l'activité BOLD de régions cibles. L'IRMf permet de mesurer de manière indirecte l'activité des neurones. Comment peut-on alors inférer à partir de données d'IRMf des propriétés excitatrices et inhibitrices (E/I) neurales au sein de régions cérébrales impliquées dans une tâche de contrôle inhibiteur ?

Nous partons d'un modèle biophysique non linéaire, hiérarchique qui décrit les évolutions temporelles des activités neurales excitatrice et inhibitrice par région ([Naskar et al., 2021]). Ces variations d'activité produisent des changements BOLD dans chaque région cérébrale. L'analyse de ce modèle nous permet de : 1) identifier des paramètres neuraux de la balance E/I ; 2) montrer que l'augmentation d'activité BOLD d'une région de contrôle ne permet pas d'abaisser l'activité BOLD d'une région cible parce que les régions sont connectées par leurs neurones excitateurs uniquement ; 3) proposer une nouvelle architecture de connectivité pour le permettre ; 4) étudier comment la baisse d'activité de la région cible dépend de la balance E/I dans la cible. Nous proposons alors une nouvelle procédure d'inversion. Nous en vérifions la fiabilité avec des simulations, avant de présenter une preuve de concept sur les données d'un sujet pendant une tâche de Think/NoThink, un paradigme d'étude du contrôle inhibiteur des intrusions mnésiques ([Mary et al., 2020]).

**Mots clés** : contrôle inhibiteur, balance locale Excitation/Inhibition, IRMf, intégration fonctionnelle en neurosciences, modèle en masses neurales, dynamique non linéaire multi-échelles, étude de sensibilités, inversion.

In psychology, inhibitory control is a cognitive mechanism that stops a motor, emotional or cognitive response from achieving a desired goal. At cerebral level, inhibitory control is associated with a network of brain regions, whose function may be measured using BOLD signals from fMRI. Prefrontal control regions lower the BOLD activity of target regions. fMRI provides an indirect measure of the activity of neurons. How can we then infer from fMRI data, neural excitatory and inhibitory (E/I) properties of brain regions involved in an inhibitory control task ?

We start with a non-linear biophysical model that describes by region the temporal evolution of neural excitatory and inhibitory activities ([Naskar et al., 2021]). These variations in activity produce BOLD changes in each brain region. Analysis of this model enables us to : 1) identify neural parameters of the E/I balance ; 2) show that increasing the BOLD activity of a control region does not lower the BOLD activity of a target region, since these regions are connected by their excitatory neurons only ; 3) propose a new connectivity architecture to enable this ; 4) study how the lowering of activity in the target region depends on the E/I balance in the target region. We then propose a new inversion procedure. We check its reliability through simulations, before presenting a proof-of-concept using real data from a subject during a Think/No-Think task, a paradigm used for studying the inhibitory control of memory intrusions ([Mary et al., 2020]).

**Keywords** : inhibitory control, local Excitation/Inhibition balance, fMRI, functional integration in neuroscience, neural mass model, multi-scale nonlinear dynamics, sensitivity study, inversion.