

III

UN MODÈLE DE DÉVELOPPEMENT ÉPIGÉNÉTIQUE DU SYSTÈME NERVEUX : RÔLE POSSIBLE DANS LA PRÉDISPOSITION AU LANGAGE

« Je dessinais distraitement des signes ennuyeux, des lettres qu'on grave sur la pierre, et aussi les abréviations correspondantes du style ordinaire. Et soudain, une parole oubliée d'Oné ou un phénomène inexplicable en moi rendit vivants les mots et les caractères. De l'image sort un mot, du mot une syllabe, de la syllabe une lettre. En associant les lettres des images on formait des mots nouveaux, étranges, qui n'avaient rien de commun avec les images. L'arroseur le plus obtus peut comprendre une image, mais seul un homme sachant lire peut déchiffrer deux images conjuguées. Je crois que tous ceux qui ont étudié l'écriture et appris à lire comprendront l'événement dont je parle. Ce fut pour moi une véritable aventure, plus passionnante et plus captivante qu'une grenade dérobée à l'étalage d'un marchand, plus douce qu'une datte sèche, délicieuse comme l'eau pour un assoiffé. »

Mika Waltari, *Sinouhé l'Égyptien*.

« 'How is the Dictionary getting on?' said Winston, raising his voice to overcome the noise.

'Slowly,' said Syme. 'I'm on the adjectives. It's fascinating. [...] It's a beautiful thing, the destruction of words. Of course the great wastage is in the verbs and adjectives, but there are hundreds of nouns that can be got rid of as well. It isn't only the synonyms; there are also the antonyms. After all, what justification is there for a word which is simply the opposite of some other word? A word contains its opposite in itself. Take "good", for instance. If you have a word like "good", what need is there for a word like "bad"? "Ungood" will do just as well—better, because it's an exact opposite, which the other is not. Or again, if you want a stronger version of "good", what sense is there in having a whole string of vague useless words like "excellent" and "splendid" and all the rest of them? "Plusgood" covers the meaning; or "doubleplusgood" if you want something stronger still. Of course we use those forms already, but in the final version of Newspeak there'll be nothing else. In the end the whole notion of goodness and badness will be covered by only six words—in reality, only one word. Don't you see the beauty of that, Winston? It was Big Brother's idea originally, of course,' he added as an afterthought. »

George Orwell, *Nineteen Eighty-Four*.

1 CONNEXIONNISME ET REPRÉSENTATION NEURALE

Il est communément admis que le cerveau code l'information au moyen des *niveaux d'activité* de cellules ou d'"assemblées de cellules" (Hebb, 1949), ce schéma de codage ayant inspiré la quasi-totalité des modèles théoriques du fonctionnement du cerveau, appelés aujourd'hui modèles *connexionnistes*. En ce sens, comme le font remarquer Fodor et Pylyshyn (1988), les modèles connexionnistes sont clairement concernés par la question de la représentation mentale, car ils supposent l'existence d'"états" de l'esprit codant les états du monde extérieur (modèles "représentationalistes" par opposition aux "éliminativistes", selon les termes de ces auteurs). Tous les chercheurs s'accordent en effet à penser qu'il existe une projection du monde extérieur sur l'ensemble physique des unités de traitement neuronales : le problème reste cependant de préciser la nature de ces états de représentation, cette question faisant tout l'objet de la présente discussion.

Partant de l'hypothèse d'un codage en niveaux d'activité, les neurophysiologistes ont jusqu'à présent principalement concentré leurs efforts sur l'étude des fréquences de décharge pouvant être évoquées par des stimuli variés, ou plus généralement enregistrées en conjonction avec certains comportements du sujet. Depuis longtemps en effet, et en particulier depuis les remarquables expériences de Hubel et Wiesel sur le cortex visuel, il est reconnu qu'un "stimulus élémentaire" ou une "action élémentaire" est en général codé par les niveaux d'activité des neurones des aires sensorielles et motrices, et que ce code est donc utilisé très largement dans les parties périphériques de notre cerveau. Cependant, rien n'indique que ce schéma doive être généralisé aux fonctions cognitives supérieures. Certaines expériences tentent aujourd'hui de mettre en évidence l'importance de l'activité neuronale individuelle dans diverses situations comportementales impliquant l'utilisation de connaissances structurées, comme la reconnaissance de stimuli complexes (visages, mains, ou mouvements) (Perrett *et al.*, 1982) : il est alors dit que la cellule *code* de façon sélective une tête ou une main connue, par exemple, et permet de ce fait une perception invariante par rotation. Ainsi, la plupart des auteurs pensent que le cerveau utilise essentiellement le même code "à tous les niveaux" : percepts, concepts, images mentales, linguistique, sont donc tous représentés par des taux d'activité neuronale (voir Barlow, 1972).

Cependant, un certain nombre de psychologues soulignent l'inaptitude des modèles connexionnistes à expliquer les fonctions cognitives supérieures, et tout particulièrement le langage. Dans leur critique *Connectionism and Cognitive Architecture*, Fodor et Pylyshyn (1988) font remarquer que les niveaux d'activation neuronaux sont dépourvus de structure et ne peuvent donc pas être "*composés*"

(c'est à dire assemblés) entre eux de façon non-triviale. Or la propriété de compositionnalité est sans doute la question fondamentale des sciences cognitives en général, et du langage en particulier. Les facultés langagières démontrent en effet clairement notre capacité à produire un nombre infini de symboles complexes à partir de symboles plus élémentaires (Chomsky, 1957; Fodor, 1975). En ce sens, le langage est un véritable jeu de construction, dans lequel des "briques élémentaires" peuvent être assemblées de différentes façon pour former des objets composites, qui seront à leur tour les constituants d'objets plus vastes. Par ailleurs, ces symboles linguistiques reçoivent également un *sens*, c'est à dire une interprétation cohérente, qui est une émergence de leur structure composée : il existe donc une *syntaxe et une sémantique combinatoires*. La notion de "constituance" est ainsi une notion centrale de psychologie cognitive : la signification d'un symbole est déterminée par ses constituants ainsi que par la façon dont ces constituants sont assemblés.

Le problème est donc que les configurations distribuées d'activités moyennes sont de simples *listes dépourvues de structure* : il est impossible de les composer sans les mélanger et perdre de ce fait l'intégrité de chaque partie ("catastrophe de superposition" soulignée par von der Malsburg, 1986). Ceci peut être illustré par l'exemple schématique suivant : si on suppose qu'une cellule ou une assemblée de cellules *T* code spécifiquement l'apparition d'un "triangle" dans le champ visuel, et que, de même, un "cercle" est codé par un ensemble d'activités *C*, si on suppose par ailleurs que les qualités "rouge" et "vert" sont codées respectivement par des assemblées *R* et *V*, alors, dans le schéma des assemblées de cellules, un "triangle vert" sera représenté par l'activation conjointe *TV*, et un "cercle rouge" par *CR*, et l'apparition simultanée du "triangle vert" et du "cercle rouge" provoquera l'activation des quatre assemblées de cellules *TCVR* : le problème vient alors de ce qu'il n'est plus possible de retrouver les paires initiales, car cette activation collective code également l'apparition simultanée d'un "triangle rouge" *TR* et d'un "cercle vert" *CV*.

On voit donc qu'une simple conjonction de propriétés ne peut suffire à rendre compte de leur association, et les représentations neurales doivent donc nécessairement être de type *relationnel*. Ainsi, le problème précédent sera résolu en codant également les relations entre caractéristiques : la conjonction *T—V, C—R* ne pourra pas être confondue avec la conjonction *T—R, C—V*.

Nous proposons alors que, dans le cortex, la structure interne des représentations neurales, dont on vient de montrer la nécessité, soit fondée sur une *structure temporelle fine* des états d'activités neuronales. Au-delà de l'échelle de temps de la cognition, ou échelle de temps "psychologique", qui est apparente au niveau du comportement, et qui est de l'ordre de la seconde ou de la fraction de seconde, il existe de toute évidence une échelle de temps "physiologique" beaucoup plus fine, qui est typiquement celle des potentiels d'action, c'est à dire des processus d'interaction rapides entre neurones, et qui est de l'ordre de la milliseconde

(propagation d'un potentiel d'action le long de l'axone; libération du neuro-transmetteur et liaison avec le récepteur provoquant la production d'un potentiel postsynaptique; propagation de ce dernier sur les dendrites; intégration somato-dendritique). Il existe plusieurs ordres de grandeur entre ces deux échelles de temps, et le système nerveux est donc parfaitement équipé pour exploiter et traiter des événements définis de façon extrêmement précise à la fois dans le temps et dans l'espace (des dizaines de milliers de terminaisons peuvent être produites par un même axone ou, au contraire, peuvent converger, en provenance d'axones différents, sur un neurone donné).

Cette hypothèse a été formulée par plusieurs auteurs, et on peut la retrouver en particulier chez Sejnowski (1981). Mais cette idée a surtout été développée par von der Malsburg, dans *The Correlation Theory of Brain Function* (1981) : sa "théorie des corrélations" propose une solution neurobiologique plausible au problème de la représentation neurale, dans le domaine général de la perception. Les arguments qui y sont avancés peuvent être résumés de la façon suivante : les objets sont représentés par un ensemble de relations entre leurs parties constitutives, c'est à dire en général par des *graphes*, et, pour assembler, segmenter, ou comparer ces objets, le cerveau dispose de "liens dynamiques", qui sont essentiellement des *niveaux de corrélations* entre activités neuronales. Ces corrélations sont soutenues par des poids synaptiques doués de *plasticité rapide*, c'est à dire obéissant à une dynamique hebbienne rapide et réversible. Ainsi, moments statistiques d'ordre élevé et modulations synaptiques rapides offrent un format naturel de représentation structurée (voir le chapitre 1 de la partie II).

La théorie des corrélations a conduit en particulier à la prédiction suivante (von der Malsburg et Bienenstock, 1986; p270) :

« Si l'on enregistrerait, chez un animal entraîné, plusieurs cellules sensorielles, activées soit par un stimulus unique perçu comme une entité (condition a), soit par plusieurs stimuli présentés simultanément mais clairement séparés et indépendants (condition b), alors on devrait trouver, dans la structure temporelle fine de cet enregistrement multiple, plus de coïncidences dans le cas a) que dans le cas b). »

Par la suite, des expériences récentes dans le cortex visuel du chat anesthésié (Gray *et al.*, 1989) ont apporté des éléments de confirmation à cette prédiction, et ont révélé une activité oscillatoire dépendant de la stimulation, qui servirait d'"onde porteuse" aux corrélations. Remarquons cependant que les oscillations (dont la réalité et la signification sont encore débattues) ne sont pas indispensables au code par corrélations, qui s'accommoderait aussi bien de processus de décharge irréguliers.

Ainsi, nous pensons que le substrat physiologique de représentation des symboles, permettant de rendre compte de leur structure combinatoire interne et de leur compositionnalité, doit être constitué par des *configurations spatio-temporelles*

précises d'activités. Ce point de vue peut être abordé globalement à travers le raisonnement suivant : pour combler la distance qui sépare les deux échelles de temps "psychologique" et "physiologique" dont il a été question plus haut, il semble naturel d'utiliser de façon générale un opérateur de *moyenne temporelle*, que l'on pourra noter $E_T[\dots]$, où T définit une fenêtre d'intégration de l'ordre d'une centaine de millisecondes. La propriété d'un tel opérateur est d'éliminer l'aspect stochastique et variable d'un ensemble complexe de processus de décharge instantanés, dont il permet d'extraire certaines caractéristiques plus durables : la configuration détaillée des activités neuronales ne sera jamais tout à fait identique d'une perception à l'autre d'un même objet, et il est raisonnable de supposer que E_T soit l'expression des "propriétés invariantes" de la représentation neurale de l'objet en question.

Le problème est alors de savoir sur quelles variables cette moyenne temporelle devra agir. Etant donné un ensemble d'états d'activité instantanés $\{x_i(t)\}$ (où x_i est une variable binaire, $x_i(t)=1$ faisant référence à la décharge d'un potentiel d'action par le neurone i dans un intervalle de $\sim 3\text{ms}$ entre t et $t+1$), le code traditionnel fondé sur des "taux d'activité", que nous avons évoqué plus haut, revient à utiliser E_T de la façon la plus simple qui soit, en définissant des états moyens individuels $\{\underline{x}_i\}$ par : $\underline{x}_i = E_T[x_i(t)]$. Ainsi, dans les "assemblées de cellules", les moyennes sont calculées séparément sur chaque composante : ce format de représentation est donc constitué de moments statistiques du premier ordre, et nous le qualifierons de "*code d'ordre 1*".

Rien n'empêche alors d'envisager une combinaison plus complexe des variables temporelles $\{x_i\}$: c'est précisément ce que propose la théorie des corrélations, dont l'intérêt porte sur les quantités du type $E_T[x_i(t)x_j(t)]$, ou plus généralement $E_T[x_i(t)x_j(t-\tau)]$ (τ représentant un délai pouvant varier entre 0 et plusieurs dizaines de millisecondes), et que nous pourrions appeler "*code d'ordre 2*". Dans ce cadre, les activités neuronales ne sont plus traitées indépendamment les unes des autres, mais on s'intéresse au contraire à leur variation solidaire. Partant, il est alors possible de généraliser l'idée d'une interaction non-triviale entre activités à un ordre plus élevé que 2, et s'intéresser à des "*événements d'ordre n*", c'est à dire des quantités du type : $E_T[x_{i_1}(t)x_{i_2}(t-\tau_2)\dots x_{i_n}(t-\tau_n)]$. Il est clair que les codes d'ordre élevé, distribués avec précision à la fois dans le temps et dans l'espace, offrent potentiellement d'immenses ressources pour la représentation relationnelle des connaissances.

En résumé, nous proposons que le substrat physique de représentation des symboles soit constitué de *configurations spatio-temporelles* (CST) précises, c'est à dire d'états d'activités caractérisés par de nombreuses corrélations entre neurones, et soutenus par une focalisation particulière des connexions. Ces états constituent donc les codes d'"ordre élevé". Les preuves expérimentales directes de l'existence de tels états font bien sûr défaut à l'heure actuelle, cependant on peut trouver des éléments de confirmation dans certaines études des aires corticales du chat ou du singe

(l'animal étant anesthésié ou en comportement) utilisant des enregistrements cellulaires multiples. En particulier, Abeles (1982; Abeles et Gerstein, 1988), l'un des chercheurs qui se sont le plus intéressés à la question des corrélations, a observé l'apparition répétitive de motifs temporels impliquant typiquement 3 à 6 neurones, avec une précision de l'ordre de la milliseconde, et des délais pouvant atteindre quelques centaines de millisecondes. De tels motifs peuvent par exemple se reproduire une douzaine de fois dans une période d'enregistrement de quelques minutes, ce qui est beaucoup plus qu'une fréquence simplement due au hasard (cette affirmation s'appuie sur une comparaison avec un calcul théorique prenant comme hypothèse des processus de décharge indépendants). Par ses expériences, Abeles a donc montré que des événements spatio-temporels reproductibles existaient dans le cortex. Ceci prouve que le cerveau dispose de circuits neuronaux capables de soutenir des événements précis à la fois dans l'espace et dans le temps, mais bien sûr ceci ne prouve pas que ces événements constituent un *code* utilisé par le cerveau (le lien entre ces événements et une éventuelle "signification" est donc loin d'être établi). Cependant, nous adopterons cette idée comme hypothèse de travail, car elle nous semble la plus plausible et la plus naturelle.

Dans son livre *Local Cortical Circuits* (1982), Abeles introduit le terme de "*synfire chains*" pour décrire les structures de connexions qui seraient susceptibles d'expliquer ses observations : ces architectures sont semblables à des "réseaux à couches" dans lesquels tous les neurones d'une couche sont synchrones (voir la figure 5, chapitre 4). Dans cette étude, nous nous concentrerons surtout sur le problème de la *croissance épigénétique* de ces chaînes de connexions servant de support aux CST. Comme nous l'expliquerons plus en détail au chapitre 2, nous supposons que le réseau cortical, initialement caractérisé par une innervation large et diffuse, se *structure* progressivement dans une étape de développement qui peut être en partie antérieure à l'apparition explicite du comportement langagier chez l'enfant (et qui peut bien sûr être en recouvrement avec les premières années où ce comportement s'exprime). Le modèle de développement épigénétique que nous proposons sera exposé au chapitre 3, puis nous mettrons en évidence les principaux mécanismes de son fonctionnement au chapitre 4, ainsi qu'au chapitre 5 à travers quelques simulations numériques. L'utilisation de ce code neural, à base de CST et de chaînes de connexions, dans la perspective de la *composition* des symboles sera ensuite abordée de façon rapide au chapitre 6.

2 LE DÉVELOPPEMENT ÉPIGÉNÉTIQUE DU SYSTÈME NERVEUX

Pour résumer les enjeux théoriques du chapitre 1, le modèle que nous proposons (qui sera exposé au chapitre 3), s'inscrit dans le projet général d'une *neurobiologie de la cognition*, et a pour but de rendre compte de l'existence d'une représentation mentale *structurée*, c'est à dire de la manipulation des symboles par le cerveau. En effet, comme nous l'avons souligné à partir de la critique de Fodor et Pylyshyn (1988), la production et la compréhension du langage doivent être modélisées par des *systèmes symboliques*, ce que les théories connexionnistes classiques ont du mal à réaliser au moyen d'un codage distribué purement spatialement. Ainsi, contrairement aux tenants du "subsymbolisme" (Rumelhart *et al.*, 1986; Smolensky 1988), qui cherchent à combler les lacunes du formalisme de l'IA par une microstructure de la cognition de type *statistique* (dont les limitations théoriques ont été étudiés dans la première partie de cette thèse), nous pensons que, à tous les niveaux de son architecture cognitive, le cerveau est essentiellement une "machine à manipuler des symboles" (que cette manipulation soit consciente ou non, ses règles explicites ou non), et que ces symboles ont une *structure combinatoire interne* qui permet leur *composition* avec d'autres symboles. Comme on l'a montré plus haut, cette nécessité d'une représentation *relationnelle* trouve son expression naturelle dans la *structure temporelle* fine de l'activité neuronale : ainsi, von der Malsburg (1981) a été l'un des premiers à suggérer un codage par les *corrélations* des décharges, c'est à dire des événements d'activité d'ordre élevé. Pour soutenir ces événements, et donc coder les symboles, le cerveau doit alors disposer d'une organisation de connexions synaptiques appropriée : autrement dit, une structuration particulière du cortex est nécessaire pour servir de support à tout l'édifice hiérarchique de la cognition et des symboles mentaux qui y sont manipulés. Dans ce cadre, notre modèle tentera de cerner les mécanismes qui peuvent être à l'origine de la mise en place des "briques élémentaires" de cet édifice que sont les CST, c'est à dire les configurations de connexions associées à des propagations régulières d'activations. Les CST constituent donc un format de "*préreprésentation*" selon le terme de Changeux (1983), mais ce format intègre également une dimension temporelle en plus de l'habituelle distribution spatiale.

Dans cette perspective, il est difficile de concevoir qu'une telle préreprésentation des symboles ne se mette en place qu'à partir du moment où apparaissent effectivement les fonctions linguistiques. Au contraire, il semble nécessaire de faire l'hypothèse que ce substrat apparaît *antérieurement* à l'expression de cette capacité cognitive, par des mécanismes qui pourraient être le prolongement de ceux qui président à l'établissement des grandes fonctions du cerveau, telles

qu'elles ont pu être étudiées dans le domaine de la psychophysiologie de la perception. En effet, comme on va le voir plus loin, il a été démontré de façon très claire que le système nerveux, au cours de son développement ontogénétique précoce, *se prépare* à son interaction future avec son environnement par la mise en place de propriétés de sélectivité fonctionnelle et par des connexions spécifiques synaptiques qui sous-tendent cette sélectivité. Il est donc naturel de penser que dans un domaine tel que celui du langage, qui fait appel à des processus cognitifs encore plus complexes que dans le domaine de la perception, cette phase préparatoire de mise en place des substrats physiologiques soit au moins aussi importante.

Cette idée d'un développement précoce de l'"organe" du langage, un peu à la manière du développement d'organes physiques tels que bras ou jambes, est défendue par des auteurs tels que Chomsky (voir à ce propos Piattelli-Palmarini, 1978). Selon Chomsky, ce développement se ferait selon un programme bien défini et relativement rigide : ainsi, l'interaction avec l'environnement jouerait un rôle bien moins important qu'on le croit d'habitude dans l'acquisition du langage. Dans cette vue, il conviendrait donc de parler de "morphogenèse" du langage plutôt que d'acquisition. L'exposition à une langue particulière servirait alors uniquement à "révéler" (au sens photographique du terme) le programme interne de développement et à lui permettre de s'exprimer, ainsi qu'à régler un certain nombre de paramètres caractéristiques d'une ou plusieurs langues particulières. Sans défendre aucunement les modèles chomskiens du "langage interne" (grammaires génératives), nous retiendrons ici l'hypothèse très raisonnable d'une continuité entre les mécanismes généraux du développement à l'œuvre dans le cerveau et en particulier dans les systèmes sensoriels, et ceux présidant à l'établissement des structures sous-jacentes au langage. Au-delà du caractère artificiel des langages chomskiens, il est donc important de s'interroger davantage sur l'idée d'une *prédisposition ontogénétique* au langage.

Nous nous proposons donc d'élaborer un modèle de développement ontogénétique dans lequel on démontrera l'émergence spontanée de CST structurées possédant des propriétés requises pour permettre au système dans son fonctionnement ultérieur de réaliser des opérations de composition telles qu'elles existent de façon évidente dans les facultés langagières. Ce modèle s'inspirera de façon précise des modèles de *développement épigénétique* des systèmes sensoriels et en particulier du système visuel. Il constituera en fait une continuation de ces modèles, que l'on obtiendra en modifiant les conditions aux limites et les règles précises de modification synaptique. La notion centrale dans cette recherche théorique est donc celle d'*algorithme épigénétique*, que l'on doit concevoir ici au sens large du terme, c'est à dire non pas seulement réservé aux processus qui révèlent la plasticité du système à travers une interaction avec l'environnement ou une manipulation par cet environnement, mais appliqué en général au cas où l'on démontre que le développement du système est déterminé autant par son

fonctionnement propre que par un ensemble d'instructions génétiques. Ainsi que le souligne Prochiantz (1985), dans son livre *La construction du cerveau* (p73) :

« [Il s'agira bientôt de comprendre] le rôle de l'activité nerveuse dans la constitution épigénétique de la structure cérébrale. En effet le nombre des neurones, la forme des réseaux, le nombre et la distribution des synapses reflètent, à l'intérieur d'une enveloppe génétiquement déterminée, l'histoire de l'individu et de l'activité de son système nerveux. On peut donc dire que, dans ses détails, le développement cérébral n'obéit à aucun automatisme, aucune fatalité (sinon celle de l'environnement sensoriel, affectif ou culturel). C'est donc une embryologie fonctionnelle qu'il s'agira de mettre en place et de relier avec les données des sciences du comportement. »

Par conséquent, on peut parler ici d'*auto-organisation* : au lieu de se donner comme contraintes a priori l'existence de deux structures organisées de façon topologique (par exemple la rétine et le tectum) qui conduisent à l'apparition d'une projection topographique, on considérera un substrat au départ inorganisé, c'est à dire un *graphe aléatoire*. Par ailleurs, on se donnera comme règle de modification une version de la règle de Hebb requérant une synchronie précise de la convergence des décharges, à travers un réseau de délais de transmission axono-somatiques cohérents. Dans ces conditions, on verra apparaître des CST qui possèdent les propriétés requises. Ainsi, de la même façon que la mise en place d'une projection topographique de la rétine sur le tectum ou sur le cortex prépare le système nerveux à son interaction avec un environnement visuel structuré en accord avec la géométrie euclidienne tridimensionnelle, la mise en place de ces CST composables avant même l'apparition du langage permettrait à celui-ci de s'exprimer dès l'instant où l'enfant atteint un degré de maturation nécessaire.

C'est également dans une perspective *phylogénétique* que l'on peut envisager ce processus comme une continuation du développement ontogénétique des systèmes sensoriels. L'une des théories modernes de l'évolution propose en effet que les relations entre ontogenèse et phylogenèse (en particulier pour ce qui concerne les mécanismes de spéciation au sommet de l'échelle phylogénétique) soient expliquées par les *hétérochronies*, c'est à dire les modifications survenues dans les séquences de développement (ou "trajectoires ontogéniques") qui caractérisent les espèces et qui sont reproduites par tous les individus (de Beer, 1958; Gould, 1977). Il semble que ces modifications soient décomposables selon quelques paramètres temporels simples affectant le développement des caractères phénotypiques : signal de début de croissance, signal de fin de croissance, et vitesse de croissance. Tout changement, même faible, de l'un de ces paramètres peut alors avoir de grandes conséquences sur la structure ou les proportions de l'organisme, ce qui expliquerait les changements morphologiques massifs observables sur les descendants par rapport aux ancêtres. La théorie des hétérochronies fournit de la sorte un schéma unitaire pour la variation simultanée de nombreux caractères non

fonctionnellement corrélés. Ainsi, un début avancé, une fin retardée ou une croissance accélérée auront pour effet l'exagération d'un caractère morphologique déjà présent, ou l'apparition d'un nouveau caractère (*péramorphose*), tandis qu'un début retardé, une fin avancée ou une croissance ralentie auront un effet contraire, c'est à dire aboutiront à une version "inachevée" (*pædomorphose*). En particulier, on soutient l'hypothèse selon laquelle l'homme serait une version *pædomorphique* du singe, ce qui est fortement suggéré par la comparaison d'un crâne humain adulte avec celui d'un jeune chimpanzé (Gould, 1977), et que cette "juvénalisation" phylogénétique serait due à une *néoténie* (terme qui désigne la diminution du taux de croissance). Par conséquent, les causes des profondes différences phénotypiques entre l'homme et le singe, dont les génotypes sont très voisins, devraient surtout être recherchées dans la modulation des *gènes régulateurs* de la dynamique temporelle du développement.

Dans cette optique, nous nous intéressons donc tout particulièrement aux conséquences que pourrait avoir le développement néoténique sur le cortex et la facultés langagières. Ainsi, nous proposons que l'émergence du langage puisse être liée à une faible modification de certains gènes de développement ayant entraîné un *retard* dans le processus d'organisation synaptique, et permis de la sorte un *niveau résiduel de plasticité* dans l'état final. En un mot : on peut dire que par la modification de l'ontogenèse, la phylogenèse a préparé le terrain d'une *logogenèse* (von der Malsburg, 1987). Comme d'ailleurs Prochiantz le souligne :

« Une des caractéristiques majeures du système nerveux réside sans aucun doute dans sa plasticité. [...] Les modifications du paysage neuronal marquent notre adaptation, nos capacités d'apprentissage et de perfectionnement qui se maintiennent jusqu'à un âge avancé de la vie, en fait, jusqu'à la mort. [...] Les mécanismes qui président à cette plasticité chez l'adulte sont, pour une partie, semblables à ceux que l'on trouve à l'œuvre dans le développement embryonnaire. En particulier certains facteurs de croissance, qui assurent la maturation des neurones restent également présents dans l'organisme adulte ou ils sont impliqués dans le maintien et la plasticité du système nerveux. »

Notre modèle s'inscrit parfaitement dans cette perspective, car il repose sur l'idée d'une plasticité synaptique qui serait à la fois retardée sur l'échelle de temps ontogénétique, et qui serait également devenue plus *rapide et réversible*. En effet, pour rendre compte des facultés compositionnelles du langage, il nous faut également utiliser l'hypothèse de "*plasticité synaptique rapide*" postulée par von der Malsburg (1981), selon laquelle les efficacités synaptiques seraient sujettes à des variations relativement grandes sur une échelle de temps de l'ordre de la fraction de seconde. Comme on le verra, cette hypothèse s'avère nécessaire pour expliquer la composition des symboles, c'est à dire le liage des CST qui les représentent. La plasticité rapide a été reprise par d'autres auteurs comme Crick (1982), qui en a proposé un mécanisme biologique ("*twitching synapses*"). Cette propriété particulière ne nécessite pas une mutation complexe : elle peut intervenir au niveau

des protéines ou enzymes impliqués dans les processus de plasticité synaptique pour accélérer sa dynamique et la rendre réversible. Une telle mutation peut donc être relativement subtile et légère.

Nous commencerons au paragraphe 2.1 par un exposé général des mécanismes d'embryogenèse et de neurogenèse : cet exposé nous est apparu nécessaire afin de placer avec précision notre travail de modélisation dans la perspective neurobiologique appropriée. Le paragraphe 2.2 s'intéressera plus particulièrement au phénomène de *focalisation de l'innervation* qui est capital dans notre modèle, puis au 2.3, nous mettrons en évidence les deux mécanismes principaux de cette focalisation : la *compétition* et la *coopération*.

2.1 Embryogenèse et neurogenèse

Les siècles derniers furent marqués par les croyances *préformistes* : on imaginait que l'être humain était déjà tout prêt à l'intérieur de l'œuf fécondé et que son développement consistait en un simple agrandissement homomorphique de ses organes. Afin d'assurer la reproduction, cet "homonculus" possédait lui-même des cellules germinales contenant d'autres "homonculi", etc., ce principe récurrent étant appelé *emboîtement des germes*. Ces conceptions fausses ont culminé avec la dispute entre "ovistes" et "spermistes" : pour les uns, l'homonculus se trouvait initialement dans l'œuf, tandis que les autres affirmaient qu'il était apporté par les têtes des spermatozoïdes (vus au microscope) !

Vers le milieu du XIX^e siècle, Les idées préformistes ont progressivement décliné et fait place aux nouvelles théories introduites en particulier par Wolff, qui a su démontrer que le développement était une véritable création de structure, s'effectuant par *additions et différenciations cellulaires* successives. Puis, parallèlement à ces conceptions "interactivistes", sont également apparues les constatations fondamentales de la *génétique*, qui démontraient l'existence d'un programme transmis d'un individu à l'autre. Historiquement, il est donc intéressant de noter que les observations que nous qualifions aujourd'hui d'"épigénétiques" ont précédé les découvertes de la génétique.

Cependant, malgré les considérables progrès techniques accomplis depuis cette époque, la plupart des grandes questions relatives au développement n'ont pas encore reçu de réponse définitive. On continue aujourd'hui de s'interroger sur les causes des deux phénomènes fondamentaux de l'embryogenèse : la *différenciation* cellulaire et la *morphogenèse* des organes.

Une unique cellule, l'œuf fécondé, est le point de départ de la longue série de divisions cellulaires à partir desquelles se développera l'embryon, et, en particulier, le cerveau. Il est bien sûr difficile d'étudier séparément le système nerveux, car

celui-ci est influencé par le développement de tous les autres organes, avec lesquels il communique, et sur lesquels il exerce lui-même une influence. On peut néanmoins distinguer cinq grandes étapes dans sa formation : différenciation des neurones, division et migration, croissance des axones, mort neuronale, et formation des synapses. Dans cette séquence temporelle, chaque étape est préparée par l'étape précédente, soit au travers d'un déterminisme strict (part génétique), soit par un choix effectué dans un éventail plus ou moins large de possibilités, sous l'influence primordiale de l'environnement (part épigénétique). La dernière grande étape, appelée *synaptogenèse*, retiendra tout particulièrement notre attention, et fera l'objet d'un exposé plus détaillé au paragraphe 2.2.

Comment une seule cellule peut-elle donner naissance à tant de variétés cellulaires et moléculaires différentes ? Comment cette construction remarquable qu'est le cerveau, dont la configuration unique des connexions détermine un individu, est-elle engendrée au départ à partir de quelques cellules ? Ce paragraphe rappelle quelques notions élémentaires d'embryogenèse et de neurogenèse, en parcourant dans un premier temps la succession des événements qui mènent de l'œuf à l'enveloppe du système nerveux.

2.1.a Différenciation neuronale

La *différenciation* est le processus par lequel une cellule se spécialise, c'est à dire acquiert certaines propriétés physiologiques distinctes de celles d'autres cellules (son *phénotype*) : elle devient ainsi cellule musculaire, cellule nerveuse, cellule gliale, etc. Ce phénomène est fondamental dans tout le développement, et s'affine au cours des étapes. Par exemple, une cellule est d'abord appelée à devenir un neurone, puis migre vers une région spécifique où sa géométrie particulière apparaît avec l'élongation de ses neurites (axone et dendrites); elle adopte également un mode de transmission, un neurotransmetteur spécifique, etc.; enfin, sa fonction se précise à travers les contacts sélectifs qu'elle établit avec les autres cellules. Il y a donc toute une gamme de différenciations secondaires complexes, faisant suite à la première orientation neurale, qui est attestée par la très grande diversité des types neuronaux dans le système nerveux mature. Ces orientations sont dictées à la fois par les gènes et par l'environnement local : autrement dit, la différenciation dépend d'un cadre *génétique* et d'un contexte *épigénétique*. Le fait que ces deux influences agissent perpétuellement l'une sur l'autre rend par ailleurs la compréhension de ce phénomène assez difficile. En général, on peut dire que le phénotype est le résultat d'une *expression sélective* du génotype, qui est le même dans toutes les cellules, et que cette expression est régulée par des signaux en provenance des cellules avoisinantes : cette influence intercellulaire est appelée *induction*.

Les premiers signes de différenciation nerveuse remontent au tout début du développement embryonnaire. Initialement, la prolifération des cellules issues de la

division de l'œuf s'organise progressivement, et donne naissance à une petite structure en forme de sphère creuse : la *blastula* (du grec *blastos*, "germe"). Déjà à ce stade, les cellules de la blastula (les *blastomères*) ne sont pas toutes équivalentes du point de vue des divisions ultérieures, et leurs particularités dépendent de leur position : ainsi, les premières cellules de type neural sont situées sur la partie dorsale de cette boule creuse (notons cependant que cette différenciation n'est pas encore irréversible). Dans une deuxième étape, la blastula se replie en surface, et par un jeu de déformations compliqué, aboutit à la création des trois feuilletts cellulaires fondamentaux de l'embryon : l'*endoderme* (couche interne), le *mésoderme* (couche intermédiaire) et l'*ectoderme* (couche externe). Au cours de cette étape très importante, appelée *gastrulation* (du latin *gastrula*, "petit vase"), les cellules de l'ectoderme dorsal sont confirmées dans leur engagement vers une différenciation nerveuse par *induction* en provenance du mésoderme sous-jacent. Cette découverte est due à Mangold et Spemann (1924; voir Spemann, 1938), qui ont réussi à provoquer l'induction d'un second système nerveux chez un amphibien, par greffe d'un supplément de mésoderme. On voit donc que, dès le départ, l'environnement extracellulaire est crucial dans la genèse des précurseurs nerveux.

Une troisième étape du développement précoce, propre aux vertébrés, est marquée par l'ébauche du système nerveux, et a reçu le nom de *neurulation*. Au cours de ce processus, la plaque neurale ectodermique s'étire et se creuse, formant une gouttière, puis elle se referme, créant le *tube neural* qui sera à l'origine du système nerveux central (encéphale et moelle épinière). D'autres cellules nerveuses bordent le tube neural de part et d'autre par la *crête neurale*, futur système nerveux périphérique. Le cerveau est issu d'une différenciation de la partie antérieure du tube neural, qui se segmente en trois vésicules : à l'avant, le *télencéphale*, dont seront originaires les hémisphères cérébrales; au milieu, le *mésencéphale*; et à l'arrière, le *rhombencéphale*, qui fournira le cervelet. La partie postérieure se développera en moelle épinière. Ainsi, la position d'une cellule le long du tube neural aura des conséquences sur son appartenance à l'un de ces domaines. Cette constatation renvoie à un autre phénomène essentiel du développement, complémentaire de la différenciation : la *morphogenèse*, qui désigne le plan d'organisation cellulaire aboutissant à la formation des organes. La structuration morphogénétique est déclenchée par des gènes du développement, appelés *gènes homéotiques*, que l'on suppose en particulier responsables de la création des différents territoires encéphaliques mentionnés ci-dessus.

2.1.b Division et migration

A l'intérieur du tube neural et de la crête neurale en développement, les cellules prolifèrent à une cadence très rapide, et connaissent également une *activité migratoire* intense : ce phénomène est un déplacement *actif*, qui doit être distingué d'un déplacement passif où chaque cellule reste immobile par rapport à son

voisinage. Dans l'épaisseur du tube, les neurones migrent globalement de la zone ventriculaire (intérieure) vers la zone marginale (extérieure) (Rakic, 1972), tandis que de la crête neurale partent deux grands courants de migration qui iront coloniser des régions périphériques de l'embryon. A de rares exceptions près, la division d'un neurone cesse toujours avant sa migration. Après quelques cycles de divisions, les cellules nerveuses immatures du tube neural, appelées *neuroblastes*, se déplacent des zones germinatives vers leur emplacement définitif, où elles vont acquérir leurs caractères de maturation et pourront commencer à faire croître leurs neurites. Il est en effet très important que les neurones atteignent d'abord les bons emplacements, avant d'établir leurs connexions, afin que les contacts se fassent sur les cellules appropriées : l'information de position joue donc un rôle critique dans le câblage du système nerveux.

Le guidage d'une cellule en mouvement se fait en référence à des indices locaux extracellulaires. Certains neurones rampent le long de cellules gliales spécialisées qui s'étendent radialement à travers le tube neural; cependant, cette charpente est plutôt une structure d'orientation qu'un véritable support. Il semble en général que les neurones se repèrent tout au long de leur déplacement par certaines propriétés *d'adhésion différentielle* du milieu : autrement dit, parmi plusieurs surfaces présentes, les neurones s'accrochent de préférence à celles qui offrent, selon eux, la meilleure adhérence en fonction de critères chimiques complexes. Arrivés finalement à leurs places respectives dans différentes parties du cerveau, les neurones forment des agrégats spécifiques avec d'autres cellules du même type, par un mécanisme de reconnaissance moléculaire superficielle en partie identifié (N-CAM : "neural cell adhesion molecules"; Edelman, 1983). On pense que la quantité et la répartition de ces molécules de surface changent de façon systématique et programmée au cours du développement : les propriétés d'adhésion des cellules sont donc modifiées et exercent ainsi une influence importante sur les événements morphogénétiques.

2.1.c Croissance des axones

Une caractéristique unique aux neurones est leur capacité d'étendre des prolongements très loin du corps de la cellule, après migration de celui-ci, et de se constituer ainsi des voies de communication particulièrement efficaces et précises. L'information sera transmise très rapidement le long de l'axone, pourra toucher plusieurs cibles ou provenir de plusieurs sources à la fois, et ces contacts pourront être sélectifs, c'est à dire opérer un choix parmi plusieurs cibles potentielles. Ce phénomène d'extension des neurites et de contacts sur d'autres cellules est appelé *innervation*.

Les neurites poussent au niveau de leur extrémité où se trouve une structure très mobile, appelée *cône de croissance*, véritable "tête chercheuse" constituée de minces filaments (les *lamellipodia* et *filopodia*), qui rayonnent vers l'avant et

explorent l'environnement local. Le mécanisme d'avancée du cône de croissance est similaire à celui d'un déplacement cellulaire complet (de type amiboïde) : les extensions filamenteuses adhèrent au substrat extracellulaire, puis tractent le cône derrière eux. Ainsi, l'axone s'allonge progressivement par un jeu d'extensions et de contractions de son extrémité, où du matériel nouveau est synthétisé en permanence, un peu à la manière d'un fil produit par une araignée.

L'aspect le plus remarquable de l'innervation reste la précision avec laquelle les axones savent *s'orienter* au cours de leur croissance, et réussissent à atteindre les destinations appropriées après avoir souvent parcouru de très longues distances. Ce phénomène est bien sûr fondamental dans le développement des circuits neuronaux, et fait suite à celui du positionnement adéquat postmigratoire : il est vital en effet que les bons axones moteurs innervent les bons muscles, ou encore que les cellules de la rétine se projettent sur des aires visuelles et non auditives... Dans la jonction neuromusculaire, par exemple, on constate que les configurations d'innervation sont très stéréotypées. Cependant, la question de l'orientation axonale est aussi l'une des plus mal comprises de la neuroembryologie. Il est clair que les axones réagissent continuellement à des indices directionnels répartis dans l'environnement local, et plusieurs mécanismes de guidage ont été envisagés : stérotropisme (existence de supports physiques directs), adhésion différentielle, galvanotropisme (faibles champs électriques), chimiotropisme (gradients chimiques), etc. Cependant, aucun ne fournit une explication complète de ce phénomène, car sans doute plusieurs de ces influences sont à l'œuvre pour assurer une navigation précise des prolongements axonaux à travers un milieu embryonnaire en perpétuelle évolution.

2.1.d Mort neuronale

Il est désormais admis que le développement procède aussi par phases de *mort cellulaire*, qui affectent tous les organes, et en particulier le système nerveux. Cette phase intervient au moment où l'ensemble des neurones forment des connexions avec les cellules cibles : suivant les régions, la perte neuronale peut varier entre 20% et 80% de la population initiale. Ce phénomène paraît étonnant si l'on songe que, par la suite, les neurones seront les seules cellules de l'organisme adulte qui ne se renouvelleront plus jamais.

La mort naturelle des cellules nerveuses est l'œuvre de stratégies différentes, qui sont fonction de la "complexité" des animaux. De façon générale chez les invertébrés, les neurones meurent par l'action d'un programme intrinsèque : les cellules individuelles descendent d'un lignage stéréotypé, dont une partie est invariablement éliminée. Chez les vertébrés, en revanche, la disparition des neurones n'est pas prédéterminée, mais dépend essentiellement de leurs interactions avec les autres cellules, et principalement d'une *compétition* entre axones au niveau des cellules innervées. Ce mécanisme de régulation est essentiel

dans les organismes complexes, où chaque fonction est assurée par un grand nombre de neurones, contrairement aux organismes simples où celles-ci ne mobilisent que quelques cellules nerveuses voire une seule. Ainsi, chez les vertébrés, la disparition neuronale lors de l'embryogenèse fait partie intégrante de la stratégie d'adaptation permanente par sélection épigénétique. En effet, il est clair que le but important de cette compétition est d'ajuster le nombre de cellules innervantes à la taille du groupe des cellules cibles. Des expériences menées sur des embryons de poulet (Hamburger, 1934) montrent bien cette proportionalité : l'amputation d'un bourgeon de membre a pour conséquence une réduction massive du nombre de motoneurones et de neurones sensitifs innervants (situés dans les ganglions spinaux), tandis que la greffe d'un membre surnuméraire provoque l'augmentation de celui-ci.

Il semble que l'objet de la compétition se trouve dans certains facteurs chimiques produits par les cellules cibles, appelés *facteurs trophiques* (du grec *trophê*, "nourriture") dont l'acquisition rétrograde par le neurone améliore ses chances de survie. Faute de pouvoir s'en procurer suffisamment, l'axone se rétracte (littéralement : *s'atrophie*) et finit par provoquer la dégénérescence du corps cellulaire entier. La protéine NGF, ou *nerve growth factor* (facteur de croissance du nerf) découverte par Levi-Montalcini (1953), est un exemple important d'agent trophique : son action a été mise en évidence dans la chaîne nerveuse sympathique et les ganglions sensoriels. De façon générale, il semble que l'acquisition de facteurs trophiques soit non seulement à la base de la survie neuronale mais aussi de la régulation ultérieure des connexions synaptiques (cf. paragraphe 2.3). Notons dès maintenant que les mécanismes de compétition joueront un rôle central dans le modèle que nous développerons au chapitre 3.

2.1.e Formation des synapses

Nous arrivons maintenant aux phénomènes qui intéresseront tout particulièrement notre modèle. Les événements précédents que nous venons de décrire ont installé le cadre général de l'innervation, qui comporte déjà, comme on l'a montré, une grande part de *spécificité* : les neurones ne migrent pas au hasard, et leurs prolongements s'orientent avec une remarquable précision pour parvenir aux tissus cibles. Puis leur nombre, a priori excessif, est adapté à la taille de la zone innervée par compétition et mort neuronale. Le système nerveux arrive alors à une étape cruciale de son développement, car il reste à former les connexions synaptiques entre neurones, ou entre neurones et organes périphériques, qui détermineront entièrement son bon fonctionnement : c'est la *synaptogenèse*. Ainsi, après les grandes orientations de l'innervation, qui sont déjà en soi porteuses d'une information particulière, des circuits neuronaux beaucoup plus fins doivent maintenant se former grâce à une distribution judicieusement réglée des connexions. Ce phénomène fondamental dans le développement du système

nerveux, que l'on nomme *focalisation de l'innervation*, sera examiné en détail au paragraphe 2.2. La présente section se contente de décrire quelques caractéristiques morphologiques de la création des contacts synaptiques.

La synaptogenèse débute très tôt et se poursuit très longtemps au cours de l'ontogénèse : les synapses commencent à croître dès l'approche des cellules cibles par les terminaisons axonales, tandis que les spécialisations présynaptiques et postsynaptiques se développent graduellement. L'ensemble est déjà fonctionnel bien avant d'avoir achevé sa maturation structurelle. L'étude de la jonction neuromusculaire a fourni un grand nombre de données expérimentales très importantes sur ce phénomène, et a permis de comprendre comment les éléments présynaptiques pouvaient *induire* certains changements sur la membrane de la cellule postsynaptique. Il est raisonnable de penser que ces résultats sont généralisables, dans une certaine mesure, aux connexions interneuronales. La réorganisation des contacts synaptiques sur les fibres musculaires est également un exemple important de focalisation de l'innervation, dont nous reparlerons plus bas.

Dans la jonction neuromusculaire, on constate que les terminaisons axonales ne recherchent pas des sites récepteurs postsynaptiques préexistants (spécifiques à l'acétylcholine), mais ont au contraire une influence très nette sur l'agrégation de ceux-ci, apparemment au moyen de signaux chimiques, mais aussi à travers l'activité électrique. Sur certaines fibres, caractérisées par une innervation multiple de toute leur longueur, les différentes afférences imposent une distance minimale de séparation entre plaques motrices, sans doute par une influence chimique réfractaire. D'autres changements postsynaptiques tels qu'un épaissement de la membrane ou l'apparition de replis postjonctionnels (sillons) sont également induits par les éléments présynaptiques. Après une déafférentiation, en revanche, les sites récepteurs initialement formés sont privilégiés par le processus de réinnervation qui s'ensuit, ce qui prouve que la cellule postsynaptique a également une influence attractive sur la cellule présynaptique.

Par ailleurs, il semble qu'il soit possible de provoquer une synaptogenèse entre divers éléments pré- et postsynaptiques, combinés de façon inhabituelle dans des expérimentations *in vitro* : dans la plupart des cas, des synapses apparemment normales sont formées, leur fonctionnement étant cependant plus ou moins altéré. Donc ces tests tendent à prouver que les signaux propres à la synaptogenèse sont à peu près uniformisés dans toutes les régions du système nerveux.

En résumé, la synaptogenèse est un cas particulier d'interaction trophique : on peut dire qu'elle est le résultat d'un *dialogue prolongé entre éléments pré- et postsynaptiques* au cours du développement.

2.2 Focalisation de l'innervation

Comme nous l'avons souligné à propos de la croissance axonale, les neurones ont la faculté unique de développer un réseau de communication intercellulaire ciblé de façon précise, et ce sur de grandes distances. Les premières étapes de la neurogenèse qui ont été décrites ci-dessus tracent les grandes lignes de ce réseau, dans lesquelles vient ensuite s'inscrire la phase de synaptogenèse, caractérisée par la formation de circuits beaucoup plus fins. L'ordre remarquable des connexions qui apparaît alors est sans doute l'aspect le plus frappant de la croissance du système nerveux. La précision de ce câblage est bien sûr d'une importance fondamentale pour l'individu, car toutes les fonctions neuronales en dépendent : il est prouvé par exemple que certaines cellules corticales possèdent une *spécificité fonctionnelle*, c'est à dire qu'elles réagissent à des stimuli spécifiques de l'environnement (tels des contours orientés), et que ces propriétés sont le résultat du développement de circuits synaptiques très particuliers.

La question est alors de comprendre comment ces connexions précises et ordonnées sont formées au cours du développement. Entre un point de vue "néo-préformiste", qui suppose que le câblage est entièrement programmé à l'avance, et un point de vue "empiriste" qui suppose au contraire qu'aucune connexion ne peut être établie sans être guidée par l'expérience sensorielle, le point de vue de compromis couramment adopté aujourd'hui envisage l'existence d'une "*enveloppe de connexions*" préalable, qui est présente avant tout fonctionnement, et qui se préciserait ensuite graduellement, à la fois par une interaction avec l'environnement, mais aussi par une auto-régulation interne.

C'est ainsi que la théorie de la "*stabilisation sélective*" élaborée par Changeux et Danchin (1976) prévoit dans un premier temps une croissance synaptique excédentaire au cours de la neurogenèse, suivie dans un deuxième temps par une phase de *sélection*, au moment où le système commence à fonctionner : au cours de cette phase, certaines synapses sont stabilisées par l'activité, tandis que d'autres disparaissent. Il s'agit donc de la régression d'une multi-innervation transitoire, c'est à dire de la *focalisation* d'une innervation initialement large et diffuse (focalisation de l'"enveloppe génétique"). Ainsi, à partir d'un état de connexions non-sélectif, le réseau fait la transition progressive vers un état *sélectif* : la compétence fonctionnelle des cellules corticales est au départ relativement étendue, puis ces cellules se différencient les unes des autres par l'acquisition de propriétés d'activation sélective (dominance oculaire ou sélectivité à l'orientation dans le système visuel). En résumé : le génome trace un cadre et ouvre des possibilités, puis l'expérience choisit. Si l'on en croit l'aphorisme de Changeux : « *Apprendre, c'est éliminer.* »

La focalisation de l'innervation est présente dans de nombreuses régions du système nerveux, et nous présenterons ici les exemples les plus étudiés : jonction

neuro-musculaire, système autonome, cervelet, cortex visuel primaire, et projection rétino-topique. Tous ces cas montrent que le fonctionnement neuronal dépend en dernier lieu de relations qualitatives et quantitatives précises entre cellules nerveuses (ou entre cellules nerveuses et fibres musculaires). Comme on le sait, dans la perspective de notre modèle, nous nous intéresserons à une généralisation de ce processus aux aires cognitives supérieures, en nous attachant davantage à la notion de "stabilisation interne", préalable à l'acquisition d'une expérience par interaction avec l'environnement.

2.2.a Jonction neuromusculaire

L'innervation des muscles squelettiques par les neurones moteurs présente une configuration différente selon qu'il s'agit d'un muscle néo-natal ou d'un muscle mature. En général, chez les mammifères adultes, chaque fibre musculaire est contactée par un seul axone, tandis qu'à la naissance elle peut être innervée par plusieurs axones provenant de motoneurones différents : la maturation est donc caractérisée par une régression de la multi-innervation au niveau de la jonction neuro-musculaire (Redfern, 1970). Expérimentalement, le nombre de fibres innervantes est mesuré en stimulant directement le nerf moteur regroupant l'ensemble des axones : une augmentation progressive de l'intensité présynaptique provoque alors des sauts discrets dans la réponse postsynaptique, ce qui s'explique par le recrutement successif de plusieurs axones différents.

Initialement, ces contacts multiples sur une fibre sont regroupés à peu près en un même endroit, qui correspond à la future et unique plaque motrice : les changements structuraux intervenus entre l'âge néo-natal et l'âge adulte sont donc le résultat d'une *élimination synaptique*. Ce phénomène commence en général avant la naissance, et se poursuit pendant les premières semaines post-natales. Il est confirmé que la perte synaptique ne s'accompagne d'aucune modification notable du nombre de neurones innervants ou de cellules cibles. En conséquence, parallèlement à la réduction de la convergence, il y a également réduction de la divergence : la taille de l'unité motrice (ensemble des fibres musculaires innervées par un même motoneurone) diminue. Cet effet peut être mesuré à travers la tension musculaire provoquée par un seul axone. Chez d'autres vertébrés que les mammifères, on trouve fréquemment des fibres innervées en de multiples sites, plus ou moins régulièrement espacés sur toute leur longueur (au lieu d'une seule plaque motrice). Là aussi, les sites en question sont initialement occupés par plusieurs axones différents, puis par un seul. Ceci prouve que le phénomène d'élimination synaptique dans la jonction neuromusculaire est très général.

2.2.b Système nerveux autonome

L'élimination post-natale des synapses n'est pas limitée aux muscles, mais

intervient également au niveau des cibles neuronales. La régression de la multi-innervation a été démontrée sur des systèmes relativement simples tels que les ganglions parasympathiques de certains mammifères (Purves et Lichtman, 1980). Dans le ganglion submandibulaire du rat, les cellules sont contactées par un seul axone préganglionnaire à l'âge adulte, mais initialement par environ 5 d'entre eux. Ce cas de figure, qui aboutit à une mono-innervation à l'instar des fibres musculaires, n'est pourtant pas la seule possibilité : par exemple, on observe une convergence finale d'environ 6 axones dans le ganglion cervical supérieur du hamster, contre une convergence initiale double; ou encore, la convergence décroît de 5 à 2 environ dans le ganglion cilié du lapin. Comme pour les muscles, cette réduction n'est pas due à une variation du nombre de cellules pré- et postsynaptiques, qui reste à peu près constant, donc, à nouveau, ceci implique que la divergence est également réduite : "l'unité neurale", c'est à dire le champ d'action d'un axone, diminue tout comme l'unité motrice dans la jonction neuromusculaire.

Remarquons cependant que, tandis que la convergence et la divergence diminuent, la surface totale de contacts synaptiques a plutôt tendance à *augmenter* : on trouve en effet plus de contacts synaptiques sur les cellules ganglionnaires de ces mammifères à un âge de quelques semaines que sur les mêmes cellules à la naissance. C'est aussi le cas de la jonction neuromusculaire où la convergence chute à un seul contact par fibre musculaire, où cet unique "contact" représente en réalité un vaste ensemble de boutons synaptiques concentrés sur la plaque motrice de la fibre. Donc, en général, la régression de la multi-innervation ne consiste pas en une simple élimination synaptique : il y a bien disparition de terminaisons axonales et de synapses en provenance d'autres axones, mais dans le même temps il y a aussi formation de nouvelles synapses en provenance de l'axone restant, qui renforce ainsi sa domination et compense largement la perte initiale. C'est pourquoi le nombre total de contacts est en réalité en augmentation globale.

La conclusion est donc la suivante : les axones forment initialement un petit nombre de synapses sur un nombre relativement élevé de cibles, puis, au fur et à mesure du développement, ces contacts initiaux sont réarrangés de telle sorte qu'un plus grand nombre de synapses est concentré sur un plus petit groupe de cellules postsynaptiques. Ce phénomène correspond précisément à la focalisation de l'innervation, et on le retrouvera dans le comportement du modèle exposé dans les chapitres suivants.

2.2.c Cervelet

Les phénomènes de focalisation sont aussi présents dans de nombreuses parties du système nerveux central. On les retrouve en particulier dans le cervelet, au niveau de l'innervation des cellules de Purkinje par les fibres grimpantes (Crepel *et al.*, 1976; Mariani et Changeux, 1981).

Le cortex cérébelleux est une structure relativement simple et uniforme, constituée de trois couches contenant en tout cinq types de neurones, parmi lesquels les *cellules de Purkinje*, dont les axones constituent l'unique voie de sortie du cervelet (leur action est inhibitrice). Ces cellules possèdent en outre une géométrie assez particulière : leur arbre dendritique très fourni est entièrement contenu dans un plan, perpendiculaire à un faisceau de *fibres parallèles* qui les innervent par milliers, à la manière de fils télégraphiques traversant des poteaux (ces fibres sont les axones d'un autre type de neurones, appelés *cellules granulaires*). En dehors des fibres parallèles et des autres afférences provenant de l'intérieur du cervelet, les cellules de Purkinje sont aussi innervées par des axones spécifiques, les *fibres grimpantes*, dont l'origine se situe dans un noyau du bulbe rachidien. Chez l'animal adulte, une même fibre grimpante peut se subdiviser pour établir des connexions avec plusieurs cellules de Purkinje (en petit nombre, de 1 à 10 environ), mais inversement, chaque cellule de Purkinje ne reçoit des connexions synaptiques que d'une et une seule fibre grimpante. Cette régularité dans la connectivité est tout à fait remarquable, et constitue par ailleurs l'un des plus forts contacts que l'on connaisse dans le système nerveux : la fibre s'enroule de nombreuses fois autour du corps cellulaire en épousant également la partie proximale de son arborisation dendritique, et, par conséquent, un seul potentiel d'action de la fibre grimpante suffit à induire systématiquement un énorme potentiel postsynaptique excitateur chez la cellule de Purkinje (appelé *PSP complexe*), qui provoque aussitôt son activation à grande échelle. Ceci peut être mis en évidence par stimulation de la fibre grimpante.

Or, les mêmes mesures électrophysiologiques faites dans le cervelet d'animaux nouveaux-nés montrent en revanche que l'innervation d'une cellule de Purkinje est assurée par *plusieurs* fibres grimpantes différentes. Ainsi, au cours des premières semaines qui suivent la naissance, ces afférences sont toutes perdues, sauf l'une d'entre elles, qui est au contraire renforcée.

2.2.d Cortex visuel primaire

D'autres phénomènes de réarrangements synaptiques plus subtils, concernant directement les facultés supérieures du cerveau, sont observables dans différentes régions du cerveau antérieur : aires sensorielles primaires du cortex (visuelle, auditive, olfactive), noyaux sous-corticaux (corps genouillé latéral), corps calleux, etc. La sélectivité de l'innervation la mieux étudiée est probablement celle du système visuel : cortex strié, corps genouillé latéral (CGL), rétine. Dans ce paragraphe, nous décrirons brièvement l'état des connaissances sur le système visuel, en montrant l'apparition d'une connectivité sélective au cours du développement, puis nous reviendrons au 2.3 sur les mécanismes de type compétitif et coopératif de cette focalisation, qui ont pu être mis en évidence à travers des expériences de privation sensorielle.

Chez l'adulte (expériences faites sur le singe), la surface du cortex est organisée en *bandes de dominance oculaire* au niveau de la couche IV, ce qui veut dire que les cellules corticales sont réparties en colonnes (structures corticales perpendiculaires aux couches) dominées alternativement par l'œil gauche ou l'œil droit, et regroupées en bandes parallèles. Ces regroupements peuvent être mis en évidence au moyen d'enregistrements électriques effectués régulièrement le long du chemin suivi par une électrode. La plupart des cellules de la couche IV sont *monoculaires*, c'est à dire qu'elles répondent préférentiellement à l'activation d'un œil ou de l'autre : ainsi, dans une pénétration perpendiculaire à la surface du cortex, l'électrode ne rencontre qu'un seul type de dominance (gauche ou droite), tandis que dans une pénétration tangentielle, la dominance oculaire alterne tous les demi-millimètres environ (Hubel et Wiesel, 1968).

Une preuve plus directe de cette organisation peut être obtenue en injectant un traceur chimique radioactif dans l'un des deux yeux : assimilé par les cellules de la rétine, ce traceur parvient au cortex visuel au bout de plusieurs jours, via le corps genouillé latéral, grâce au transport axonique antérograde. Une autoradiographie du cortex montre alors nettement une alternance de bandes claires et sombres, ces dernières correspondant à l'œil marqué. Par les mêmes procédés, on constate alors que les domaines oculaires sont quasi inexistant 6 semaines avant la naissance, et qu'il sont seulement faiblement apparents à une semaine prénatale (Hubel, Wiesel et LeVay, 1977). Ainsi, chez le singe nouveau-né, la plupart des neurones de la couche IVc du cortex visuel sont *binoculaires*, c'est à dire indifféremment sensibles à chacun des yeux. Il est donc clair que les cellules corticales se spécialisent progressivement vers une dominance oculaire à travers un réarrangement synaptique qui se déroule au début de la vie post-natale, et qui aboutit à la formation de groupes fonctionnellement distincts. L'interprétation classique de ce phénomène envisage une *rétractation* des axones du CGL innervant la couche IV : les efférences des axones ipsi- et contralatéraux sont initialement en large recouvrement sur les cellules corticales, puis la divergence est réduite et les domaines d'innervation se séparent.

Comme avec la dominance oculaire, la sélectivité de l'innervation dans les aires sensorielles primaires du cortex peut être mise en évidence de façon générale à travers les *propriétés fonctionnelles* des neurones : dans le système visuel, ces propriétés peuvent être résumées par la notion de *champ récepteur*, dont on rappelle ici la définition. Dans la rétine, les premières cellules traitant l'information en provenance du monde extérieur sont les *photorécepteurs* (cônes et bâtonnets) : ceux-ci effectuent la transduction de l'intensité lumineuse en potentiel électrique, qu'ils transmettent ensuite aux cellules *bipolaires*, lesquelles activent à leur tour les cellules *ganglionnaires*. Les axones de ces dernières se rassemblent pour former le nerf optique, qui sort de la rétine et se projette sur le CGL (situé dans le thalamus), que l'on peut assimiler à un "relais". Puis, des couches du CGL partent des faisceaux

de fibres (les radiations optiques), qui iront établir des contacts sur le cortex visuel primaire situé dans le lobe occipital (aire 17), appelé également *cortex strié*. En général, les axones en provenance du CGL innervent préférentiellement la couche IV, mais pas uniquement. Dans ce cadre, étant donné une cellule quelconque du système visuel primaire (rétine, CGL, ou cortex visuel), le "champ récepteur" de cette cellule est défini comme étant la région de la rétine qui contribue à son activation. Pour déterminer celle-ci, on stimule sélectivement des groupes restreints de photorécepteurs au moyen de spots lumineux projetés sur un écran vers lequel est fixé le regard d'un animal, en général anesthésié et paralysé pour éviter les mouvements oculaires qui rendraient difficile la détermination précise du champ : il existe donc une correspondance biunivoque entre la surface rétinienne et l'écran. Pendant qu'une électrode mesure l'activité du neurone étudié, on fait varier la position, la taille et la forme de la tache de lumière pour mettre en évidence les zones qui ont une influence sur son activation. On réussit de cette façon à délimiter progressivement le champ récepteur. Cependant, plus encore que l'étendue de ce champ, on constate que l'état de la cellule est aussi très fortement dépendant de la *forme* particulière du stimulus lumineux. Ainsi, certaines cellules ganglionnaires de la rétine sont spécifiquement sensibles à un spot lumineux ponctuel entouré d'une couronne d'obscurité, tandis que d'autres répondent à la configuration inverse, à savoir un spot sombre entouré d'une couronne lumineuse. Des cellules du cortex strié (en dehors de la couche IV) présentent une réponse encore plus spéciale : elles réagissent sélectivement à la présence d'une petite barre lumineuse (ou inversement, sombre) ayant une certaine *orientation*. Ces importantes découvertes sont dues à Hubel et Wiesel (1959, 1962, 1968, 1974).

La question qui nous intéresse ici est donc celle des mécanismes présidant à la mise en place des connexions responsables de cette sélectivité. Des expériences de privation visuelle, telle qu'une privation monoculaire (Hubel et Wiesel, 1970) ou un environnement visuel anormal constitué de lignes verticales (Blakemore et Cooper, 1970), ont montré que la dominance oculaire et la sélectivité à l'orientation s'installaient progressivement pendant une *période critique* chez l'animal jeune : ainsi, ces observations faites en intervenant sur l'environnement sensoriel de l'animal ont permis d'isoler la partie épigénétique des facteurs contribuant au développement de ces propriétés. Pour expliquer ces phénomènes, l'hypothèse d'une "stabilisation sélective" est donc très vraisemblable : les cellules du cortex visuel primaire auraient, avant d'être complètement spécifiées du point de vue de leur sélectivité à l'orientation, une compétence potentielle relativement large, qui serait assurée par une innervation diffuse (ce qu'on appelle "enveloppe génétique"); puis, au cours des premières étapes du fonctionnement du système, ces connexions synaptiques se focaliseraient, pour permettre à chaque cellule d'acquérir sa singularité, c'est à dire par exemple, ses propriétés de sélectivité à l'orientation ou de dominance oculaire. Bien entendu, cette sélectivité est programmée dans le code

génétique, mais pour s'exprimer, il lui faut une interaction normale avec l'environnement : les expériences de privation sensorielle montrent qu'une expérience précoce anormale conduit à l'établissement de configurations nerveuses tout aussi anormales.

2.2.e Projection rétinotopique

Des observations similaires ont été faites concernant la propriété très importante de *rétinotopie* : il s'agit ici de l'application de la surface rétinienne sur le tectum, qui est la structure visuelle centrale des poissons et des amphibiens (ou sur le cortex visuel par l'intermédiaire du CGL chez les mammifères). La projection des fibres rétiniennes sur le tectum a l'aspect d'une véritable carte topographique, comme c'est également le cas d'autres projections neurales : la propriété essentielle de cette organisation est de conserver au niveau du tectum les relations de voisinage entre cellules ganglionnaires de la rétine. La correspondance peut être reconstituée point par point en mesurant l'activité électrique des cellules tectales au moment où une région particulière de la rétine est éclairée par un spot lumineux (recherche des champs récepteurs).

On s'est donc posé la question de savoir par quels mécanismes cette application continue se mettait en place. Sperry fut l'un des premiers chercheurs à s'être intéressé à cette question du point de vue expérimental (dans les années 1940; voir Sperry, 1963). Ses travaux sur la grenouille l'ont conduit dans un premier temps à l'hypothèse d'une *chémo-affinité*, c'est à dire d'une identification des cellules de la rétine par un marqueur chimique qui leur permettrait de reconnaître leurs partenaires dans le tectum marqué par la même molécule ou par une molécule complémentaire. Il s'agissait donc d'un marquage rigide qui devait expliquer les premières observations de Sperry : une section du nerf optique était suivie d'une régénération et d'une réinnervation sélective dans laquelle les cellules de la rétine retrouvaient les partenaires tectaux avec lesquels elles avaient initialement établi leurs contacts.

Par la suite, ces résultats ont été complétés grâce à des expériences plus détaillées (sur la grenouille et le poisson rouge), où l'on permettait au système de se réorganiser : ces expériences ont alors montré que la projection rétinotopique est en réalité capable d'une flexibilité tout à fait remarquable, contredisant ainsi l'hypothèse d'un câblage rigide par marquage chimique préétabli. Par exemple, l'ablation d'une moitié de la rétine ou d'une moitié du tectum donne lieu à une *compensation* de la projection : au début de la réinnervation, les axones retrouvent leurs cibles initiales (pouvant faire croire à une certaine prédilection de la projection pour son état antérieur), mais cette projection s'étend peu après, et finit par occuper tout le territoire resté disponible (i.e. toute la rétine sur un demi-tectum ou une demi-rétine sur le tectum entier). De même, des "yeux composites" peuvent être créés en réunissant deux moitiés rétiniennes nasales ou deux moitiés temporales :

dans ce cas, si les marqueurs chimiques étaient uniques, ces deux parties devraient se projeter sur une seule moitié du tectum; or elles innervent finalement la totalité de celui-ci. Ainsi, la spécificité du système rétino-tectal n'est pas due, comme on a d'abord pu le croire, à des signaux cytochimiques qui attireraient invariablement les fibres vers la formation d'un diagramme rigide : ces expériences, tout en confirmant l'idée d'une affinité sélective à base chimique, prouvent que la reconnaissance est *relative* et non absolue.

Différents modèles ont été proposés pour expliquer la grande flexibilité de la rétinitopie, et en particulier le modèle de *marqueurs induits* de Willshaw et von der Malsburg (1979) : l'hypothèse d'un marquage chimique déterminant l'identité de chaque cellule est toujours présent, cependant, les marqueurs sont synthétisés uniquement au niveau de la rétine. Les molécules sont alors transportées de façon antérograde par les axones, et diffusent dans le tectum : chaque cellule tectale peut ainsi être marquée de façon non-équivoque. Autrement dit, les différentes molécules n'indiquent pas une position dans l'absolu, mais l'information positionnelle est codée de façon *relative* à travers le rapport des concentrations : les afférences se placent les unes par rapport aux autres, en respectant la continuité d'une "carte chimique" bidimensionnelle. Ce marquage dépendant lui-même de l'application rétino-tectale, on est donc en présence d'un système proprement *auto-organisateur*. Ce système est également capable de compensation, ainsi que l'ont montré en détail Willshaw et von der Malsburg dans une étude numérique approfondie de leur modèle (reproduction de toutes les expériences d'ablation d'une demi-rétine ou d'un demi-tectum suivies de régénération, ou de rotation ou implémentation d'un œil ectopique). Les deux grands principes sous-jacents à l'auto-organisation de la carte rétinotopique sont donc la compétition et la coopération, qui sont mis en œuvre ici à travers des marqueurs chimiques.

Par ailleurs, le principe selon lequel les cellules topologiquement voisines coopèrent pour établir des contacts synaptiques fonctionnels dans la région du tectum a également été réalisé dans d'autres modèles (Willshaw et von der Malsburg, 1976) en utilisant une identité *temporelle* et non plus chimique : donc le code de la reconnaissance intercellulaire se trouve ici dans la configuration précise d'activité portée dans chaque cellule. Depuis, il a été confirmé que c'est bien un marquage de type temporel qui est à l'œuvre dans certains cas (Schmidt, 1985) : des expériences menées sur le poisson rouge, élevé dans un environnement visuel anormal, montrent qu'une altération de la structure temporelle de l'activité des fibres rétiniennes empêche la projection rétinotectale de se développer de façon satisfaisante. Parmi les expériences consistant à bloquer ce processus de focalisation, une illumination stroboscopique a été utilisée pour provoquer la corrélation artificielle de toutes les activités des cellules ganglionnaires : cette forte corrélation brise alors la structure de corrélation fine dépendante de l'espace dans la rétine, et

interfère de cette façon avec l'établissement de la projection rétinotopique, qui devient beaucoup moins précise que dans un développement normal s'appuyant sur un régime d'activation "topologique" interne à la rétine. Ainsi, les expériences de Schmidt prouvent que l'*activité corrélée* joue également un rôle très important dans la focalisation de la projection diffuse pendant le développement précoce.

En conclusion, comme le souligne Schmidt, la chémo-affinité invoquée en premier lieu semble surtout intervenir dans l'organisation d'une carte rétinotopique relativement grossière, tandis que le raffinement de cette carte, c'est à dire l'établissement de connexions plus précises, doit être obtenu par un mécanisme de stabilisation sélective fondé sur les *corrélations d'activité* des cellules. Autrement dit, deux mécanismes seraient à l'œuvre dans le développement des projections rétinotopiques, et sans doute dans l'organisation d'autres cartes sensorielles : un mécanisme chimique qui détermine l'orientation de l'application, et un mécanisme de corrélations temporelles qui détermine le réglage fin de cette application. C'est en particulier ce dernier mécanisme coopératif que nous retiendrons et que nous développerons dans notre modèle.

2.3 Mécanismes du développement de la sélectivité

2.3.a Nécessité d'un mécanisme compétitif

Il est certain que le phénomène le plus important à la base de la redistribution des synapses est une *compétition* entre axones innervant la même cellule cible : certaines terminaisons se rétractent, d'autres ayant eu plus de "succès" qu'elles. Par exemple, chez le rat, si on effectue à la naissance une dénervation des muscles lombaires de telle sorte qu'un seul axone soit sauvegardé, on constate que la taille de l'unité motrice restante ne diminue pas comme elle devrait le faire normalement, faute de compétition en l'absence d'axones voisins. D'autre part, il est impossible d'expliquer autrement que par un mécanisme compétitif l'innervation véritablement *bijective* qui existe au niveau de certains ganglions autonomes (un axone innerve un et un seul neurone). En effet, si la rétractation des axones n'était due qu'au hasard, les cellules cibles n'auraient aucune raison de recevoir exactement une afférence : certaines d'entre elles se retrouveraient avec plus d'une afférence, tandis que d'autres n'en posséderaient aucune. Dans certains cas, le réarrangement synaptique est également influencé par la forme géométrique du neurone cible, et en particulier du nombre de ses dendrites : ainsi, plus la surface offerte est grande, plus les terminaisons présynaptiques d'origines différentes peuvent se déployer, et éviter d'entrer en compétition les unes avec les autres sur les mêmes sites (l'innervation de la cellule de Purkinje par la fibre grimpante fait

figure d'exception). Il existe de même une distance minimale de séparation des plaques motrices sur les fibres musculaires multiples innervées.

Comme on l'a évoqué plus haut (§2.2.d), le rôle de la compétition dans la redistribution des synapses au cours du développement a pu être mis en évidence par des expériences de privation visuelle, telles que la *privation monoculaire* chez le singe (œil droit maintenu fermé par suture, de l'âge de 2 semaines à l'âge de 18 mois) (Hubel, Wiesel et LeVay, 1977) : une autoradiographie du cortex montre alors une largeur beaucoup plus importante des bandes de dominance oculaire correspondant à l'œil resté ouvert. Un autre exemple spectaculaire de ségrégation des afférences induite par compétition a été obtenu en greffant un troisième œil à un têtard (Constantine-Paton et Law, 1978). Normalement, la projection rétino-tectale de la grenouille se fait sans interférence entre les chemins optiques : l'œil gauche se projette entièrement sur le tectum droit et réciproquement. Or, lorsqu'une troisième rétine envoie également des axones vers le toit optique, ceux-ci doivent alors entrer en compétition avec l'un des deux yeux pour une partie de la zone à innover. Au cours de la maturation, cet événement force l'apparition de bandes de dominance oculaire très semblables à celles du singe ou du chat, mais absentes chez la grenouille normale, où les fibres optiques en provenances des deux yeux ne sont pas en compétition.

Comme on l'a déjà mentionné à propos de la mort neuronale (§2.1.d), l'objet de cette compétition est probablement un support trophique produit en quantité limitée par les cellules postsynaptiques et assimilé dans le sens rétrograde par les terminaisons présynaptiques (Purves et Lichtman, 1980, 1985) : ainsi, les mêmes molécules qui seraient nécessaires à la survie de la cellule joueraient également un rôle dans la régulation des contacts synaptiques. Mais dans ce cas, il reste à expliquer la tendance systématique à la *régression* d'une multi-innervation qui semble caractériser l'ensemble des phénomènes de focalisation. En effet, le principe d'une compétition suppose seulement une limitation de la quantité d'afférences, mais n'a pas de conséquences plus précises sur la configuration de l'innervation : avec une contrainte quantitative globale, rien n'indique si les afférences doivent se concentrer sur une seule cible ou se répartir et coexister sur plusieurs cibles à la fois.

Il faut donc approfondir l'interprétation de la façon suivante : il semble que les contacts créés par un neurone présynaptique aient davantage de chances d'être stabilisés s'ils se concentrent sur un nombre restreint de cellules cibles que s'ils se maintiennent sur une région étendue. Cette constatation suppose alors l'existence d'un *mécanisme non-linéaire*, c'est à dire de type *coopératif*.

2.3.b Nécessité d'un mécanisme coopératif

Ainsi, le mécanisme compétitif est insuffisant pour expliquer la focalisation très spécifique de l'innervation. De plus, comme on l'a fait remarquer au 2.2.b, la surface des contacts synaptiques est généralement en augmentation globale sur la

cellule cible privilégiée par un axone : ceci veut donc dire que le facteur trophique existe en réalité en plus grande quantité que ce qui était nécessaire aux connexions initiales.

Par contre, il semblerait que ce support fasse justement défaut lorsqu'il s'agit de soutenir des contacts appartenant à plusieurs axones *différents*. Il faut donc rechercher une explication dans l'existence d'une "identité" synaptique : les boutons synaptiques appartenant aux fibres afférentes en provenance d'un même neurone, ou d'une même région de cellules-sources (région rétinienne, par exemple), sont vraisemblablement traités comme un *ensemble* cohérent au cours de l'innervation. Quel serait alors le codage permettant de réaliser cette discrimination ? Un marquage chimique l'explique difficilement, en revanche, le point commun évident à toutes les synapses d'un même axone est la *synchronie de leur activité électrique*.

On sait déjà le rôle crucial joué par l'activité dans les réarrangements synaptiques, et de nombreuses expériences sur la jonction neuromusculaire le prouvent (une paralysie des nerfs entraîne un ralentissement de l'élimination synaptique sur la fibre musculaire, et, au contraire, des stimulations répétées accélèrent ce phénomène). Cependant, plus que la quantité totale d'activité, c'est l'activité *relative* entre axones compétiteurs qui peut servir de guide à la focalisation. L'expérience la plus édifiante illustrant ce dernier point a consisté à *décorrélér* les stimuli parvenant aux deux yeux en créant un strabisme par section des muscles chez le chat (Hubel et Wiesel, 1965). Chaque œil est stimulé globalement autant que l'autre et peut fonctionner normalement chez l'animal adulte : pourtant, on observe une forte augmentation du nombre de cellules corticales monoculaires par rapport au cortex visuel normal du chat (qui est dans l'ensemble binoculaire, contrairement au singe). Il faut donc en conclure que le caractère *asynchrone* de l'activité (dû au fait que les yeux ne voient jamais la même portion de l'espace) au niveau des différents axones innervant un même neurone du cortex, *renforce la compétition* entre ces axones, et aboutit à une ségrégation accrue de leurs afférences. Dans le cas normal, au contraire, la coopération des signaux issus d'yeux synchrones permet de maintenir une focalisation plus large et en recouvrement (l'interprétation est la même pour la greffe d'une troisième rétine à la grenouille).

Quels sont alors les détails du mécanisme de "récompense" par coopération ? De façon générale, il est admis que les effets de l'activité électrique se trouvent dans une *corrélacion des signaux pré- et postsynaptiques*. L'importance d'une coïncidence entre terminaisons présynaptiques et cellules postsynaptiques dans la synaptogenèse est reconnue depuis longtemps, et a été invoquée en particulier par Hebb (1949) pour expliquer, dans le cadre de sa théorie, la formation d'"assemblées de cellules" destinées à coder l'information de façon distribuée. C'est ainsi qu'il énonça le principe devenu célèbre :

« When an axon of cell A is near enough to excite a cell B and repeatedly and persistently takes part in firing it, some growth or metabolic change

takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased ».

Cette idée très générale est donc celle d'un "feed-back" positif, c'est à dire d'un renforcement des connexions à travers les succès des transmissions qu'elles assurent. Un mécanisme plus précis expliquant ce phénomène a été plus tard proposé par Stent (1973) : selon Stent, les récepteurs situés au voisinage d'une synapse en activité seraient *protégés* par la coïncidence des activités pré-/post-, et, au contraire, si la synapse n'est pas active mais si le neurone postsynaptique est dépolarisé, ces récepteurs seraient déstabilisés et éliminés à cause d'une inversion de potentiel transmembranaire plus importante. Dans leur théorie de la stabilisation sélective, Changeux et Danchin (1976) ont suggéré un mécanisme de potentiation à travers une interaction biochimique entre signaux antérogrades et rétrogrades ("facteur de couplage interne"), déterminant le passage des récepteurs d'une forme labile à une forme stable. Selon Purves et Lichtman (1980), la libération postsynaptique et la capture présynaptique d'un facteur trophique nécessaire à la survie des terminaisons synaptiques serait dépendante de l'activité, en ce sens que la "récompense présynaptique" serait plus grande pour des signaux afférents coïncidents que pour les mêmes signaux décorrélés.

En conclusion, il est reconnu que la *structure temporelle précise* de l'activité joue un rôle fondamental dans la focalisation de l'innervation, et on peut en résumer la logique de la façon suivante : (a) la potentiation synaptique requiert une corrélation entre l'activité présynaptique et la dépolarisation postsynaptique; (b) or cette corrélation sera d'autant plus probable que les signaux afférents sont eux-mêmes synchronisés.

3 PRÉSENTATION DU MODÈLE

Dans le but de fournir une vision synthétique de notre modèle, nous commencerons au §3.0 par une description succincte de celui-ci, en exposant les principaux paramètres du réseau ainsi que les équations d'évolution qui caractérisent son comportement. Puis les paragraphes 3.1 et 3.2 reprendront cette présentation de façon plus détaillée, en précisant certains aspects qui seront également commentés d'un point de vue neurobiologique.

3.0 Résumé préliminaire

On considère un réseau de N neurones, excitateurs et inhibiteurs. La connectivité de ce réseau n'est pas nécessairement complète, et sera modélisée par un graphe aléatoire contenant environ pN^2 connexions, où $0 < p \leq 1$. L'échelle de temps est la milliseconde : chaque neurone i est caractérisé par un état d'activité instantané binaire $x_i(t) \in \{0, 1\}$ qui symbolise le déclenchement d'un potentiel d'action (un "spike") sur l'axone de i . Les synapses sont schématisées par des valeurs de connexions réelles $\{W_{ij}\}$ comprises entre 0 et 1 (sous la contrainte de connectivité du graphe définie plus haut) qui varient sur la même échelle de temps que les activités, leurs incréments étant toutefois plus faible que 1 (de l'ordre de 0,1). Donc à tout instant t , $W_{ij}(t)$ représente l'efficacité synaptique du neurone i vers le neurone j : elle est positive si i est excitateur et négative si i est inhibiteur. Par ailleurs, chaque connexion (i,j) est caractérisée par un délai de transmission τ_{ij} (entier supérieur à 1), tiré aléatoirement une fois pour toutes au début de la simulation.

L'état global du réseau est à chaque instant : $(\mathbf{x}(t), \mathbf{W}(t))$. Cet état évolue en fonction de trois lois qui impliquent une interaction permanente entre activités neuronales et connexions, à savoir : (a) une loi d'*activation* des neurones; (b) une loi de *potentiation* synaptique; (c) une loi de *compétition* intersynaptique. Remarquons que, dans notre modèle, seules les connexions reliant les neurones excitateurs entre eux pourront varier. Les afférences et efférences des neurones inhibiteurs auront quant à elles des valeurs constantes, déterminées au départ en fonction d'un tirage aléatoire.

(a) Un neurone j (excitateur ou inhibiteur) est caractérisé par un seuil θ_j (réel positif), et sa loi de décharge s'écrira :

$$\text{proba}[x_j(t)=1] = \frac{1}{1 + e^{-(V_j(t) - \theta_j)/T}}$$

où V_j est le potentiel qu'il reçoit à l'instant t de la part de autres neurones i :

$$V_j(t) = \sum_i W_{ij}(t) x_i(t-\tau_{ij})$$

et T un paramètre ("température") mesurant le degré d'agitation dans le réseau. La dynamique des activités neuronales est synchrone : tous les (x_i) seront remis à jour en même temps (instant t), en tenant compte de l'état du réseau aux instants précédents (instants $t-\tau_{ij}$).

(b) D'autre part, les variations des poids des connexions seront fonction des *corrélations* entre signaux pré- et postsynaptiques, selon un schéma de type hebbien. Pour modéliser ce mécanisme rétro-actif, on posera schématiquement :

$$W_{ij}(t) = W_{ij}(t-1) + A_{ij}(t)$$

avec :

- $x_i(t-\tau_{ij}) = 1$ et $x_j(t) = 1 \Rightarrow A_{ij}(t) = +\alpha$
- $x_i(t-\tau_{ij}) = 1$ et $x_j(t) = 0 \Rightarrow A_{ij}(t) = -\beta$
- $x_i(t-\tau_{ij}) = 0$ et $x_j(t) = 1 \Rightarrow A_{ij}(t) = -\beta$
- $x_i(t-\tau_{ij}) = 0$ et $x_j(t) = 0 \Rightarrow A_{ij}(t) = 0$

où α et β sont des réels positifs (typiquement : $\alpha \sim 10^{-1}$ et $\beta \sim 10^{-2}$) : α fait référence au phénomène de *potentiation* synaptique, et β au phénomène de *dépression* (α est toujours non-nul; par contre, on pourra poser $\beta=0$).

(c) La compétition sera quant à elle réalisée par l'adjonction d'un deuxième terme, $B_{ij}(t)$, qui représente le gradient d'une fonction-coût pénalisant les écarts entre les sommes de connexions (afférentes ou efférentes), et des valeurs "idéales" s_0 et s_0' . Autrement dit :

$$B_{ij}(t) = - \left(\frac{\partial H}{\partial W_{ij}} \right)_{W(t-1) + A(t)}$$

avec :

$$H(W) = \gamma \sum_i \left[\left(\sum_j W_{ij} \right) - s_0 \right]^2 + \gamma' \sum_j \left[\left(\sum_i W_{ij} \right) - s_0' \right]^2$$

Cette formulation suppose une remise à jour synchrone des connexions : la matrice A est calculée complètement avant de participer au calcul de la matrice B (mais il est également possible d'envisager un processus séquentiel, dans lequel chaque calcul de A_{ij} est immédiatement suivi par le calcul de B_{ij}).

Finalement, la dynamique des connexions s'écrira :

$$W_{ij}(t) = W_{ij}(t-1) + A_{ij}(t) + B_{ij}(t)$$

en "tronquant" par ailleurs W_{ij} à chaque instant de façon à obtenir : $0 \leq W_{ij}(t) \leq 1$.

3.1 Caractéristiques du réseau

3.1.a Neurones et connexions

On considère un réseau de neurones formel, R , composé de N neurones, excitateurs et inhibiteurs. L'état d'un neurone $i \in R$ est caractérisé par la variable x_i qui représente son *activité instantanée*. Pour simplifier, nous adopterons le formalisme binaire de McCulloch et Pitts (1943) : ainsi, à chaque instant, on ne reconnaîtra que deux états possibles à un neurone i , un état *actif*, pour lequel on posera $x_i = 1$, et un état *inactif* pour lequel on posera $x_i = 0$. L'état global du réseau R est alors représenté par le vecteur : $\mathbf{x} = (x_i)_{i \in R}$ qui pourra parcourir 2^N configurations possibles : $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \{0, 1\}^N$.

L'interprétation que nous donnons à x_i est cependant différente de celle qui prévaut habituellement, car l'échelle de temps sur laquelle nous nous plaçons est celle de la *milliseconde*, et non pas de la centaine de millisecondes. Dans la proposition classique de McCulloch et Pitts, et dans la plupart des modèles connexionnistes actuels, la variable x_i représente implicitement une *fréquence moyenne de décharge*, calculée sur une période de quelques centaines de millisecondes (pour plus de commodité, cette valeur est ensuite seuillée, donc ramenée à 0 en-deçà d'un certain seuil de fréquence, et à 1 au-delà). Dans notre modèle, en revanche, x_i schématise la *décharge instantanée* du neurone i , c'est à dire l'apparition ponctuelle d'un potentiel d'action (PA) sur le site générateur se trouvant à la base de l'axone de i , appelé également *hille axonique*. Autrement dit, le fait que $x_i = 1$ ne signifie pas que le neurone i est globalement excité pendant une période de 100ms, mais indique que ce neurone déclenche à un instant précis une dépolarisation maximale sur son hille, et, inversement, $x_i = 0$ indique que la membrane axonale a retrouvé provisoirement une polarisation normale. Bien sûr, l'instant de cette décharge n'est pas parfaitement ponctuel, et recouvre en réalité un petit intervalle de temps d'environ 1ms.

Comme on le sait, ces échelles de temps différentes sont liées par ailleurs à des interprétations différentes du réseau connexionniste. Le modèle que nous proposons situe l'activité neuronale à un niveau temporel plus fin que celui des modèles connexionnistes traditionnels car nous ne donnons pas à cette activité le rôle prédominant qu'elle reçoit d'ordinaire. Rappelons que McCulloch et Pitts voyaient essentiellement dans un neurone une *unité logique*, dont l'intérêt était de participer à la réalisation générale de fonctions booléennes. De leur côté, les modèles neuronaux actuels ont une approche très semblable, puisqu'ils assimilent les neurones à des "détecteurs de caractéristiques", et codent les objets du monde extérieur sur des assemblées d'unités binaires. Comme on l'a expliqué au chapitre 1, cette conception est donc assez différente de celle qui caractérise notre modèle : selon nous, un neurone (cellule réelle ou nœud formel d'un réseau) n'est pas chargé d'une signification particulière, mais contribue à des états macroscopiques *relationnels* par le jeu des corrélations temporelles qu'il entretient avec les autres neurones. Ainsi, nous pensons que l'activité neuronale est pertinente dans ses *variations* temporelles (codes d'ordre supérieur à 1), et non pas dans sa stabilité moyenne (code d'ordre 1).

Une même modélisation élémentaire peut donc se trouver à la base de deux théories différentes : il est toujours possible de symboliser l'activité d'un neurone réel i par une variable binaire x_i , mais cette symbolisation est "multivalente", car elle peut servir aussi bien dans un contexte statique (moyenne temporelle) que dans un contexte dynamique (corrélations temporelles). Comme on le verra plus bas, nous adopterons également la règle d'activation à seuil de McCulloch et Pitts, dans laquelle x_i vaut 0 ou 1 selon que le potentiel postsynaptique est plus grand ou plus petit qu'un seuil propre à i . Cette schématisation très simple n'aura pas non plus besoin d'être perfectionnée pour les besoins de notre modèle et suffira à produire les phénomènes d'auto-organisation que nous voulons mettre en évidence (chapitres 4 et 5).

Les neurones communiquent à travers un ensemble de contacts synaptiques qui seront schématisés par des connexions formelles $(i,j) \in R \times R$. L'efficacité de transmission du groupe des synapses qui relie un neurone présynaptique i à un neurone postsynaptique j est représentée globalement par une valeur de connexion W_{ij} . De ce fait, on conserve aussi le sens de transmission des synapses réelles : ainsi, la connexion (i,j) représente un groupe de synapses dont les terminaisons axonales appartiennent au neurone i et dont les régions réceptrices sont sur la membrane dendritique du neurone j , tandis que (j,i) représente un groupe de synapses entièrement différent, réalisant des contacts de j sur i .

L'état des connexions du réseau est représenté par la matrice $W = (W_{ij})_{i,j \in R}$: idéalement, ces connexions varient sur la même échelle de temps que l'état des activités neuronales, c'est à dire au niveau de la milliseconde. Les W_{ij} représentent donc des efficacités synaptiques *instantanées*. Etant donné un neurone i , on dira que la ligne $(W_{ij})_{j \in R}$ de cette matrice constitue l'ensemble des connexions sortantes, ou *efférences* du neurone i . De même, pour un neurone j , la colonne $(W_{ij})_{i \in R}$ constitue l'ensemble des connexions entrantes ou *afférences* de j . Comme un neurone forme rarement des synapses sur lui-même, on posera $W_{ii} = 0$ pour tout $i \in R$ (la matrice W est donc à diagonale toujours nulle).

3.1.b Excitation et inhibition

Les valeurs de connexions W_{ij} sont réelles et leur signe dépendra du neurone *présynaptique* i : W_{ij} sera positif ou négatif selon que i est respectivement excitateur ou inhibiteur. Par conséquent, un neurone excitateur ne possède que des efférences excitatrices et un neurone inhibiteur que des efférences inhibitrices, mais il n'existe pas de neurone qui possède les deux types d'efférences à la fois.

Dans notre modèle, l'information est codée par les neurones *excitateurs* et par les connexions qui relient les *excitateurs entre eux*. Les neurones inhibiteurs, ainsi que leurs efférences ou afférences, n'auront qu'un rôle secondaire, et interviendront occasionnellement pour réprimer certaines configurations d'activation : ils seront minoritaires, et leur présence installera une contrainte supplémentaire sur l'auto-organisation des autres neurones.

On notera I l'ensemble des inhibiteurs et E celui des excitateurs, leurs tailles respectives étant N_I et N_E , soit : $R = E \cup I$, et $N = N_E + N_I$ avec $N_I < N_E$. Dans toute la suite, on fera la distinction entre les 4 classes de synapses suivantes : $E \times E$, $E \times I$, $I \times E$ et $I \times I$, et, pour simplifier, on supposera que seuls les contacts $E \times E$ (contacts "interexcitateurs"), sont doués de plasticité : comme on le verra en détail au §3.2, leur variation sera fonction des corrélations temporelles entre impulsions nerveuses. Ainsi, dans le réseau de l'innervation corticale, nous nous retrouvons aux phénomènes de focalisation interexcitateurs : les CST susceptibles d'apparaître sur le réseau seront uniquement composées des connexions $E \times E$, tandis que les connexions impliquant les cellules inhibitrices, aussi bien comme cibles que comme sources, formeront un câblage fixe.

3.1.c Contraintes morphologiques

Ce développement est par ailleurs soumis à certaines contraintes anatomiques et morphologiques sur la géométrie des contacts synaptiques. Les neurones n'auront pas la possibilité de former des contacts fonctionnels avec tous les autres neurones, ne serait-ce qu'en raison de l'encombrement spatial provoqué par la croissance de milliards de fibres nerveuses. De cette constatation, nous retiendrons simplement le fait que l'architecture du réseau cortical n'est pas un graphe complet, et ceci sera traduit de façon très schématique dans le modèle par la dilution du graphe $R \times R$ en un *sous-graphe aléatoire homogène*. Donc on supposera que le réseau ne présente pas de topologie particulière : la probabilité pour qu'un neurone i forme des synapses sur j ne dépendra pas de la position relative de i et j (le graphe des connexions reste de dimension infinie : le nombre de nœuds qui peuvent être atteints à partir d'un seul nœud varie exponentiellement, et non pas géométriquement, en fonction du nombre k d'arêtes parcourues pour les atteindre, pour k petit).

Pendant, cette densité uniforme des connexions ne sera pas la même à l'intérieur des quatre ensembles de synapses mentionnés plus haut : on utilisera ainsi quatre coefficients différents, p_{EE} , p_{EI} , p_{IE} et p_{II} (réels positifs inférieurs à 1), qui seront interprétés comme des probabilités d'existence uniformes pour les connexions. Leur action se traduira formellement par le tirage aléatoire d'une "matrice de contacts" à composantes binaires, $C = (C_{ij})_{i,j \in R}$ qui indiquera s'il existe effectivement des contacts synaptiques entre les neurones de R . Le fait que $C_{ij} = 1$ signifiera que le neurone i peut innerver le neurone j , donc contribuer à son activation, et le fait que $C_{ij} = 0$ signifiera que les terminaisons axonales de i n'atteignent pas j (ou bien que les contacts membranaires ne se différencient pas en sites synaptiques), donc que j reste indifférent à l'activité de i . Dans le modèle, les coefficients binaires de C seront donc choisis de la façon suivante :

$$\forall (i,j) \in E \times E, \text{ proba}[C_{ij}=1] = p_{EE} \text{ et } \text{proba}[C_{ij}=0] = 1-p_{EE}$$

et il en va de même pour les trois autres catégories de connexions, $E \times I$, $I \times E$ et $I \times I$. Ainsi, la matrice C formalise l'"enveloppe génétique" du réseau : en particulier, les réarrangements synaptiques d'un neurone excitateur i vers un neurone j ne pourront avoir lieu que sous la condition $C_{ij} = 1$. Rappelons qu'avec la non-symétrie naturelle du réseau il est tout à fait possible que $C_{ij} = 1$ et $C_{ji} = 0$ (la cellule i innerve la cellule j , mais pas l'inverse). D'autre part, l'interdiction de former des synapses sur soi-même se traduit encore par : $C_{ii} = 0$.

En résumé, C est une matrice binaire divisée en quatre secteurs qui se distinguent par des densités de "1" différentes. Les nombres moyens de contacts dans chacun de ces quatre ensembles de synapses sont : $p_{EE}N_E(N_E-1)$, $p_{EI}N_EN_I$, $p_{IE}N_IN_E$ et $p_{II}N_I(N_I-1)$, leur somme étant inférieure au nombre total maximum de contacts $(N_E+N_I)(N_E+N_I-1) = N(N-1)$ (obtenus pour les quatre p égaux à 1). Cependant, pour simplifier, nous supposons dans toute la suite que les neurones inhibiteurs n'ont pas le pouvoir de s'inhiber mutuellement, et nous poserons par conséquent : $p_{II} = 0$. Leurs efférences seront donc dirigées exclusivement vers les neurones excitateurs, et leurs afférences proviendront exclusivement de ces mêmes neurones. Finalement, chaque inhibiteur innerve en moyenne $p_{IE}N_E$ excitateurs, mais aucun inhibiteur; un neurone excitateur, quant à lui, innerve en moyenne $p_{EE}(N_E-1)$ excitateurs et $p_{EI}N_I$ inhibiteurs.

Etant donné cette architecture de contacts (C_{ij}), les valeurs réelles des efficacités de transmission (W_{ij}) sont alors soumises à la contrainte suivante :

$$\forall (i,j) \in R \times R, C_{ij} = 0 \Rightarrow W_{ij} = 0$$

Avec ces valeurs obligatoirement nulles, la matrice W sera utilisée dans son ensemble pour calculer les potentiels postsynaptiques (PPS) comme s'il existait virtuellement N^2 connexions, en effectuant la combinaison linéaire $W.x$. Les connexions W_{ij} autorisées par $C_{ij} = 1$ auront des domaines de variation différents dans chacune des trois combinaisons possibles, $E \times E$, $E \times I$ et $I \times E$: ce seront des intervalles $[w_1, w_2]$ dont les bornes dépendront de la nature de i et j . La situation est résumée de la façon suivante :

- $i \in E$ et $j \in E \Rightarrow 0 \leq W_{ij} \leq C_{ij}$
- $i \in E$ et $j \in I \Rightarrow 0 \leq w_{1EI} C_{ij} \leq W_{ij} \leq w_{2EI} C_{ij}$
- $i \in I$ et $j \in E \Rightarrow w_{1IE} C_{ij} \leq W_{ij} \leq w_{2IE} C_{ij} \leq 0$
- $i \in I$ et $j \in I \Rightarrow C_{ij} = W_{ij} = 0$

Donc, pour les connexions $E \times E$ telle que $C_{ij} = 1$, la borne inférieure 0 représente une *absence* de transmission, tandis que la borne supérieure 1 représente une *saturation* de la transmission. Les connexions invariables $E \times I$ et $I \times E$ seront

distribuées au départ sur le réseau, en fonction d'un tirage aléatoire uniforme à l'intérieur de leurs domaines respectifs d'appartenance. Si on note w_0 la valeur centrale d'un intervalle $[w_1, w_2]$ et Δw sa largeur, on prendra par exemple : $w_{0EI} = +1$ et $w_{0IE} = -5$, avec $\Delta w_{EI} = \Delta w_{IE} = 1$ (sachant que l'intervalle des connexions $E \times E$ a été normalisé à $[0, 1]$).

Dans ce modèle, nous nous intéresserons donc particulièrement à la sous-matrice $N_E \times N_E$ de \mathbf{W} qui est constituée des connexions interexcitatrices, et que l'on notera \mathbf{W}^{EE} . La sous-matrice correspondante dans \mathbf{C} sera notée de même \mathbf{C}^{EE} , et la contrainte géométrique sur ces connexions peut alors s'écrire en abrégé : $0 \leq \mathbf{W}^{EE} \leq \mathbf{C}^{EE}$, où 0 représente une matrice $N_E \times N_E$ uniformément nulle. Une connexion de $E \times E$ telle que $C_{ij} = 1$ sera qualifiée de *forte* si W_{ij} est proche de 1, et *faible* si W_{ij} est proche de 0. Formellement, \mathbf{W}^{EE} appartient à l'hypercube réel de dimension N_E^2 compris entre 0 et 1 : $[0, 1]^{N_E \times N_E}$. Mais, plus précisément, la contrainte imposée par \mathbf{C}^{EE} restreint \mathbf{W}^{EE} à l'intérieur d'un sous-hypercube K défini par : $\mathbf{W}^{EE} \in K \Leftrightarrow 0 \leq \mathbf{W}^{EE} \leq \mathbf{C}^{EE}$.

3.1.d Variables temporelles et délais

L'axe des temps est discrétisé pour les besoins des simulations numériques : on le divise régulièrement en intervalles élémentaires de l'ordre d'une milliseconde. Les variables temporelles du modèle sont les activités des neurones $x(t)$ (on notera en particulier $x^E(t)$ les activités des neurones excitateurs et $x^I(t)$ celles des inhibiteurs) et les valeurs des synapses interexcitatrices $\mathbf{W}^{EE}(t)$. A tout instant t , l'état du réseau R est donc représenté par l'ensemble des valeurs $(x^E(t), x^I(t), \mathbf{W}^{EE}(t))$. Cet état global est entièrement recalculé entre deux instants consécutifs t et $t+1$, et on dira que ce calcul constitue une *itération élémentaire* de l'évolution du réseau.

Les connexions synaptiques sont caractérisées par des *délais de transmission* non-uniformes sur le réseau. Le délai particulier d'une connexion (i, j) correspond au temps total qui s'est écoulé entre le déclenchement d'un PA sur le hille axonique de i , et l'arrivée du PPS provoqué par ce signal au centre d'intégration somatique de j , c'est à dire juste avant que j ne s'apprête à décharger à son tour. Ce délai de transmission dépend à la fois des neurones pré- et postsynaptique, et on le notera τ_{ij} . En général, τ_{ij} peut prendre des valeurs assez différentes d'une connexion à l'autre, et pourra varier de 1 milliseconde à 15 ou 20 millisecondes : dans notre modèle, un délai est donc un entier strictement positif compris entre 1 et 20. Comme les autres paramètres, les valeurs des délais seront tirées aléatoirement dans des intervalles qui dépendront du type des neurones (en général, les transmissions inhibitrices $I \times E$ seront plus rapides que les transmissions excitatrices $E \times E$ ou $E \times I$).

Ce délai global τ_{ij} est un paramètre schématique qui inclut en réalité plusieurs composantes temporelles : la durée de propagation sur l'axone de i (délai axonal), la durée de la transmission synaptique de i vers j (délai synaptique), et la durée de propagation entre les sites récepteurs dendritiques sur j et le site générateur de j

(délai somato-dendritique). Le délai de propagation axonal est assez court, et en général dépasse rarement une milliseconde : les axones relativement longs (plus d'un millimètre) sont entourés d'une gaine de myéline qui augmente considérablement leur vitesse de conduction. On peut donc considérer qu'un PA engendré par i atteint toutes les terminaisons axonales simultanément. Les délais des transmissions synaptiques sont eux aussi très courts, et pour ainsi dire tous identiques : le temps nécessaire à la libération des molécules de neurotransmetteur dans la fente synaptique et à leur réception sur la membrane postsynaptique est également de l'ordre d'une milliseconde. Finalement, des trois composantes de τ_{ij} , c'est le délai somato-dendritique qui est le plus long et surtout le plus variable : avant d'atteindre le centre intégrateur, la propagation des potentiels postsynaptiques le long des ramifications dendritiques et du soma peut prendre des formes très diverses (potentiels décrémenteurs relativement lents, impliquant des diffusions d'ions et de seconds messagers, ou au contraire PA dendritiques rapides). De plus, les contacts synaptiques (W_{ij}) afférents à j occupent des positions très différentes sur l'arbre dendritique de ce neurone. Le délai τ_{ij} dépendra donc essentiellement des propriétés de membrane au site synaptique (i,j) et de la distance qui le sépare du hille axonique de j . En résumé, τ_{ij} est composé d'environ une milliseconde axonale, une milliseconde synaptique, et plusieurs millisecondes somato-dendritiques : on notera cette dernière quantité τ_{ij}' , et on a donc schématiquement $\tau_{ij} \approx 2 + \tau_{ij}'$.

Dans ce cadre, l'état d'activité d'un neurone j (excitateur ou inhibiteur), subira l'influence de l'état d'activité de chaque fibre afférente i avec un retard de τ_{ij} millisecondes propre à la jonction (i,j) : autrement dit, $x_j(t)$ dépendra des états $\{x_i(t-\tau_{ij})\}_{i \in R}$ (les termes de cette dépendance seront détaillés au §3.2.a). De plus, il dépendra aussi de l'état de toutes les afférences (i,j) (excitatrices ou inhibitrices) par lesquelles transitent ces signaux : il s'agit donc des valeurs W_{ij} prises aux instants $t-\tau_{ij}'$ où les potentiels postsynaptiques sont engendrés sur les sites (i,j) : $\{W_{ij}(t-\tau_{ij}')\}_{i \in R}$. Cependant, par souci de simplification, les W_{ij} seront alignés sur un instant commun qui sera choisi égal à t , c'est à dire l'instant de la décharge de j . Ceci constitue une approximation tout à fait acceptable dans la mesure où les variations des connexions W^{EE} sont relativement lentes comparées aux fluctuations des activités : en effet, comme on le verra au §3.2.b, la modulation synaptique entre neurones excitateurs procède par incréments et décréments qui restent petits devant la valeur maximale 1. Donc en général, $W_{ij}^{EE}(t-\tau_{ij}')$ pourra être assimilé à $W_{ij}^{EE}(t)$ sans grande erreur, ce qui fait que $x_j(t)$ dépendra globalement de $\{W_{ij}(t)\}_{i \in R}$.

De leur côté, les synapses W^{EE} auront un comportement de type hebbien, c'est à dire qu'elle varieront en fonction des *corrélations temporelles* récentes entre neurones pré- et postsynaptiques, calculées avec les décalages τ_{ij} . Ainsi, pour tout $(i,j) \in E \times E$, le changement de $W_{ij}(t)$ dépendra des couples d'états $(x_j(t'), x_i(t'-\tau_{ij}))$ en des instants t' compris dans l'intervalle $]t-\Delta t, t]$. La durée Δt représente la période d'intégration de la corrélation : elle est plus petite que les délais (de l'ordre de quelques millisecondes) et sera généralement réduite à 1. Cette dépendance

hebbienne à travers le délai total τ_{ij} est le résultat d'une schématisation primaire dont les détails seront développés au §3.2.b et au §3.2.c.

Donc, en résumé, si on note τ_{\min} le plus petit délai du réseau et τ_{\max} le plus grand délai, l'état d'activation $x(t) = (x^E(t), x^I(t))$ dépend de tous les états antérieurs $x(t-\tau_{\min}), \dots, x(t-\tau_{\max})$, et de toutes les connexions $W(t)$; par ailleurs, pour $\Delta t = 1$, l'état des connexions $W^{EE}(t)$ dépend de $x(t), x(t-\tau_{\min}), \dots, x(t-\tau_{\max})$ ainsi que de l'état précédent $W^{EE}(t-1)$ (le processus conjoint (x, W) est markovien).

3.2 Equations d'évolution

Nous abordons maintenant les règles précises d'évolution des états $x^E(t), x^I(t)$ et $W^{EE}(t)$, c'est à dire la façon dont les activités neuronales et les efficacités synaptiques s'influencent mutuellement. Comme nous l'avions annoncé au paragraphe 3.1, les afférences des neurones inhibiteurs W^{EI} (qui sont excitatrices), ainsi que leurs efférences W^{IE} (qui innervent exclusivement les neurones excitateurs), forment une configuration nerveuse fixe. On écrira donc :

$$\forall (i,j) \in E \times I, \forall (i,j) \in I \times E, \forall t \geq 0, W_{ij}(t) = W_{ij}(0).$$

où l'instant 0 marque symboliquement le début de l'évolution du réseau. On rappelle que ces connexions sont aléatoires et valent en moyenne $w_{0EI} \geq 0$ et $w_{0IE} \leq 0$ (il s'agit bien sûr des connexions pour lesquelles $C_{ij} = 1$).

Le câblage des neurones inhibiteurs étant achevé, nous nous intéresserons au développement des contacts synaptiques entre neurones excitateurs, $W^{EE}(t)$, à travers les deux phénomènes fondamentaux qui dirigent leur évolution : la *potentiation* et la *compétition*. Ces phénomènes neurobiologiques font respectivement référence à l'augmentation des efficacités synaptiques par la transmission des signaux électriques, et aux contraintes trophiques ou géométriques limitant le nombre des contacts afférents et efférents. Dans notre modèle, ils seront intégrés sous la forme de deux termes de variation, A_{ij} et B_{ij} , qui viendront s'ajouter à chaque étape aux connexions W_{ij} . Au §3.2.b, nous présenterons la formulation schématique du mécanisme de potentiation qui sera adoptée dans toute la suite : en résumé, le terme A_{ij} est un incrément ou un décrément appliqué à la connexion (i,j) , qui dépendra de la corrélation des activités x_i et x_j avec le déphasage τ_{ij} . Puis, au §3.2.c, nous verrons en quoi cette formulation représente une simplification des données neurobiologiques, et comment la potentiation synaptique devrait être en réalité modélisée à un niveau *local* : l'augmentation des efficacités dépend des dépolarisations locales provoquées par des groupements de synapses indépendants, et ces dépolarisations dépendent elles-mêmes de la coopération des signaux à l'intérieur de chaque groupement. Cependant, bien que plus complète et plus réaliste, cette deuxième version et les complications qu'elle

apporte seront abandonnées au profit de l'interprétation simplifiée du §3.2.b. Enfin, au §3.2.d, on modélisera la compétition intersynaptique à travers une loi de conservation des sommes d'afférences et d'efférences au cours du temps. Cette contrainte sera traduite dans une fonction-coût H de type quadratique, dont la minimisation sera équivalente à la normalisation de ces sommes à une valeur idéale s_0 : pour cela, on utilisera la méthode du gradient, qui sera appliquée aux connexions à travers le terme B_{ij} . Enfin, le §3.2.e récapitulera l'action formelle des matrices A et B sur W à l'intérieur des limites de l'hypercube K . Mais tout d'abord, au §3.2.a, nous rappelons la loi classique d'activation des neurones, utilisée sous diverses variantes dans la quasi-totalité des modèles connexionnistes, en y adjoignant pour notre part les délais de transmission introduits au §3.1.d.

3.2.a Activités neuronales

L'état d'activité d'un neurone j , $x_j(t)$, c'est à dire le fait qu'il produit ou non un potentiel d'action à l'instant t , obéira à une loi probabiliste qui dépend avant tout du potentiel postsynaptique total induit par les autres neurones à travers ses afférences. Cette loi est la même pour tous les neurones, et il n'y aura pas lieu de faire la distinction entre $j \in E$ et $j \in I$: les caractéristiques différentes des excitateurs et des inhibiteurs sont inscrites dans les valeurs de leurs efférences (positives ou négatives), dans les différents délais des connexions, ainsi que dans leurs seuils.

On rappelle que la transmission des potentiels d'action à travers les synapses excitatrices et inhibitrices provoque des dépolarisations et hyperpolarisations locales, en différents endroits de la membrane réceptrice du neurone. Celles-ci se propagent le long des ramifications dendritiques et convergent vers le corps cellulaire où elles sont intégrées et produisent une dépolarisation globale. Si l'intensité de cette dernière est suffisante, c'est à dire dépasse un certain *seuil* de déclenchement, le neurone déclenche à son tour un PA qui se propage très rapidement sur son axone en direction des cellules cibles, et contribue à l'activation ou l'inhibition de celles-ci. Contrairement aux PPS, les PA sont constitués de *spikes* ("pics de décharge") très brefs, d'une durée inférieure à 2ms, et d'amplitude constante, de l'ordre de +70mV (la polarisation normale de la membrane est à environ -40mV). Il est remarquable que ces spikes ne subissent aucune altération notable au cours de leur propagation sur l'axone : leur base ne s'étale pas dans le temps, et leur amplitude ne faiblit pas.

Pour modéliser ce processus, nous reprendrons le schéma classique du *neurone binaire à seuil*, proposé par McCulloch et Pitts (1943). Cependant, comme nous l'avons déjà fait remarquer au §3.1, l'apparente identité des équations ne devra pas faire oublier les différences d'échelle de temps qui existent entre notre modèle et la plupart des modèles connexionnistes traditionnels. Bien sûr, dans tous les cas, un neurone peut être grossièrement schématisé par un sommateur linéaire à seuil, mais ici, la réponse binaire $x_j(t)$ représente l'activité instantanée du neurone j , et

non pas son activité moyenne : ainsi, $x_i(t) = 1$ formalise le fait qu'un PA de +70mV est engendré à la racine de l'axone à l'instant t , et $x_j(t) = 0$ signifie que la membrane de l'axone est polarisée normalement à cet instant, aux alentours de -40mV. La validité de l'approximation $x_j(t) = 0$ ou 1 concerne donc un intervalle de temps très court, $[t, t+1]$, d'une milliseconde environ, alors que dans la majorité des modèles, cet intervalle est beaucoup plus long et vaut quelques centaines de millisecondes (sur lesquelles le signal est moyenné et binarisé).

Par conséquent, $x_j(t)$ dépend d'un PPS *instantané*, noté $V_j(t)$, qui parvient au centre d'intégration somatique à l'instant t . Ce potentiel global est la somme d'une multitude de potentiels locaux v_{ij} apparus au niveau des dendrites, qui se trouvent converger au même moment sur le soma, chaque v_{ij} étant issu de la transmission d'un PA engendré par un neurone i à travers la connexion W_{ij} (si W_{ij} est excitatrice, v_{ij} est un "PPSE", et si W_{ij} est inhibitrice, v_{ij} est un "PPSI"). Pour compléter le schéma de causalité qui va de x_i à x_j , il faut encore remonter dans le temps jusqu'au moment de la naissance du potentiel d'action x_i sur le hille axonique de i . Comme on l'a vu au §3.1.d, le délai total qui s'est alors écoulé entre l'émergence d'un spike en i et ses conséquences sur l'émergence d'un spike en j est noté τ_{ij} , et sa composante somato-dendritique τ_{ij}' . Donc, en toute rigueur, on devra écrire :

$$v_{ij}(t-\tau_{ij}') = W_{ij}(t-\tau_{ij}') x_i(t-\tau_{ij})$$

ce qui signifie qu'un PA engendré sur l'axone i à l'instant $t-\tau_{ij}$ est transmis sur la connexion W_{ij} à l'instant $t-\tau_{ij}'$ (c'est à dire 1 ou 2 millisecondes plus tard), et que l'amplitude du PPS local v_{ij} qui en résulte sur j est proportionnelle à l'efficacité de cette connexion (le coefficient de proportionalité étant déjà inclus dans W_{ij}). S'il n'y a pas de transmission, ce potentiel vaut 0. L'intégration de tous les v_{ij} avec leurs délais de convergence respectifs τ_{ij}' fournit alors le PPS global $V_j(t)$:

$$V_j(t) = \sum_{i \in R} v_{ij}(t-\tau_{ij}') = \sum_{i \in R} W_{ij}(t-\tau_{ij}') x_i(t-\tau_{ij})$$

V_j est donc une combinaison linéaire des efficacités synaptiques W_{ij} , pondérées par des activités instantanées valant 0 ou 1. Dans toute la suite, pour plus de commodité, on fera l'approximation annoncée au §3.1.d, c'est à dire que les W_{ij} seront pris au *même instant* t :

$$V_j(t) = \sum_{i \in R} W_{ij}(t) x_i(t-\tau_{ij})$$

On verra en effet que la modulation synaptique est suffisamment lente pour autoriser un tel décalage. Finalement, le neurone binaire j décide de déclencher un PA selon que V_j est plus grand ou plus petit qu'un seuil de déclenchement θ_j propre à ce neurone. Si l'on s'en tient à un déterminisme strict, le spike sera engendré à la

racine de l'axone j si et seulement si V_j est plus grand que θ_j :

$$\begin{aligned} x_j(t) &= 1 \text{ si } V_j(t) \geq \theta_j \\ x_j(t) &= 0 \text{ si } V_j(t) < \theta_j \end{aligned}$$

Dans ce modèle, les seuils formels sont distribués aléatoirement sur les neurones, leurs valeurs étant à nouveau tirées dans un intervalle uniforme $[\Theta_1, \Theta_2]$ (les neurones excitateurs et les neurones inhibiteurs se distinguant par des domaines de seuils différents). Notons que ces valeurs sont toujours positives ici, puisque, par convention, $V_j = 0$ correspond à une polarisation normale de la membrane (c'est à dire environ -40mV dans la réalité) : dans ce contexte, un seuil négatif voudrait dire que le neurone peut décharger spontanément en l'absence de potentiel, et doit être hyperpolarisé, c'est à dire inhibé, pour redevenir silencieux.

Cette loi de décharge peut encore s'écrire : $\text{proba}[x_j(t)=1] = h(V_j(t) - \theta_j)$, où h est la fonction de Heavyside, caractéristique des réels positifs (elle vaut 1 si son argument est positif, et 0 s'il est négatif). Cependant, le processus peut être rendu véritablement *probabiliste* en remplaçant h par une fonction continue f à valeurs réelles dans l'intervalle $[0, 1]$: $\text{proba}[x_j(t)=1] = f(V_j(t) - \theta_j)$. Cette fonction doit alors tenir compte de la force du potentiel V_j , c'est à dire de son éloignement du seuil : plus $V_j - \theta_j \geq 0$ est grand, plus f doit être proche de 1, et inversement, plus $\theta_j - V_j \geq 0$ est grand, plus f doit être proche de 0. Le plus simple est d'adopter une fonction croissante entre 0 et 1, et l'exemple communément choisi pour f est une *sigmoïde*, que l'on notera ici σ :

$$\sigma_T(z) = \frac{1}{1 + e^{-z/T}}$$

Cette fonction est un lissage de la fonction h : elle a aussi la forme d'une "marche" avec deux paliers asymptotiques en 0 et 1, mais ses bords sont arrondis et la pente de la transition de 0 à 1 est finie. Elle possède un centre de symétrie en $z = 0$ et dépend d'un paramètre T strictement positif qui règle la raideur de sa pente en ce point : $(d\sigma_T/dz)(0) = 1/4T$. A la limite où T tend vers 0, on retrouve la fonction h (sauf en $z = 0$) : T est appelé *température* par analogie avec la physique. La formule finale de $x_j(t)$ est donc donnée par :

$$\begin{aligned} \text{proba}[x_j(t)=1] &= \sigma_T(V_j(t) - \theta_j) \\ &= \frac{1}{1 + e^{-(V_j(t) - \theta_j)/T}} \\ &= \frac{1}{1 + \exp\left(-\frac{1}{T} \left(\sum_{i \in R} W_{ij}(t)x_i(t-\tau_{ij}) - \theta_j\right)\right)} \end{aligned}$$

En plus de cette loi de décharge probabiliste, nous tiendrons également compte de la propriété de réfraction des neurones réels. Après chaque événement $x_j(t) = 1$, c'est à

dire chaque apparition d'un PA sur j , la membrane axonale de ce neurone traverse une phase au cours de laquelle elle ne peut plus être dépolarisée, quel que soit le potentiel postsynaptique qui parvient au soma. Cette phase est appelée *période réfractaire*, et dure quelques millisecondes. Donc, dans notre modèle, chaque neurone j de R sera caractérisé par un nombre entier positif ou nul ρ_j représentant sa période réfractaire, qui affectera l'état x_j à l'instant t de la façon suivante :

- si $\exists t' = t - \rho_j \dots t - 1$, tel que $x_j(t') = 1$, alors $x_j(t) = 0$ obligatoirement
- si $\forall t' = t - \rho_j \dots t - 1$, $x_j(t') = 0$, alors $x_j(t) = 1$ avec la probabilité $\sigma_T(V_j(t) - \theta_j)$

Si $\rho_j = 0$, la membrane n'est jamais réfractaire et peut décharger à tout instant avec la probabilité $\sigma_T(V_j(t) - \theta_j)$. De même que les seuils θ_j , les paramètres ρ_j peuvent être différents d'un neurone à l'autre, et seront tirés aléatoirement dans un intervalle constant $[P_1, P_2]$ (bornes différentes selon que j est excitateur ou inhibiteur).

3.2.b Modèle simple de potentiation synaptique

La potentiation synaptique à long terme, ou "LTP", est un phénomène rapide et persistant. L'expression de la LTP semble résulter dans la plupart des cas d'une modification au niveau présynaptique, c'est à dire d'une augmentation de la quantité de neuro-transmetteur libéré. Cependant, on pense que ce phénomène fait également intervenir des modifications postsynaptiques, telles qu'une diminution de la résistance de l'épine dendritique ainsi qu'une augmentation du nombre de récepteurs sur la membrane postsynaptique (ces récepteurs existeraient déjà mais de façon "occulte", et seraient alors révélés au moment de la LTP). Comme nous l'avons évoqué plus haut (§2.3.b), divers mécanismes ont été proposés à la suite du principe de Hebb (Stent, 1973; Changeux et Danchin, 1976; Purves et Lichtman, 1980). Depuis une période récente, on pense que la potentiation synaptique fait intervenir des récepteurs "NMDA" (N-méthyl-D-aspartate), associés à des canaux ioniques contrôlés par le glutamate (neurotransmetteur excitateur très répandu dans le système nerveux) : un influx de calcium (Ca^{2+}) à travers ces canaux résulterait en une augmentation transitoire et localisée de sa concentration à l'intérieur de l'épine dendritique, ce qui déclencherait une série de réactions ayant pour conséquence un accroissement de l'efficacité synaptique ("spine model").

Dans tous les cas, comme nous l'avons fait remarquer au §2.3.b, la condition de potentiation est à rechercher principalement dans une "spécificité spatio-temporelle" de l'activité électrique, c'est à dire une *synchronie des signaux pré- et postsynaptiques*. Ceci est attesté par des expériences au niveau cellulaire : par exemple, un blocage de l'activité de la cellule ("voltage-clamp") empêche la LTP de se produire, tandis que la LTP peut être au contraire induite par la production d'une activité présynaptique en corrélation avec une activité postsynaptique.

Par ailleurs, on pense qu'il existe également un phénomène inverse de

dépression synaptique à long terme (LTD), qui serait la conséquence d'un couplage entre des signaux excitateurs afférents et une hyperpolarisation de la membrane postsynaptique.

De tous ces phénomènes, nous retiendrons le principe général d'une "modulation rétro-active" en fonction des décharges pré- et postsynaptiques, que nous modéliserons très schématiquement de la façon suivante : le fait qu'un neurone j décharge un PA $x_j = 1$ grâce à un PPS V_j élevé, aura pour conséquence de renforcer toutes les afférences excitatrices qui avaient participé à ce potentiel en transmettant un PA $x_i = 1$, c'est à dire d'augmenter les valeurs de connexions W_{ij} correspondantes (*potentiation*). Au contraire, une mauvaise synchronisation entre la décharge de i et la réponse de j (excitation de i non-suivie d'une excitation de j , ou bien décharge de j alors que i était silencieux) aura pour conséquence d'affaiblir le contact (i,j), ce qui se traduira par la diminution de W_{ij} (*dépression*).

Ainsi, en tenant compte du délai τ_{ij} inhérent à la connexion (i,j), nous supposerons que la variation de W_{ij} répercute *quasi instantanément* le degré de corrélation entre l'activité du neurone i et l'activité du neurone j , c'est à dire la corrélation entre les deux processus suivants :

$$(x_j(t-\Delta t+1), \dots, x_j(t-1), x_j(t)) \text{ et } (x_i(t-\tau_{ij}-\Delta t+1), \dots, x_i(t-\tau_{ij}-1), x_i(t-\tau_{ij}))$$

où Δt représente une période de quelques millisecondes. Pour cela, on définira un terme A_{ij} qui sera ajouté à W_{ij} tous les Δt , soit :

$$W_{ij}(t) = W_{ij}(t-\Delta t) + A_{ij}(t)$$

avec :

$$A_{ij}(t) = \frac{1}{\Delta t} \sum_{t'=t-\Delta t+1}^t \alpha [x_i(t'-\tau_{ij})=x_j(t')=1] - \beta [x_i(t'-\tau_{ij}) \neq x_j(t')]$$

où α et β sont deux coefficients positifs, fractions de la valeur maximale 1 (typiquement : $\alpha \sim 10^{-1}$ et $\beta \sim 10^{-2}$). Les propositions logiques entre crochets valent 1 si elles sont vraies et 0 si elles sont fausses : comme elles s'excluent mutuellement, il ne peut y avoir que 3 possibilités de variation à chaque instant, $+\alpha$, $-\beta$ ou 0, et par conséquent : $-\beta \leq A_{ij} \leq \alpha$. Ainsi, avec une unité de temps Δt , la modulation synaptique est un peu moins rapide que les activations neuronales.

Par cette règle, l'augmentation de W_{ij} est donc proportionnelle au nombre d'événements "réussis" $x_i(t'-\tau_{ij})=1 \rightarrow x_j(t')=1$ (que nous noterons plus simplement $i|1 \rightarrow j|1$), apparus dans une petite fenêtre de temps $]t-\Delta t, t]$, et inversement, sa diminution est fonction des événements de transmission "ratés" tels que $i|1 \rightarrow j|0$ ou $i|0 \rightarrow j|1$. Dans le cas d'un "non-événement" $i|0 \rightarrow j|0$, c'est à dire si aucun des deux neurones ne décharge, W_{ij} ne varie pas. Remarquons dès à présent que les

éventuelles diminutions de W_{ij} seront également contrôlées par la *compétition* intersynaptique (cf. §3.2.d) : cette compétition mettra en relation les augmentations avec les diminutions, de sorte que les incréments $+\alpha$ récompensant une connexion (i,j) se feront nécessairement au détriment des connexions (i,j') ou (i',j) qui n'ont pas transmis d'événement $1 \rightarrow 1$.

Pour simplifier les calculs, nous poserons dans toute la suite : $\Delta t = 1$. Dans ce cas, les synapses W_{ij} varient sur la même échelle de temps que les activités et répondent instantanément à l'événement de transmission le plus récent : $x_i(t-\tau_{ij}) \rightarrow x_j(t)$ (propriété de "*plasticité immédiate*"). On a donc : $W_{ij}(t) = W_{ij}(t-1) + A_{ij}(t)$, où le terme A_{ij} est réduit à une seule contribution, $+\alpha$, $-\beta$ ou 0 :

$\forall (i,j) \in E \times E,$

- $x_i(t-\tau_{ij}) = 1$ et $x_j(t) = 1 \Rightarrow A_{ij}(t) = +\alpha$
- $x_i(t-\tau_{ij}) = 1$ et $x_j(t) = 0 \Rightarrow A_{ij}(t) = -\beta$
- $x_i(t-\tau_{ij}) = 0$ et $x_j(t) = 1 \Rightarrow A_{ij}(t) = -\beta$
- $x_i(t-\tau_{ij}) = 0$ et $x_j(t) = 0 \Rightarrow A_{ij}(t) = 0$

En conclusion, toute connexion $(i,j) \in E \times E$ (autorisée par $C_{ij} = 1$), est modulée rétro-activement à chaque milliseconde par le succès des transmissions dont elle est le siège. Dans la version "immédiate" du modèle ($\Delta t = 1$), cette connexion augmente de α à la suite d'un événement $i|1 \rightarrow j|1$, diminue de β à la suite de $i|1 \rightarrow j|0$ ou $i|0 \rightarrow j|1$, et ne change pas pour $i|0 \rightarrow j|0$.

On voit donc que la potentiation liée à la synchronie des décharges "pré/post" suppose implicitement une coopération entre décharges présynaptiques pour créer un PPS global suffisamment fort sur j .

3.2.c Modèle des groupements synaptiques

Comme on l'a fait remarquer plus haut, ce modèle de variation synaptique est une schématisation très élémentaire des phénomènes de potentiation/dépression, la simplification la plus nette se situant au niveau du calcul des corrélations : en effet, à travers le terme A_{ij} , on fait dépendre les variations synaptiques locales ΔW_{ij} de la décharge d'un PA x_j , provoqué par un PPS global, dont le site générateur se trouve "en aval" sur j , relativement éloigné des synapses. Dans une version plus réaliste, il conviendrait donc plutôt de relier cette variation au succès de la *dépolarisation dendritique locale* qu'au succès de l'activation globale de j : ainsi, le mécanisme rétro-actif n'aurait pas à attendre un signal tardif pour s'exprimer. Cependant, il est par ailleurs important de conserver le principe d'une *coopération* entre signaux afférents en provenance de différents neurones. A l'inverse, on ne peut donc pas restreindre la corrélation à un événement individuel $x_i \rightarrow v_{ij}$, où $v_{ij} = W_{ij}x_i$ est un PPS local ne concernant que la connexion W_{ij} : il faut que la réponse de j dépende de

façon *non-linéaire* des stimulations afférentes (x_i).

C'est pourquoi, si l'on veut retrouver une dépendance locale "pré/post" tout en conservant une dépendance collective "pré/pré", il faut envisager un *niveau intermédiaire* entre d'une part la connexion individuelle W_{ij} et d'autre part l'ensemble de toutes les connexions afférentes de j , $(W_{ij})_{i \in R}$: ce niveau est celui du "*groupement synaptique*". Ainsi, au lieu de connexions séparées (i,j) , la configuration de l'innervation sur un neurone j sera décrite en termes de groupements locaux (G, j) , où $G = \{i, i', i'', \dots\}$ représente l'ensemble des neurones présynaptiques dont les contacts sur j sont concentrés sur une même région de la membrane (pour simplifier, on fera l'hypothèse qu'un neurone i ne peut participer qu'à un seul groupe afférent sur j : dans le cortex, il est en effet assez rare qu'un neurone i innerve plusieurs fois un neurone j). Ces groupements sont situés en différents endroits de l'arbre dendritique, et n'interagissent pas entre eux du point de vue de la modulation : ce sont des "satellites" d'arrivée indépendants sur j , qui unissent leur action seulement dans le but de faire décharger j .

Dans ce cadre, il devient difficile de délimiter un effet v_{ij} propre à la connexion (i,j) et de le séparer des contacts (i',j) appartenant au même groupement G . Tous ces contacts contribuent en réalité à un PPS commun v_{Gj} , qui dépend de façon *non-linéaire* de l'ensemble des activités $(x_i)_{i \in G}$ et des connexions $(W_{ij})_{i \in G}$: v_{Gj} n'est donc pas simplement la somme des anciens v_{ij} sur $i \in G$, mais c'est une fonction d'"ordre élevé" sensible à la *simultanéité* des PA x_i afférents.

En revanche, du point de vue de la potentiation, les connexions W_{ij} gardent leur autonomie à l'intérieur du groupement G , et dépendent alors des corrélations $x_i \rightarrow v_{Gj}$: si $x_i = 1$ et $v_{Gj} > 0$, la connexion W_{ij} sera augmentée proportionnellement à v_{Gj} ; inversement, si $x_i = 1$ et $v_{Gj} = 0$, ou si $x_i = 0$ et $v_{Gj} > 0$, elle sera diminuée. Ainsi, le renforcement de la connexion (i,j) dépend de la dépolarisation locale à laquelle elle participe, mais cette dépolarisation dépend elle-même de l'arrivée *synchrone* de plusieurs potentiels d'actions sur un ensemble de synapses (i',j) situées dans le voisinage immédiat de (i,j) . Autrement dit, un neurone présynaptique i n'est pas seul pour dépolariser localement j , mais il doit coopérer avec d'autres neurones présynaptiques i' dont les contacts sont adjacents aux siens.

Quant à la décharge de j , elle reste subordonnée à la convergence de tous les potentiels locaux v_{Gj} , en provenance des différents groupements G , chacun étant caractérisé par un unique délai somato-dendritique τ_{Gj}' : donc x_j dépend de V_j qui est la somme de tous les v_{Gj} sur G avec leurs retards respectifs τ_{Gj}' . Pour obtenir x_j en fonction des x_i , il faut donc ajouter un retard constant, d'environ 2ms, qui correspond à la somme du délai axonal et du délai synaptique pour chaque fibre i et ses contacts (i,j) . La loi d'activation de j est donc la même qu'au §3.2.a, en remplaçant la somme sur i par une somme sur G .

En résumé, grâce à ce modèle des groupements, moins simpliste que celui des connexions indépendantes, on réussit à concilier l'idée d'une potentiation locale

avec celle d'une coopération entre plusieurs neurones présynaptiques différents. Cette coopération s'effectue à deux niveaux : au niveau du groupement local pour la dépolarisation et la potentiation; et au niveau global pour la décharge du neurone postsynaptique. Bien sûr, le modèle des groupements est lui aussi la schématisation d'une réalité beaucoup plus complexe : chaque neurone peut posséder des centaines de groupements, mais ceux-ci ne sont pas séparés les uns des autres de façon aussi stricte. Les synapses sont plus ou moins loin les unes des autres et interagissent plus ou moins fortement entre elles pour dépolariser la membrane, ces interactions évoluant de façon non-linéaire dans le temps, et pouvant être aussi non-symétriques (une synapse est située en amont, une autre en aval).

Comme nous l'avons annoncé plus haut, nous n'approfondirons pas davantage cette version plus complexe du modèle de potentiation, et nous nous contenterons de l'approximation simple présentée au §3.2.b.

3.2.d Compétition intersynaptique

La loi de potentiation synaptique adoptée plus haut (cf. §3.2.b) incrémente les contacts $(i,j) \in E \times E$ de $+\alpha$ à chaque transmission réussie $i \rightarrow j$. Or, si tous ces contacts pouvaient croître indépendamment les uns des autres, ils finiraient inévitablement par se renforcer mutuellement, grâce à des décharges de plus en plus fréquentes des neurones postsynaptiques, et les connexions W_{ij}^{EE} atteindraient rapidement la valeur de saturation à 1. L'ensemble tomberait alors dans un état d'auto-renforcement permanent, où tous les neurones déchargeraient sans discontinuer. Donc, il est clair que les variations des contacts synaptiques interexcitateurs doivent être en réalité liées les unes aux autres, et toutes les observations neurobiologiques exposées au chapitre 2 attestent en effet l'existence de contraintes métaboliques qui impliquent une *compétition* entre axones afférents sur les cellules cibles.

Les divers exemples exposés au §2.2 montrent qu'une innervation au départ large et diffuse est l'objet d'une *focalisation progressive*, c'est à dire d'un rétrécissement du champ innervant de chaque neurone, accompagné d'une intensification des contacts dans ce champ, puis, au §2.3, on a mis en évidence les deux mécanismes fondateurs de ce phénomène : la coopération et la compétition. Ainsi, pour atteindre la précision remarquable que l'on peut observer dans certaines configurations nerveuses (propriétés fonctionnelles spécifiques), cette focalisation doit nécessairement s'appuyer sur une compétition entre candidats afférents au départ trop nombreux. Seuls les axones qui réussissent à synchroniser leurs signaux à l'arrivée sur le neurone postsynaptique récolteront un gain (voir plus haut), et cette récolte se fera alors *au détriment* de ceux qui n'étaient pas en phase au même instant.

Dans le modèle des groupements synaptiques exposé au §3.2.c, la compétition

entre afférences excitatrices est présente à l'intérieur de chaque groupement mais aussi entre groupements voisins : les contacts (i,j) qui ne sont pas en phase avec l'ensemble (G,j) auquel ils appartiennent régressent progressivement à cause de la potentiation des contacts voisins, et, de même, les groupements qui sont actifs moins fréquemment que des groupements voisins, perdent du terrain en faveur de ces derniers. Donc les groupements synaptiques stables forment des "îlots de coopération", et les îlots proches qui ne sont pas en synchronisation sont forcément en compétition. Finalement, à cause de la compétition, l'alternative qui s'impose aux synapses est soit de s'unir soit de disparaître.

Dans le modèle simplifié du §3.2.b qui ne tient pas compte de l'existence des groupements locaux, la lutte entre afférences excitatrices sera généralisée directement à toute la surface dendritique du neurone postsynaptique j, c'est à dire à l'ensemble des connexions $(W_{ij})_{i \in E}$ (celles qui sont rendues possibles par $C_{ij} = 1$). Pour formaliser cette compétition, le plus simple est alors de contraindre la somme de ces afférences à être *globalement conservée* ("sum rule"; von der Malsbug, 1973). Ainsi, on notera S_j' la somme de toutes les connexions excitatrices ayant j comme neurone postsynaptique (S_j' est considérée un opérateur s'appliquant à un état W), c'est à dire :

$$S_j'(W(t)) = \sum_{i \in E} W_{ij}(t)$$

et on posera :

$$\forall j \in E, \forall t \geq t_0, S_j'(W(t)) = S_j'(W(0))$$

$t=0$ désignant l'instant initial de l'évolution du réseau (la constante $S_j'(W(0))$ sera également notée $S_j'^0$ en abrégé). On dira de (i,j) qu'elle est une *connexion entrante* de j, et on appellera S_j' *somme entrante* de j (sous-entendu : somme des connexions entrantes excitatrices). La compétition entre afférences excitatrices est donc interprétée comme une conséquence de la quantité limitée de certains facteurs trophiques pouvant être fournis par le neurone postsynaptique j. De plus, comme on l'a fait remarquer au §2.3.a, cette limitation est également reliée à des contraintes d'ordre géométrique : sur un neurone possédant de nombreuses ramifications dendritiques, les contacts auront plus d'espace pour se développer. Toutes ces caractéristiques sont donc symbolisées par une quantité globale d'afférences excitatrices, $S_j'^0$, qui représente la somme de toutes les efficacités des contacts synaptiques excitateurs admissibles sur les dendrites de j : elle peut être répartie sur plusieurs axones afférents, mais doit être à chaque instant conservée.

Ainsi, l'augmentation de (i,j) provoquera nécessairement la diminution de tous les concurrents (i',j) qui n'augmentent pas au même instant. Bien sûr, cette règle formelle de conservation des sommes entrantes est une schématisation assez radicale des phénomènes réels, mais elle permet de traduire de façon simple le principe fondamental de la compétition, qui peut se résumer ainsi : il n'y a pas de place pour le développement de tous les contacts synaptiques, car les ressources sont

limitées, et seuls ceux qui réussiront à coopérer pourront se partager ces ressources et survivre.

Par ailleurs, pour des raisons similaires, nous introduirons également dans ce modèle une contrainte sur l'ensemble des *efférences* d'un neurone excitateur : de même qu'un neurone ne peut pas entretenir une trop grande quantité de contacts afférents sur ses dendrites, il ne pourra pas non plus entretenir une trop grande quantité de terminaisons axonales, c'est à dire une innervation large sur de nombreuses cellules cibles, même si des transmissions réussies vers ces cellules lui permettraient de s'y maintenir. Donc, les restrictions ne se trouvent pas seulement au niveau postsynaptique : les neurones présynaptiques devront eux aussi faire un choix parmi les cibles qu'ils ont la possibilité d'innover.

Comme précédemment, les raisons en sont vraisemblablement géométriques et métaboliques : en général, la divergence des ramifications axonales reste localisée dans certaines régions de l'espace, et de plus, le matériel métabolique nécessaire à l'entretien de l'arborescence axonale d'un neurone ne peut pas être fourni en quantité illimitée. Cependant, la compétition entre efférences a reçu moins d'évidences neurobiologiques que la compétition entre afférences, les origines de l'une et de l'autre n'étant sans doute pas de même nature. Quoi qu'il en soit, le principe de la compétition dans le sens divergent sera formalisé exactement de la même manière que dans le sens convergent. Ainsi, on notera S_i la somme des efférences du neurone i :

$$S_i(\mathbf{W}(t)) = \sum_{j \in E} W_{ij}(t)$$

et on posera à nouveau :

$$\forall j \in E, \forall t \geq t_0, S_i(\mathbf{W}(t)) = S_i(\mathbf{W}(0))$$

(la constante étant notée encore S_i^0). On appellera aussi (i,j) *connexion sortante* de i , et S_i *somme sortante* de i . Remarquons que cette somme sortante est restreinte aux cibles de type excitateur : il n'est en effet pas utile d'y inclure les neurones inhibiteurs $j \in I$, puisque leurs contacts afférents sont supposés invariables. Dans la somme entrante S_j' , en revanche, la restriction aux excitateurs était due au fait que les synapses inhibitrices ne pouvaient pas se trouver en compétition avec des synapses excitatrices à cause de leur nature très différente (donc, finalement, si les afférences et efférences des neurones inhibiteurs pouvaient varier, la somme S_i serait étendue à tout $j \in R$, tandis que la somme S_j' resterait limitée à $i \in E$).

Ainsi, l'ensemble des contacts synaptiques sera contraint non seulement du point de vue postsynaptique mais aussi du point de vue présynaptique, et ceci est à nouveau schématisé par la conservation d'une efficacité globale, qui s'applique cette fois-ci aux sommes sortantes. Encore une fois, il s'agit d'une vue très schématique de la réalité : par exemple, cette règle est neutre du point de vue de la répartition spatiale des efférences, et, à la limite, elle autorise l'innervation de tous les neurones

du réseau par des efférences virtuellement très faibles se partageant la valeur totale S_i^0 (à l'intérieur de $C_{ij}=1$).

Le respect de la contrainte stricte de conservation des sommes par la matrice W s'exprime donc sous la forme d'un système de $2N_E$ équations à $\approx p_{EE}N_E(N_E-1)$ variables, qui sont les W_{ij}^{EE} tels que $C_{ij}=1$, les autres connexions $E \times E$ devant toujours rester nulles. Dans toute la suite, le nombre de ces variables, c'est à dire les connexions $E \times E$ non obligatoirement nulles, sera noté c_E . La condition sine qua non de cohérence du système est évidemment que la somme des sommes entrantes S_j^{-0} soit égale à la somme des sommes sortantes S_i^0 . En effet, toute matrice W vérifie :

$$\sum_{i \in E} S_i(W) = \sum_{j \in E} S_j'(W) = \sum_{(i,j) \in E \times E} W_{ij}$$

et, donc, il doit en être de même pour les constantes de conservation S_j^{-0} et S_i^0 . Or, cette condition est bien remplie ici, puisque ces constantes sont elles-mêmes les sommes d'une matrice $W(0)$: ainsi, l'existence d'au moins une solution (ici, $W(0)$), garantit la non-contradiction du système. Si le nombre de variables est inférieur au nombre d'équations, c'est à dire si $c_E \leq 2N_E$, le système est surdéterminé, et il se peut qu'il n'admette aucune solution en dehors de $W(0)$. Cependant, cette situation suppose que p_{EE} soit inférieur à $2/N_E$, c'est à dire que le graphe des connexions (i,j) telles que $C_{ij}=1$ soit très dilué par rapport au graphe complet $E \times E$: en effet, la connectivité moyenne serait alors plus petite que 2 connexions entrantes et 2 connexions sortantes par neurone, ce qui est bien trop peu pour notre modèle. On se trouve donc dans le cas où $c_E > 2N_E$, et même $c_E \gg 2N_E$ (ce qui est facilement réalisé, ne serait-ce qu'avec $N_E=100$ et $p_{EE}=0,2$). Le système est alors largement sous-déterminé, et il existe une infinité de solutions dans l'espace vectoriel des W , que l'on appellera *matrices parfaites*. Leur ensemble dépend de toutes les constantes (S_i^0) et (S_j^{-0}) , et son intersection avec l'hypercube K (cf. §3.1.c) forme un sous-ensemble M . En d'autres termes, M contient les matrices parfaites dont certains coefficients sont obligatoirement nuls (ceux pour lesquels $C_{ij}=0$), et les autres sont bornés entre 0 et 1 (ceux pour lesquels $C_{ij}=1$). Remarquons que M est *convexe*, donc a fortiori *connexe* : en effet, tout barycentre de matrices parfaites bornées est également une matrice parfaite bornée.

Dans ce cadre, la dynamique des connexions peut être décrite comme suit : partant de $W(0)$, le point W se déplace sous l'impulsion du terme A , mais doit également rester à l'intérieur des frontières de M . Ce n'est donc plus seulement l'hypercube K qui constitue le domaine de W , mais sa restriction au sous-ensemble des matrices parfaites M . Par conséquent, le problème est le suivant : il s'agit de concilier les petits déplacements ΔW occasionnés par les incréments et décréments contenus dans le terme A , avec le respect de la contrainte de conservation. C'est ici qu'intervient le terme B : étant donné un état $W(t-1) \in M$, et étant donné le déplacement $A(t)$, on obtient un nouvel état de connexions, $W(t-1)+A(t)$, qui se trouve en général à l'extérieur de M : le terme correctif $B(t)$ est alors chargé de le ramener vers un état "parfait", c'est à dire de réaliser : $W(t-1)+A(t)+B(t) \in M$. Comment est alors calculé $B(t)$? Il serait possible d'élaborer un algorithme itératif qui réussirait à faire converger les connexions vers M , par itérations successives à partir de $W(t-1)+A(t)$. Cependant, les calculs que cet algorithme nécessiterait seraient certainement plus coûteux que l'enjeu ne le mérite réellement : on se contentera donc d'une *approximation* pour B , comme on l'explique dans les lignes suivantes.

Dans ce contexte, pour faire en sorte qu'un état de connexions W satisfasse *au mieux* les équations de conservation précédentes, sans pour autant exiger leur strict respect, nous allons transformer cette contrainte stricte en contrainte faible, en définissant une *fonction-coût* sur les connexions $E \times E$. Cette fonction, notée H , aura pour tâche de mesurer le degré de violation de la règle de conservation : en d'autres termes, étant donné M , l'ensemble des états qui obéissent parfaitement aux équations ("états parfaits"), on définit une distance de W à M , à travers la quantité suivante :

$$H(W) = \gamma \sum_{i \in E} (S_i(W) - S_i^0)^2 + \gamma' \sum_{j \in E} (S_j'(W) - S_j'^0)^2$$

où γ et γ' sont des paramètres réels positifs. Il s'agit donc d'une "énergie quadratique", qui vaut par définition 0 sur tout l'ensemble M (dont $W(0)$ fait partie), et qui augmente d'autant plus que W s'éloigne de M , c'est à dire que les sommes sortantes et entrantes de W s'écartent des valeurs initiales S_i^0 et $S_j'^0$. Donc, nous n'exigerons pas que les valeurs réelles $S_i(W)$ et $S_j'(W)$ soient exactement égales aux valeurs désirées S_i^0 et $S_j'^0$, mais nous pénaliserons les différences entre les premières et les deuxièmes par une augmentation de H .

D'autre part, puisque la rigueur est maintenant délaissée au profit d'une optimisation plus souple, la situation sera davantage simplifiée en adoptant des valeurs uniformes de sommes désirées pour tous les neurones, d'une part dans le sens entrant, et d'autre part dans le sens sortant. Autrement dit, on remplacera désormais les $2N$ paramètres (S_i^0) et $(S_j'^0)$ par 2 constantes seulement, notées respectivement s_0 et s_0' :

$$H(W) = \gamma \sum_{i \in E} (S_i(W) - s_0)^2 + \gamma' \sum_{j \in E} (S_j'(W) - s_0')^2$$

Il y a deux conséquences à cette simplification. D'une part, l'état initial $W(0)$ n'est plus solution *a priori* des équations de conservation; d'autre part, si les valeurs s_0 et s_0' sont différentes, alors le système en question n'est plus cohérent, et par conséquent n'admet aucune solution, c'est à dire : $M = \emptyset$. En effet, comme on l'a signalé plus haut, la condition de non-contradiction du système est que la somme des sommes entrantes soit égale à la somme des sommes sortantes, ce qui entraîne ici : $Ns_0' = Ns_0$. Cependant, dans la mesure où on a remplacé ce système strict par la minimisation de la fonction-coût H , le cas conflictuel $s_0' \neq s_0$ se résume simplement au fait que H ne pourra pas atteindre la valeur 0, c'est à dire que $H_{\min} > 0$. Dans ce cas, on peut redéfinir M comme étant l'ensemble des états qui réalisent H_{\min} : ce sont des états de compromis entre les demandes incompatibles : $S_i(W) = s_0$ et $S_j'(W) = s_0'$. Leurs sommes seront toutes égales et prendront une valeur intermédiaire entre s_0 et s_0' pondérée par les coefficients γ et γ' . Comme on le verra tout de suite, les états en dehors de M , c'est à dire correspondant à une pénalité plus élevée que H_{\min} , seront systématiquement réorientés vers M grâce à la minimisation de H (bien sûr, on peut faciliter les choses dès le départ, en choisissant $W(0) \in M$).

Le but est donc de *minimiser* H à chaque instant. Pour simplifier, on posera dans toute la site : $s_0 = s_0'$, ce qui place le minimum absolu H_{\min} à la valeur 0. Cependant, pour gagner en temps de calcul, on ne cherchera pas à atteindre précisément $H = H_{\min} = 0$, et, en pratique, on se contentera d'un seul pas dans cette direction. Autrement dit, partant de l'état provisoire $\mathbf{W}(t-1) + \mathbf{A}(t)$ dont l'énergie H est supérieure à 0, on ajoutera un terme correctif $\mathbf{B}(t)$, proportionnel au gradient de H calculé au point $\mathbf{W} + \mathbf{A}$. Ceci est suffisant pour garantir le retour vers un état satisfaisant du point de vue de la conservation des sommes, c'est à dire proche de 0, car \mathbf{B} vient en réaction contre les déplacements provoqués par \mathbf{A} , lesquels ne sont pas assez grands pour perturber sensiblement l'équilibre conservatif. Finalement, pour tout $(i,j) \in E \times E$ tel que $C_{ij} = 1$, on a :

$$\begin{aligned} B_{ij}(t) &= - \left(\frac{\partial H}{\partial W_{ij}} \right)_{\mathbf{W}(t-1) + \mathbf{A}(t)} \\ &= - 2\gamma (S_i(\mathbf{W}(t-1) + \mathbf{A}(t)) - s_0) - 2\gamma' (S_j'(\mathbf{W}(t-1) + \mathbf{A}(t)) - s_0) \end{aligned}$$

Au chapitre 4, nous aurons l'occasion de mieux apprécier la double influence des termes \mathbf{A} et \mathbf{B} sur la variation des connexions, à travers les valeurs ponctuelles que celles-ci prendront d'une itération sur l'autre.

3.2.e Bilan

En résumé, on peut considérer formellement que l'état des connexions entre neurones excitateurs \mathbf{W}^{EE} est soumis à deux "forces" simultanées : d'une part, une force de "rétro-action" en provenance des activités neuronales, faisant immédiatement suite à la modification de celles-ci et dépendant de leurs corrélations temporelles, et, d'autre part, une force de "réaction" accompagnant cette influence et tendant à rééquilibrer la somme des afférences et des efférences qui ont été provisoirement perturbées, cette réaction étant dérivée d'une compétition intersynaptique due à une loi de conservation des sommes. La première force (corrélations) donne ainsi la véritable "impulsion" aux connexions, et déplace leur point caractéristique \mathbf{W}^{EE} dans l'hypercube K , tandis que la deuxième force (compétition) corrige ce déplacement, en réagissant à cette impulsion initiale.

Finalement, l'état du réseau $(\mathbf{x}(t), \mathbf{W}(t))$ est composé d'une partie dynamique et d'une partie statique : $(\mathbf{x}^E(t), \mathbf{x}^I(t), \mathbf{W}^{EE}(t))$ et $(\mathbf{W}^{EI}, \mathbf{W}^{IE})$, la partie dynamique obéissant aux trois équations d'évolution exposées au §3.2.a, §3.2.b, et §3.2.d. Ainsi, le système total fonctionne selon un mouvement ternaire : (a) "action" (nouvelle activation des neurones), (b) "rétro-action" (effet de leurs corrélations sur les connexions), (c) "réaction" (compétition entre connexions).

Dans ce modèle, les deux influences agissant sur les connexions ont été symbolisées par deux termes, **A** et **B**, dont la somme représente la variation totale de W^{EE} . Pour plus de commodité, cette variation sera étendue formellement à toute la matrice **W**, et on la notera ΔW . L'évolution des connexions s'écrit donc :

$$\Delta W_{ij}(t) = W_{ij}(t) - W_{ij}(t-1)$$

avec :

$$\forall (i,j) \in R \times R, \Delta W_{ij}(t) = A_{ij}(t) + B_{ij}(t)$$

où A_{ij} et B_{ij} seront formellement nuls pour $(i,j) \in E \times I$ et $(i,j) \in I \times E$, et pour tout (i,j) tel que $C_{ij} = 0$. Cependant, les connexions W^{EE} doivent également respecter les bornes minimales et maximales, 0 et 1 : il faut donc s'assurer que la variation globale ΔW ne déplace pas **W** en dehors de l'hypercube **K**. Pour cela, on procédera simplement par *troncation* : ainsi, dans le cas où la nouvelle valeur d'une connexion $(i,j) \in E \times E$ sortirait de l'intervalle $[0, 1]$, elle serait ramenée à la borne la plus proche, 0 ou 1. Par conséquent, W_{ij} doit être rectifiée en W_{ij}^* de la façon suivante :

$$\forall (i,j) \in E \times E, W_{ij}^*(t) = \sup\{0, \inf\{W_{ij}(t), 1\}\}$$

Cette dynamique sous contrainte pourra encore s'écrire : $W(t) = W(t-1) + A(t) + B(t)$, avec $0 \leq W^{EE}(t) \leq C^{EE}$.

4 ANALYSE DU MODÈLE : PRINCIPES DE FONCTIONNEMENT

Dans ce chapitre, nous analyserons le comportement du modèle sous des conditions paramétriques simplifiées. Il s'agit essentiellement d'une description "intuitive" des phénomènes d'organisation, utilisant des outils statistiques élémentaires, et qui s'accorde de façon claire avec les simulations. Cette étude ne prétend donc pas fournir un cadre théorique complet au fonctionnement de ce modèle, mais elle est destinée à mettre en évidence les principaux aspects de l'évolution du réseau R , c'est à dire les phénomènes de comportement collectif qui peuvent émerger des lois élémentaires exposées au §3.2, et qui aboutiront à la construction d'une "configuration spatio-temporelle" (CST). Des cas de figures plus complexes seront ensuite abordés à travers les simulations du chapitre 5, puis, au chapitre 6, nous traiterons de la croissance simultanée des CST ainsi que de leurs propriétés compositionnelles.

Nous verrons comment, à partir d'un état initial informe caractérisé par des contacts synaptiques aléatoires et des activités neuronales désordonnées, une structure cohérente de connexions et de corrélations peut graduellement apparaître : d'une part, comment une innervation au départ large et diffuse peut se focaliser par le renforcement spécifique de certaines connexions et l'affaiblissement des connexions concurrentes, à travers le jeu des coïncidences temporelles et de la compétition; d'autre part, comment, parallèlement, un état d'activités stochastique, ne présentant aucune synchronisation entre neurones, peut évoluer vers un état de *fortes corrélations*, grâce à la focalisation de l'innervation, et faire ainsi la transition d'une propagation aléatoire à une propagation régulière. Enfin, et surtout, nous verrons comment ces réarrangements synaptiques et ces configurations spatio-temporelles peuvent se *soutenir mutuellement* pour provoquer une véritable *auto-organisation* du réseau.

Ainsi, à travers des exemples simples, nous tenterons d'analyser le mécanisme de base de cette structuration. La loi de décharge des neurones et la loi de modulation synaptique avec compétition fournissent le principe d'un renforcement réciproque sous contrainte : la croissance des contacts favorise la coïncidence des signaux pré- et postsynaptiques, et cette coïncidence encourage à son tour le développement des synapses, l'ensemble étant soumis à une limitation sur la quantité des contacts. Le réseau devra donc opérer un choix parmi toutes les combinaisons viables de connexions et corrélations. Par le jeu du hasard, certaines deviendront prédominantes devant d'autres et imposeront leur orientation au réseau. Dans ce système évolutif, la composante stochastique introduite par la température jouera le rôle de *diversification*, en proposant des ébauches de ces

combinaisons, tandis que les lois coopératives/compétitives exerceront une *sélection* parmi ces possibilités.

Les états stables de la double dynamique synapto-neuronale seront des structures de connexions fortes reliant des neurones dont les processus de décharge seront synchronisés de façon optimale : les déphasages constants entre neurones correspondront précisément aux délais de transmission des connexions. Ces états stables seront donc caractérisés par une cohérence interne dans la cartographie des divers chemins synaptiques qui conduisent les potentiels d'action : pour aller d'un neurone i à un neurone j , les signaux pourront emprunter de multiples voies divergentes/convergentes qui auront toutes le *même délai total*. Le graphe des délais est donc "transitif", c'est à dire qu'il vérifie : $\tau_{ik} + \tau_{kj} = \tau_{ik'} + \tau_{k'j}$ (se reporter à la figure 23, chapitre 6). Or, justement, cette cohérence garantit en retour la coïncidence des signaux afférents qui permet aux neurones de décharger et par conséquent de perpétuer la stabilité de ce câblage. Un câblage non-transitif ne serait pas viable en raison du trop grand nombre d'échecs de coïncidences : autrement dit, les connexions doivent coopérer pour assurer leur survie.

En résumé, les états stables de connexions fortes supporteront des *ondes de propagation régulière* de l'activité neuronale, qui, précisément, entretiendront en retour ces connexions. Ces propagations d'activité seront donc caractérisées par des "motifs rythmiques" reproductibles, que nous avons appelés plus haut *configurations spatio-temporelles* (CST). Elles pourront être de "dimension" différente, suivant la connectivité du graphe qui les porte et l'étalement de la propagation sur les neurones : ainsi, certaines CST seront qualifiées de "linéaires" et d'autres de "planes" (par exemple, des modèles de la projection rétinotopique utilisent deux CST planes, dont les coïncidences temporelles permettent la croissance de liens parallèles; voir aussi la partie II de cette thèse, concernant les appariements de graphes).

Dans le présent modèle, nous nous intéresserons aux CST de type linéaire, que nous appellerons également "*tresses*", parmi lesquelles on distinguera une famille particulière de structures, appelées "*chaînes*". Les chaînes sont des versions schématiques des tresses, la différence pouvant se résumer par la nature des délais, qui sont entiers dans les premières et réels dans les deuxièmes. Ainsi, les chaînes sont plus aisément accessibles par les simulations : pour une même connectivité, la transitivité des délais est plus facilement atteinte dans un cas que dans l'autre. Une chaîne est constituée d'une succession de *maillons* distincts, chacun étant composé de neurones non-connectés entre eux et toujours actifs simultanément. La propagation consiste donc en une activation ordonnée de ces maillons, qui sont reliés les uns aux autres par des connexions dont les délais sont cohérents (c'est à dire proportionnels à la "distance" entre groupes le long de la chaîne). Les "*synfire-chains*" d'Abeles (1982) sont les exemples les plus simples de chaînes : leurs délais sont uniformes, ce qui fait que chaque maillon est uniquement relié au maillon suivant.

Dans ce chapitre 4, nous montrerons dans quelles conditions le modèle peut rendre compte de l'auto-organisation des synfire-chains : pour cela, les formules du §3.0 seront développées dans un contexte simplifié. Tout d'abord, nous prendrons des *délais identiques* pour toutes les connexions, et, pour plus de commodité, nous les supposerons tous égaux à 1 : ainsi, $x(t)$ ne dépendra que de $x(t-1)$. D'autre part, en plus de cette simplification sur les délais, nous ne tiendrons pas compte de l'action des neurones inhibiteurs. Ceci pourra se traduire par des connexions W^{EI} et W^{IE} très faibles (proches de 0), ou par la coupure des contacts entre excitateurs et inhibiteurs : $p_{EI} = p_{IE} = 0$. Il sera cependant plus clair de supprimer directement les inhibiteurs du réseau en posant : $I = \emptyset$ et $R = E$, c'est à dire $N_I = 0$ et $N = N_E$. Par ailleurs, nous supposerons que les contraintes génétiques sont suffisamment larges pour autoriser des connexions entre tous les neurones excitateurs restants (sauf d'un neurone vers lui-même), ce qui se traduit par : $p_{EE} = 1$ (la mention "EE", devenue superflue, sera omise dans toute la suite). La connectivité sera donc complète : $\forall (i,j) \in R \times R; i \neq j, C_{ij} = 1$. Enfin, sauf indication contraire, les propriétés de réfraction des membranes axonales seront en général négligées : $\forall i \in R, \rho_{ij} = 0$.

Il s'agira donc d'une étude formelle, destinée avant tout à comprendre les principes organisateurs issus du modèle, et à montrer comment les autres paramètres du réseau doivent être choisis pour assurer les meilleures conditions de croissance des chaînes. Le comportement du modèle complet, comportant délais non-uniformes, neurones inhibiteurs ou une connectivité partielle, sera exploré au chapitre 5 à travers les simulations numériques.

On commencera par étudier au §4.1 le comportement stochastique des activités dans un ensemble de connexions homogènes, afin d'obtenir un état initial "calme", donc favorable au développement des chaînes. Puis, le §4.2 sera consacré à l'étude d'une étape élémentaire dans la croissance des connexions : on y calculera en détail la matrice de variation $\Delta W(t+1)$ due à une transition de l'état des activités $x(t) \rightarrow x(t+1)$ a priori quelconque. Dans un paragraphe intermédiaire §4.3, on s'intéressera aux conséquences d'un régime d'activations forcé sur l'organisation du réseau, en envisageant deux cas extrêmes opposés : d'une part des processus de décharge totalement aléatoires, qui provoqueront une désorganisation généralisée des connexions et leur maintien dans un état statistiquement homogène; d'autre part des transitions d'activités très régulières, repassant systématiquement par les mêmes groupes de neurones, et qui aboutiront au contraire à la formation d'une chaîne. Ces deux états sont en réalité des états stables de la double dynamique, et le régime forcé d'activations ne leur est pas nécessaire : en effet, un réseau inorganisé perpétue les activations aléatoires qui le rendent homogène, et, de même, une chaîne peut soutenir elle-même les "ondes d'activation" qui contribuent au renforcement de ses connexions. Le premier cas (désorganisation aléatoire) ne sera pas davantage approfondi, par contre on s'interrogera sur les conditions de stabilité

des chaînes au paragraphe §4.4. Enfin, on verra au §4.5 comment ces chaînes peuvent s'auto-organiser grâce au soutien mutuel de la focalisation et des corrélations entre décharges. Leur croissance se cristallisera à partir d'une inhomogénéité introduite dans le réseau, sous forme d'un neurone possédant des efférences particulièrement fortes, et qui aura pour conséquence la formation d'un groupe de "neurones pulsatiles" synchrones : ces neurones donneront l'impulsion nécessaire à la propagation des ondes d'activation qui entretiendront la chaîne et l'inciteront à se développer.

Tout d'abord, nous récapitulons ici les paramètres qui caractérisent le réseau R , en précisant aussi les ordres de grandeur qui seront envisagés dans les simulations. On rappelle que R sera uniquement composé de neurones excitateurs, en nombre N (entre 100 et 1000). A chaque neurone correspond un seuil θ_i , valeur réelle positive tirée aléatoirement dans un intervalle $[\Theta_1, \Theta_2]$, de moyenne Θ_0 (par exemple $\Theta_0=3$). Comme on l'a précisé plus haut, les composantes C_{ij} sont toutes égales à 1 (sauf pour $i=j$), ce qui autorise $N(N-1)$ connexions entre les N neurones du réseau (cependant, au moment des simulations, si la taille du réseau est relativement grande ($N > 200$), il sera utile d'appliquer le coefficient de dilution uniforme "p" pour ramener le nombre total de connexions à $pN(N-1)$ et alléger les calculs; en fonction de N , p peut aller de 100% à 10%). La température T qui bruite les processus de décharge sera en général inférieure à 1 (dans l'argument de la sigmoïde, elle divise les différences entre potentiels et seuils ($V_j - \theta_j$) qui sont de l'ordre de 1). On rappelle que l'efficacité maximale des connexions est prise égale à 1 : $\forall (i,j) \in R \times R; i \neq j, 0 \leq W_{ij} \leq 1$. Dans ce contexte, l'incrément α et le décrétement β de la règle de potentiation des synapses seront respectivement de l'ordre de 0,1 et 0,01; les coefficients de pénalisation γ et γ' de la règle de compétition de l'ordre de 0,001, et la valeur de référence s_0 des sommes entrantes et sortantes de l'ordre de 10. L'état initial du réseau est : $(x(0), W(0))$: en général, $x(0)$ sera uniformément nul (neurones silencieux), et les connexions seront tirées aléatoirement dans un intervalle $[w_1, w_2]$, à l'intérieur de $[0, 1]$, de moyenne w_0 et de largeur Δw petite devant 1. La valeur w_0 sera choisie de telle sorte que l'état initial $W(0)$ soit satisfaisant du point de vue de la conservation des sommes : donc, sachant que les valeurs moyennes des sommes entrantes et sortantes sont les mêmes et valent Nw_0 , il faudra faire en sorte que : $s_0 = (N-1)w_0 \approx Nw_0$ (par exemple : $N=100, s_0=10$ et $w_0=0,1$).

4.1 Activités seules dans un réseau homogène

Le réseau R est au départ totalement désordonné, caractérisé par l'état de connexions aléatoire et homogène défini ci-dessus. Chaque neurone a créé des ébauches de contacts avec tous les autres neurones, mais les efficacités de transmission sont trop faibles pour assurer une activité moyenne productrice de

coïncidences et d'organisation. L'innervation est encore complète et très diffuse, et, parallèlement, les propagations de potentiels sont rares et décorréélées. Dans ce paragraphe, nous étudierons spécifiquement cet état d'inorganisation primitive, c'est à dire le comportement isolé des activités neuronales communiquant à travers le réseau de connexions homogène. Pour cela, on ne tiendra pas compte de la modulation synaptique : on négligera provisoirement l'influence des corrélations et de la compétition en posant : $\alpha = \beta = 0$ et $\gamma = \gamma' = 0$. Notre but est donc d'estimer le *taux moyen d'activité spontanée* du réseau afin de réaliser les conditions paramétriques dans lesquelles cette activité soit faible, mais non nulle. Ceci doit ensuite permettre aux connexions de s'organiser dans un environnement favorable à leur développement, c'est à dire au milieu d'un ensemble de processus de décharge qui ne soient ni trop forts ni totalement silencieux.

On rappelle que le réseau est caractérisé par les paramètres N, T, Θ_0, w_0 , et par le tirage de deux familles de composantes aléatoires : les seuils (θ_j) et les connexions (W_{ij}^0), de valeurs moyennes respectives $\langle \theta \rangle = \Theta_0$, et $\langle W^0 \rangle = w_0$. On notera $n(t)$ l'activité instantanée totale du réseau, c'est à dire : $n(t) = \sum x_i(t)$. Donc, dans un état stationnaire de connexions, $W_{ij}(t) \approx W_{ij}^0$, on cherchera la loi d'évolution de l'activité moyenne instantanée $\langle n(t) \rangle$, et, en particulier, le niveau moyen de cette quantité dans le temps, n^* .

4.1.a Première itération

On suppose qu'au départ aucun neurone n'est actif, c'est à dire $n(0) = 0$. Donc à l'instant suivant, les potentiels postsynaptiques sont tous nuls : $V_j(1) = 0$, et les éventuelles activations ne peuvent être dues qu'à la part de hasard introduite par la température T . Ainsi, à l'instant 1, le nombre moyen de neurones actifs vaut :

$$\langle n(1) \rangle = \sum_j \text{proba}[x_j(1)=1] = \sum_j \frac{1}{1 + e^{\theta_j/T}}$$

Par ailleurs, comme les seuils (θ_j) sont N variables aléatoires indépendantes issues d'une même distribution, l'espérance de $\langle n(1) \rangle$ sur différents tirages de ces paramètres s'écrit :

$$\langle n(1) \rangle = \sum_j \left\langle \frac{1}{1 + e^{\theta_j/T}} \right\rangle \approx \frac{N}{1 + e^{\langle \theta \rangle/T}}$$

avec $\langle \theta \rangle = \Theta_0$. Cette approximation se fonde sur l'hypothèse que l'écart-type de la distribution seuils $\Delta\Theta$ est suffisamment petit pour pouvoir considérer que la sigmoïde est pratiquement linéaire dans leur domaine. Donc cette égalité est exacte dans le cas où tous les seuils sont identiques ($\theta_j = \Theta_0$), et elle reste à peu près valable dans le cas où ils sont distribués autour de Θ_0 avec un étalement petit devant le domaine de transition de la sigmoïde (si on définit ce domaine comme l'intervalle

entre les abscisses qui réalisent $\sigma_T = 10\%$ et $\sigma_T = 90\%$, sa largeur vaudra environ $4T$: donc, par exemple, pour $T=0,5$ on devra avoir $\Delta\Theta \ll 2$, sachant d'autre part que Θ_0 est de l'ordre de quelques unités).

4.1.b Itérations suivantes

Ensuite, à l'instant $t=2$, les quelques neurones actifs en nombre $n(1)$ vont propager dans le réseau des potentiels faibles mais non nuls, $V_j(2)$, qui provoqueront une légère augmentation des probabilités d'activation, donc du nombre moyen de neurones actifs :

$$\langle n(2) \rangle = \sum_j \text{proba}[x_j(2)=1] = \sum_j \frac{1}{1 + e^{-(V_j(2) - \theta_j)/T}} > \langle n(1) \rangle$$

et, comme précédemment, avec un $\Delta\Theta$ suffisamment petit, θ_j pourra être remplacé par Θ_0 . Cependant, à la différence de $V_j(1)$, les potentiels $V_j(2)$ ne sont pas nuls, et il faut donc calculer l'espérance de cette quantité selon V_j , c'est à dire :

$$\langle n(2) \rangle \approx \sum_j \left(\sum_V \text{proba}[V_j(2)=V] \frac{1}{1 + e^{-(V - \Theta_0)/T}} \right)$$

Or le potentiel parvenant au neurone j à l'instant $t=2$ est déterminé par les connexions (W_{ij}^0) et par la configuration d'activités précédente $x_i(1)$:

$$V_j(2) = \sum_i W_{ij}^0 x_i(1)$$

Donc, faire la moyenne de $\langle n(2) \rangle$ sur ces potentiels nécessite la connaissance de la distribution conjointe des états d'activité et des connexions : $\text{proba}[x(1), \mathbf{W}^0]$. C'est pourquoi, afin d'éviter un calcul trop compliqué, nous appliquerons à ces potentiels la même approximation qu'aux seuils, et nous ne tiendrons compte que du *potentiel moyen* $\langle V_j(2) \rangle$ en chaque neurone j (nous verrons que cette simplification est acceptable en la comparant aux simulations). Ce potentiel moyen vaut :

$$\langle V_j(2) \rangle \approx \sum_i \langle W_{ij}^0 \rangle \langle x_i(1) \rangle \approx w_0 \langle n(1) \rangle$$

en supposant que les connexions restent proches de leur valeur moyenne $\langle W_{ij}^0 \rangle$ (et en négligeant le fait que $W_{ii}^0=0$), ce qui implique que les probabilités d'activations sont uniformes sur le réseau (c'est particulièrement vrai à l'instant $t=1$ puisqu'il n'y avait aucune activité à $t=0$). Ainsi, les neurones j ont à peu près tous le même potentiel postsynaptique moyen $\langle V_j(2) \rangle$, qui est simplement égal au produit de

l'activité moyenne précédente $\langle n(1) \rangle$ et de la valeur typique d'une connexion w_0 . Donc finalement, le nombre moyen de neurones actifs à l'instant $t=2$ s'écrit :

$$\langle n(2) \rangle \approx \frac{N}{1 + e^{-(w_0 \langle n(1) \rangle - \Theta_0)/T}}$$

En réitérant ces calculs à toutes les étapes ultérieures $[t, t+1]$, on obtient alors une formule récurrente approximative, qui à chaque instant relie le nombre moyen de neurones actifs à celui de l'instant précédent, à travers la fonction-sigmoïde σ_T :

$$\forall t \geq 0, \langle n(t+1) \rangle \approx \frac{N}{1 + e^{-(w_0 \langle n(t) \rangle - \Theta_0)/T}}$$

Les hypothèses que nous avons faites pour y parvenir concernent la distribution des paramètres du réseau, seuils et connexions initiales : ces distributions doivent être relativement serrées autour de leurs valeurs moyennes respectives Θ_0 et w_0 . Nous avons aussi supposé que l'étalement des valeurs de potentiels était restreint.

4.1.c Stabilité de l'activité moyenne

On cherche à présent les conditions de stabilité du réseau, c'est à dire les conditions dans lesquelles un même nombre moyen de neurones (petit devant le nombre total N) déchargent à chaque instant. Ainsi, partant d'un état totalement vide, le réseau voit son activité moyenne augmenter graduellement, et atteindre en quelques itérations une certaine valeur limite, notée n^* , qui dépendra des paramètres N , T , w_0 et Θ_0 . On s'intéressera aux cas où cette valeur est faible mais non-nulle, en cherchant avant tout à éviter la situation "catastrophique" dans laquelle tous les neurones déchargent en même temps, c'est à dire $n^* = N$. En effet, il est clair qu'en l'absence de processus inhibiteur, cette situation constitue un état stable de la dynamique : les neurones s'excitent mutuellement à travers des potentiels très élevés, et le bruit introduit par la température n'est pas suffisant pour faire redescendre le réseau vers une activité globale plus raisonnable. Toute possibilité ultérieure d'organisation est alors bloquée.

Donc nous voulons : $\forall t, \langle n(t) \rangle \approx n^* \ll N$, ce qui implique a fortiori : $\langle n(t+1) \rangle \approx \langle n(t) \rangle \approx n^*$. Par conséquent, n^* est un point fixe de la relation de récurrence trouvée plus haut :

$$n^* \approx \frac{N}{1 + e^{-(w_0 n^* - \Theta_0)/T}}$$

Il s'agit de l'équation du régime stationnaire du réseau, lorsque les neurones déchargent dans un ensemble homogène de connexions faibles, sous la température T . Cette équation s'écrit en résumé : $n^* \approx \sigma(w_0 n^* - \Theta_0)$, où la sigmoïde σ , en tant que fonction de n^* , vaut : $N\sigma_T(w_0 n^* - \Theta_0)$.

Les conditions d'existence d'un tel équilibre bruité sont clairement montrées par une étude graphique (figure 1) : sur un système d'axes comportant en abscisses n^* et en ordonnées l'intervalle $[0, N]$, on trace d'une part la sigmoïde $n = \sigma(n^*)$ et d'autre part la bissectrice principale $n = n^*$, puis on examine les points d'intersection de ces deux courbes. On voit alors qu'il y a toujours au moins une solution, mais que celle-ci est stable à condition que la pente de σ en ce point soit inférieure à 1. Le point de symétrie Ω de la sigmoïde (correspondant à un argument nul dans l'exponentielle) a pour abscisse Θ_0/w_0 et pour ordonnée $N/2$ (lorsque T varie, la sigmoïde "pivot" autour de Ω).

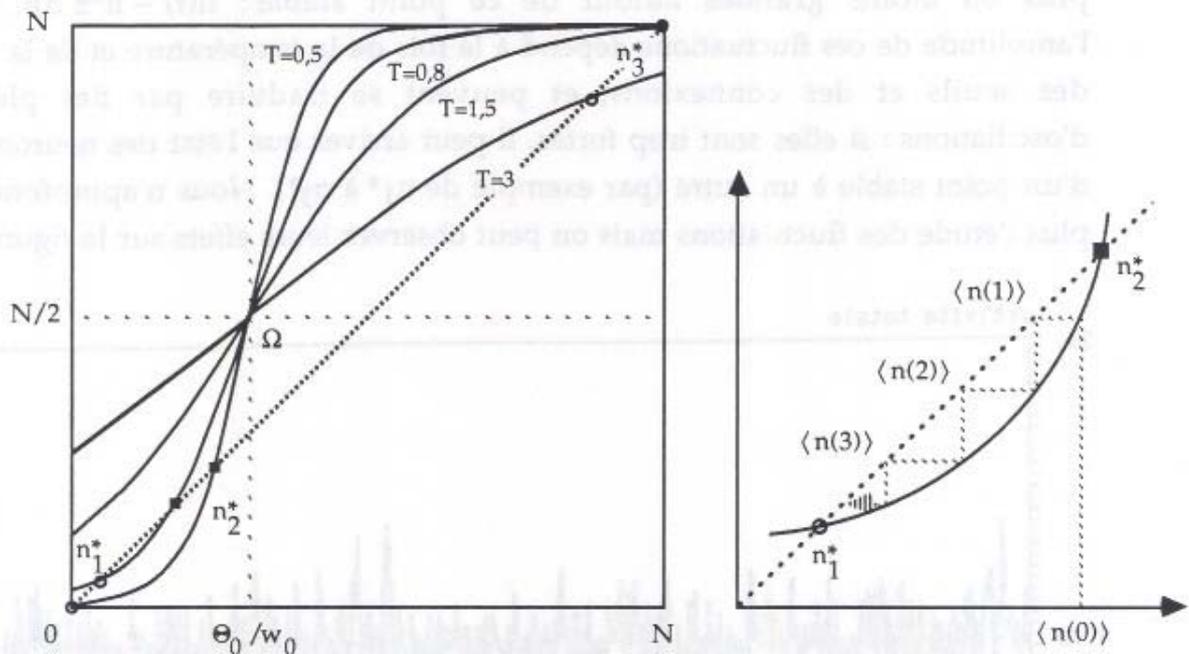


Figure 1 : Etude graphique de l'activité moyenne du réseau. On a représenté ici quatre sigmoïdes correspondant à quatre valeurs différentes de température. Les paramètres sont $\Theta_0=3$, $w_0=0,1$ et $N=100$: donc le point de symétrie de la sigmoïde a pour coordonnées $\Omega(30,50)$. On voit que pour $T=0,5$ et $T=0,8$, il existe trois points d'intersections : n_1^* (stable, proche de 0), n_2^* (instable, proche de 30) et n_3^* (stable, proche de 100). Ainsi, partant d'un état nul $n(0)=0$, c'est vers le point stable inférieur que l'activité moyenne va converger, c'est à dire un état satisfaisant où 0 à 3 neurones sont actifs à chaque instant. Comme le montre la figure de droite, le comportement serait d'ailleurs le même pour une initialisation $n(0)$ comprise entre n_1^* et n_2^* . En revanche, pour une température plus grande, $T=1,5$ ou $T=3$, il n'existe qu'un seul point d'intersection, stable et très proche de 100 : cette situation conduit donc inévitablement à l'"embrasement" du réseau, car, même en partant de $n=0$, l'activité atteindra $n=100$ en quelques itérations. L'effet serait d'ailleurs le même si Θ_0/w_0 était plus petit (i.e. des seuils plus petits ou des connexions plus fortes). Remarquons que si Θ_0/w_0 était au contraire plus grand que $N/2$ (grands seuils ou connexions faibles), le point Ω se trouverait à droite de la bissectrice, et les grandes températures auraient alors un effet inverse, ne laissant que le seul point d'intersection n_1^* .

La pente de σ est maximale en Ω et vaut : $(d\sigma/dn^*)_{\Omega} = Nw_0/4T$. Dans la recherche des points-solutions stables, on peut alors distinguer deux cas : (a) si $Nw_0/4T \leq 1$, il y a un seul point d'intersection, qui est stable; (b) si $Nw_0/4T > 1$, il peut y avoir 1, 2 ou 3 points d'intersections, selon l'abscisse de Ω (1 : stable; 2 : stable et métastable; 3 : stable, instable et stable).

En conclusion, cette petite étude statistique montre qu'avant d'envisager tout processus d'organisation des connexions, il faut régler correctement les paramètres T , Θ_0 , et w_0 pour se placer dans un domaine raisonnable d'évolution des activités. Pour obtenir la relation récurrente entre $\langle n(t+1) \rangle$ et $\langle n(t) \rangle$, et donc l'équation du point fixe $n^* = \mathcal{O}(n^*)$, nous avons dû procéder à quelques approximations, qui sont toutes fondées sur l'hypothèse d'un faible écart-type des distributions de seuils et de connexions, ainsi que des potentiels postsynaptiques.

Le nombre n^* n'est qu'une activité globale *moyenne*, c'est à dire représente $\langle n(t) \rangle$: de son côté, l'activité globale *instantanée*, $n(t)$, est le produit de fluctuations plus ou moins grandes autour de ce point stable : $n(t) \sim n^* \pm \Delta n$. Au total, l'amplitude de ces fluctuations dépend à la fois de la température et de la dispersion des seuils et des connexions, et peuvent se traduire par des phénomènes d'oscillations : si elles sont trop fortes, il peut arriver que l'état des neurones bascule d'un point stable à un autre (par exemple de n_1^* à n_3^*). Nous n'approfondirons pas plus l'étude des fluctuations mais on peut observer leurs effets sur la figure 2.

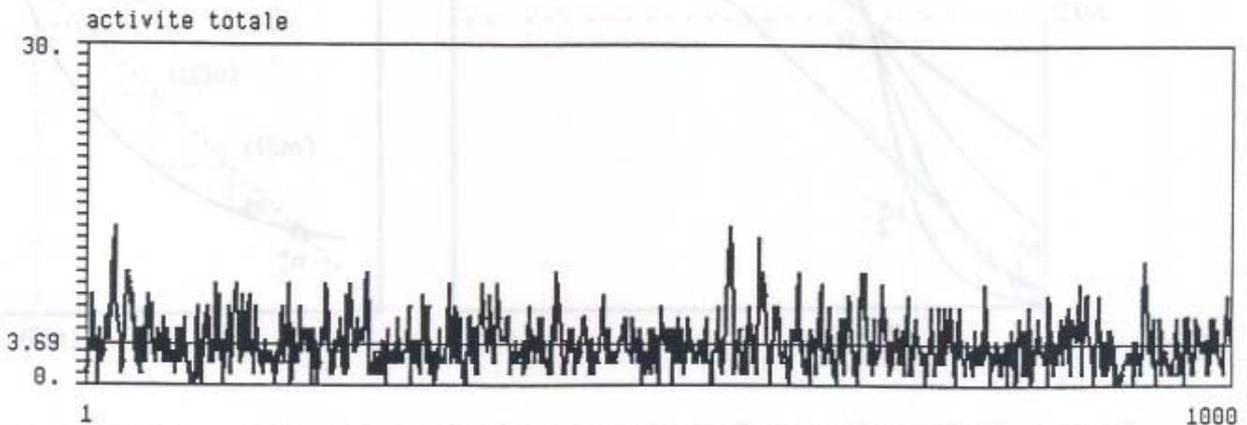


Figure 2 : Fluctuations de l'activité totale du réseau autour d'une valeur moyenne. Ce graphique représente l'évolution de $n(t)$ sur une période de 1000 itérations. Les paramètres sont ici : $\Theta_0=3$, $w_0=0,1$, $N=100$ et $T=0,8$, ce qui donne une valeur moyenne de $\approx 3,69$. La valeur prédite par l'équation $n^* = \mathcal{O}(n^*)$ est $n_1^* \approx 3,5$.

4.2 Une itération élémentaire

Dans un contexte d'activité moyenne faible, $n(t) \ll N$, nous nous intéressons à présent à la modulation synaptique, et plus précisément à une *étape élémentaire* de modification des connexions : autrement dit, nous cherchons à calculer les conséquences d'une transmission des activités $x(t) \rightarrow x(t+1)$ sur W , à travers la règle de potentiation et la règle de compétition formalisées au §3.2 (avec $\alpha > 0$, $\beta > 0$ et $\gamma > 0$). On rappelle que dans la version simplifiée du modèle que nous étudions ici, les neurones sont tous connectés les uns aux autres, et que les délais de transmission sont uniformes et valent 1. Par conséquent, tout événement de corrélation entre activités instantanées consécutives, même fortuit, aura une influence rétro-active

sur les connexions : il suffira que $x_i(t) = 1$ et $x_j(t+1) = 1$ pour que W_{ij} soit aussitôt légèrement augmenté. Ces incréments seront d'autre part soumis aux compétitions entre afférences et efférences et ne pourront pas être répercutés directement sans être corrigés : ils devront donc interagir et adopter des valeurs de compromis. Les variations finales ΔW_{ij} sont données par la somme des termes A_{ij} et B_{ij} .

L'état courant des connexions $W(t)$ sur lequel est greffé ce calcul d'une itération élémentaire pourra être a priori quelconque, dans la mesure où ses sommes entrantes et sortantes ne sont pas trop éloignées de la valeur normative s_0 . En effet, comme on le sait, les valeurs des connexions (W_{ij}) n'interviendront pas individuellement dans le calcul de leurs variations (ΔW_{ij}), mais seulement "collectivement" à travers leurs sommes S_i et S_j' , par différence avec s_0 (dans le terme **B**). Par conséquent, le degré d'hétérogénéité des connexions importera peu, dans la mesure où $S_i \approx S_j' \approx s_0$, et le calcul qui suivra sera donc valable aussi bien pour un réseau homogène que pour un réseau contenant déjà des structures particulières de connexions.

Pour simplifier, nous supposons dans un premier temps que $W(t)$ respecte *parfaitement* la conservation des sommes à s_0 , c'est à dire : $S_i(W(t)) = S_i'(W(t)) = s_0$ pour tout i . En effet, ceci aura l'avantage de mettre clairement en évidence les déviations apportées à ces sommes par le terme **A**, ainsi que l'effet réparateur, mais imparfait, du terme **B** qui les ramènera à des valeurs très proches de s_0 . On donnera ensuite l'expression de ΔW dans le cas général où les sommes de $W(t)$ sont seulement approximativement égales à s_0 : un terme correctif ϵ viendra alors s'ajouter à **A** et **B**.

4.2.a Effet des corrélations

Le groupe des neurones qui déchargent à l'instant t sera noté P , et le groupe des neurones qui déchargent à l'instant $t+1$ sera noté P' : ces groupes contiennent respectivement $n = n(t)$ et $n' = n(t+1)$ neurones. On suppose que n et n' sont petits devant N et que $P \cap P' = \emptyset$. Par conséquent, en vertu de la "règle de potentiation" appliquée sur cette échelle de temps (cf. §3.0), on observe immédiatement une augmentation de toutes les connexions sortant de P et entrant dans P' :

$$\forall i \in P, \forall j \in P', A_{ij}(t+1) = +\alpha$$

et on notera cet ensemble de connexions : $L_+ = P \times P'$. Il s'agit donc des connexions (i,j) concernées par un événement de transmission "réussi" $i|1 \rightarrow j|1$ dans l'itération $[t, t+1]$, c'est à dire telles que : $x_i(t) = 1$ et $x_j(t+1) = 1$.

D'autre part, en vertu de la "règle de dépression", on observe parallèlement une diminution des connexions qui sortent de P vers l'extérieur de P' , ainsi que des connexions qui entrent dans P' en provenance de l'extérieur de P :

$$\forall i \in P, \forall j \in P', i \neq j, A_{ij}(t+1) = -\beta$$

$$\forall i \notin P, \forall j \in P', i \neq j, A_{ij}(t+1) = -\beta$$

En effet, les premières sont le siège d'un événement "raté" de type $i|1 \rightarrow j|0$, et les deuxièmes de type $i|0 \rightarrow j|1$, c'est pourquoi toutes ces connexions sont décrémentées de β . On notera leurs ensembles respectifs L_- et L'_- (sont exclues de ces ensembles les connexions (i,i)).

Finalement, il reste une grande majorité de connexions qui ne sont concernées ni par le groupe P en sortie ni par le groupe P' en entrée, et qui sont donc inchangées ("non-événements" $i|0 \rightarrow j|0$):

$$\forall i \notin P, \forall j \in P', A_{ij}(t+1) = 0$$

Leur ensemble sera noté L_0 (c'est le complémentaire des trois autres sous-ensembles de connexions : $L_0 = R \times R \setminus (L_+ \cup L_- \cup L'_-)$, donc, formellement, il contient aussi toutes les connexions (i,i)).

En résumé, la corrélation immédiate entre l'activation des neurones de P à l'instant t et celle des neurones de P' à l'instant $t+1$, provoque une petite augmentation des efficacités de toutes les synapses qui connectent les premiers aux deuxièmes (connexions de L_+). Cette corrélation peut être fortuite ou, au contraire, induite par certains chemins synaptiques plus efficaces que les autres, mais quoiqu'il en soit, l'origine de cet événement n'est pas prise en compte dans la potentiation. Quant aux autres connexions, elles diminuent ou ne varient pas.

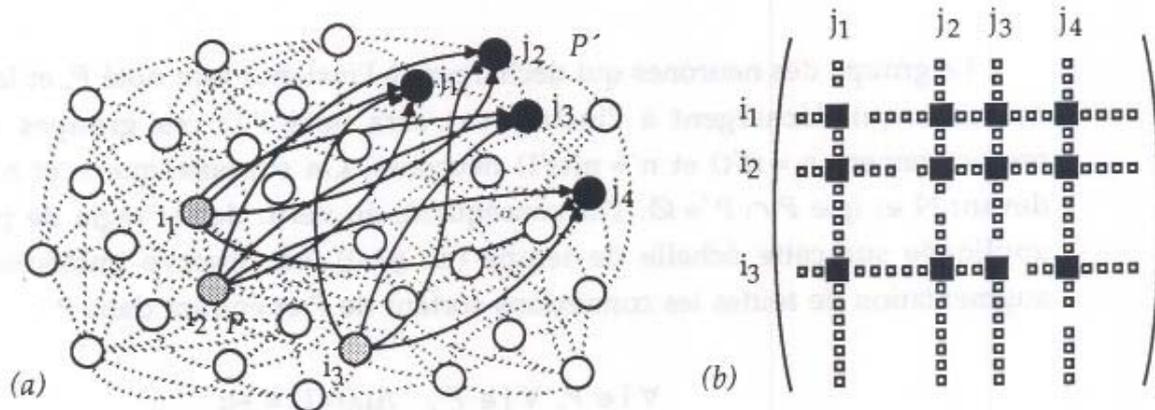


Figure 3 : Illustration de la potentiation des connexions. Sur cet exemple, trois neurones sont actifs à l'instant t : $P = \{i_1, i_2, i_3\}$, et quatre à l'instant $t+1$: $P' = \{j_1, j_2, j_3, j_4\}$. (a) Vue schématique du réseau, pour $\beta=0$: les neurones de P sont représentés en gris et ceux de P' en noir. La transmission des activités de P à P' provoque l'augmentation de toutes les connexions reliant P à P' (symbolisées par des arcs en trait noir), tandis que les autres connexions sont inchangées (symbolisées par des arcs en pointillés; mais pour ne pas surcharger le dessin, la connectivité n'est pas complète et les orientations ne sont pas indiquées). (b) Aspect général de la matrice A : aux intersections des lignes de P et des colonnes de P' se trouvent des carrés noirs symbolisant une augmentation $+\alpha$. Entre ces points, on a également fait figurer des carrés blancs symbolisant des diminutions $-\beta$, dans le cas où $\beta \neq 0$. La diagonale ainsi que toutes les autres composantes sont nulles.

Pour récapituler, on a donc :

- $(i,j) \in L_+ \Rightarrow A_{ij}(t+1) = +\alpha$
- $(i,j) \in L_- \Rightarrow A_{ij}(t+1) = -\beta$
- $(i,j) \in L'_- \Rightarrow A_{ij}(t+1) = -\beta$
- $(i,j) \in L_0 \Rightarrow A_{ij}(t+1) = 0$

L'aspect matriciel de \mathbf{A} ainsi que son effet sur le réseau sont illustrés sur la figure 3. Pour simplifier le calcul du terme \mathbf{B} , on supprimera l'effet pénalisateur de \mathbf{A} dans toute la suite en posant : $\beta = 0$.

4.2.b Effet de la compétition

Le nouvel état des connexions est donc provisoirement : $\mathbf{W}(t)+\mathbf{A}(t+1)$, et cette variation a pour conséquence de perturber l'équilibre de conservation des sommes entrantes et sortantes. En effet, l'augmentation des connexions de L_+ fait que les neurones de P ont maintenant une quantité efférente supérieure à s_0 , et que, de même, les neurones de P' ont une quantité afférente supérieure à s_0 . Précisément, les nouvelles valeurs de ces sommes (sortantes pour $i \in P$ et entrantes pour $j \in P'$) sont :

$$\begin{aligned} \forall i \in P, S_i(\mathbf{W}(t)+\mathbf{A}(t+1)) &= s_0 + n'\alpha \\ \forall j \in P', S_j'(\mathbf{W}(t)+\mathbf{A}(t+1)) &= s_0 + n\alpha \end{aligned}$$

tandis que les autres sommes n'ont pas été modifiées et valent toujours s_0 . Ces écarts par rapport à la valeur idéale s_0 vont donc provoquer l'apparition d'une "force de réaction", due à la compétition entre connexions, qui tendra à ramener les sommes vers s_0 . Cette force est formalisée par le terme \mathbf{B} , dont on rappelle ici l'expression (cf. §3.0) :

$$B_{ij}(t+1) = -2\gamma (S_i(\mathbf{W}(t)+\mathbf{A}(t+1)) - s_0) - 2\gamma (S_j'(\mathbf{W}(t)+\mathbf{A}(t+1)) - s_0)$$

Par conséquent, pour calculer les valeurs détaillées de $B_{ij}(t+1)$ sur toutes les connexions du réseau, on doit à nouveau distinguer quatre cas, suivant l'appartenance de i à P et de j à P' . On obtient alors :

- $i \in P$ et $j \in P' \Rightarrow B_{ij}(t+1) = -2\alpha\gamma(n+n')$
- $i \in P$ et $j \notin P' \Rightarrow B_{ij}(t+1) = -2\alpha\gamma n'$
- $i \notin P$ et $j \in P' \Rightarrow B_{ij}(t+1) = -2\alpha\gamma n$
- $i \notin P$ et $j \notin P' \Rightarrow B_{ij}(t+1) = 0$

et on retrouve exactement les quatre sous-ensembles de connexions L_+ , L_- , L'_- et L_0 définis plus haut.

Ainsi, on voit que la réaction du réseau à l'augmentation soudaine $+\alpha$ d'une

connexion (i,j) de L_+ , se traduit par l'affaiblissement de toutes les connexions qui partagent un neurone avec celle-ci, à savoir : les connexions $(i,j')_{j' \neq j}$ de L_- et les connexions $(i',j)_{i' \neq i}$ de L_-' . Donc ces variations vont bien dans le sens d'un rétablissement de la valeur idéale des sommes, s_0 . Remarquons par ailleurs que le contre-coup sur les connexions de L_+ est encore plus important que celui qui affecte L_- ou L_-' , car il conjugue les pénalités de ces deux ensembles. Cependant, les connexions de L_+ devront finalement l'emporter grâce au gain $+\alpha$ du terme A , ce qui constitue l'objet de la discussion suivante.

4.2.c Variation totale des connexions

Finalement, la variation totale des connexions entre les instants t et $t+1$ est donnée par la somme du terme de corrélation et du terme de compétition : $\Delta W_{ij}(t+1) = A_{ij}(t+1) + B_{ij}(t+1)$, c'est à dire :

- $(i,j) \in L_+ \Rightarrow \Delta W_{ij}(t+1) = \alpha(1-2\gamma(n+n'))$
- $(i,j) \in L_- \Rightarrow \Delta W_{ij}(t+1) = -2\alpha\gamma n'$
- $(i,j) \in L_-' \Rightarrow \Delta W_{ij}(t+1) = -2\alpha\gamma n$
- $(i,j) \in L_0 \Rightarrow \Delta W_{ij}(t+1) = 0$

et le nouvel état des connexions à $t+1$ vaut alors : $W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t+1)$, sous la condition $0 \leq W_{ij}(t+1) \leq 1$. Par conséquent, on voit que l'événement constitué par l'activation de P' consécutive à celle de P a pour effet principal de faire grandir les connexions de L_+ , c'est à dire l'ensemble des connexions de P à P' , au détriment des autres connexions sortant de P ou entrant dans P' . Toutefois, cette croissance ne sera appréciable que si $2\gamma(n+n') \ll 1$ (les connexions de L_+ ne devant pas être paradoxalement victimes d'une variation négative, à cause d'un terme B_{ij} trop fortement répressif). Sachant que n et n' sont petits devant N , nous prendrons alors : $\gamma = 1/2N$, ce qui est bien en accord avec la condition ci-dessus (les raisons de ce choix particulier seront commentées plus bas). Finalement, avec cette valeur de γ , la variation totale de t à $t+1$ s'écrit :

- $(i,j) \in L_+ \Rightarrow \Delta W_{ij}(t+1) = \alpha(1-(n+n')/N)$
- $(i,j) \in L_- \Rightarrow \Delta W_{ij}(t+1) = -\alpha n'/N$
- $(i,j) \in L_-' \Rightarrow \Delta W_{ij}(t+1) = -\alpha n/N$
- $(i,j) \in L_0 \Rightarrow \Delta W_{ij}(t+1) = 0$

Exemple numérique : avec $N=100$, $n=n'=5$ et $\alpha=0,1$, on obtient respectivement : $+0,09$, $-0,005$, $-0,005$ et 0 (variations à l'intérieur de l'intervalle $[0, 1]$). Il est évident que l'effet principal de ΔW provient de A , car si l'on négligeait les termes d'ordre (n/N) , il ne resterait que : $+\alpha$, 0 , 0 et 0 . Ainsi, au niveau de chaque connexion (i,j) , la part de B est négligeable devant celle de A .

En revanche, au niveau des *sommes* de connexions, on peut vérifier que le terme **B** est justement du même ordre de grandeur **A**, et contrebalance son action pour rétablir la norme s_0 . Calculons par exemple la variation des sommes sortantes S_i entre t et $t+1$:

$$\Delta S_i = S_i(\mathbf{W}(t+1)) - S_i(\mathbf{W}(t)) = S_i(\Delta \mathbf{W}(t+1)).$$

(on ignore ici les problèmes de troncation, en supposant que les W_{ij} se trouvent encore éloignés des extrêmes 0 et 1). Tout neurone i possède n' efférences qui vont vers P' , et $N-n'$ qui vont vers l'extérieur de P' , donc on a :

- $i \in P \Rightarrow \Delta S_i \approx n'[\alpha(1-(n+n')/N)] + (N-n')[-\alpha n'/N] = -\alpha n n'/N$
- $i \notin P \Rightarrow \Delta S_i \approx n'[-\alpha n/N] + (N-n')[0] = -\alpha n n'/N$

(pour simplifier davantage, on a négligé le fait que $W_{ii} = 0$). Un calcul symétrique fournit alors le même résultat pour les sommes entrantes S_j' , et finalement, on obtient :

$$\forall i \in R, \forall j \in R, \Delta S_i \approx \Delta S_j' \approx -\alpha n n'/N$$

Donc ceci prouve que les perturbations des sommes sont toutes du même ordre de grandeur, et que cette valeur commune est faible : par exemple, pour $N=100$, $n=n'=5$ et $\alpha=0,1$, ΔS_i et $\Delta S_j'$ valent $-0,025$, c'est à dire une variation relative de 0,25% pour une norme typique de $s_0=10$. S'il n'y avait pas la compétition, les écarts des sommes seraient beaucoup plus grands : $+n'\alpha$ en sortie sur P et $+n\alpha$ en entrée sur P' , c'est à dire 5% de s_0 , et leur accumulation finirait par déstabiliser le réseau.

En conclusion, si le réseau des connexions est initialement dans un état parfait à l'instant t (toutes les sommes valant s_0), il se retrouve à l'instant $t+1$, après action de **A** et **B**, dans un état légèrement imparfait, mais encore très satisfaisant. On peut aussi mesurer cet effet sur les valeurs de la pénalité quadratique H : avec les valeurs numériques ci-dessus, le calcul donne : $H(\mathbf{W}) = 0$, puis $H(\mathbf{W}+\mathbf{A}) = 1,25 \cdot 10^{-2}$, puis $H(\mathbf{W}+\mathbf{A}+\mathbf{B}) = 6,25 \cdot 10^{-4}$. Donc le terme **B** parvient à réduire notablement la pénalité provoquée par **A** (ici, d'un facteur 20), et à la ramener très près de 0.

Nous appellerons la matrice $\Delta \mathbf{W}(t+1)$ calculée ci-dessus "*empreinte de liaison*" entre les groupes de neurones P et P' , et dans toute la suite on la notera spécifiquement : $\mathbf{Z}[P \rightarrow P']$. Donc avec cette notation, l'évolution des connexions entre t et $t+1$ pourra s'écrire de façon générale (toujours dans les limites $0 \leq W_{ij} \leq 1$) :

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \mathbf{Z}[P(t) \rightarrow P(t+1)]$$

où $P(t)$ désigne l'ensemble des neurones actifs à l'instant t : $P(t) = \{i \in R; x_i(t) = 1\}$. Ce module matriciel \mathbf{Z} sera présent tout au long de l'évolution de \mathbf{W} : il interviendra

systematiquement entre deux groupes de neurones successivement activés, par addition sur les valeurs courantes des connexions, en suivant la propagation des activations d'une itération à l'autre. Son rôle est donc celui d'un *élément de base de la construction des chaînes*. Pour résumer, les caractéristiques principales d'une telle matrice sont : (a) des valeurs positives aux intersections des lignes $i \in P$ et des colonnes $j \in P'$ (connexions sortant de P et entrant dans P'); (b) des valeurs négatives, plus faibles que les précédentes, sur la différence symétrique entre les lignes $i \in P$ et les colonnes $j \in P'$ (autres connexions sortant de P ou entrant dans P'); (c) des valeurs nulles partout ailleurs (figure 4).

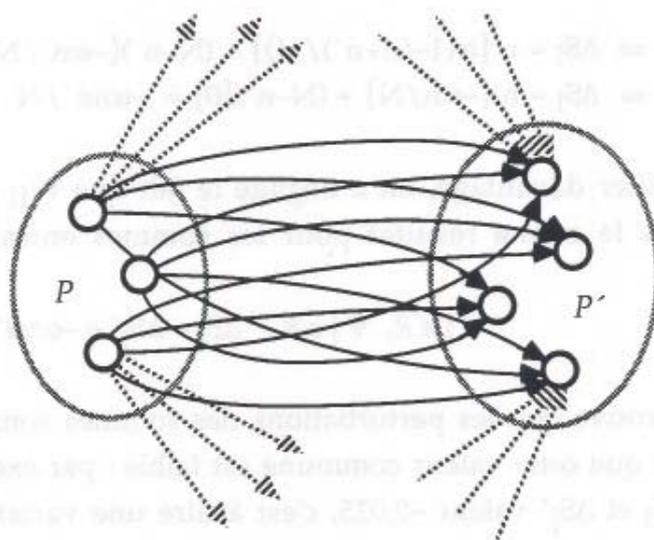


Figure 4 : Illustration de l'"empreinte de liaison" $Z[P \rightarrow P']$. Les connexions de P à P' sont renforcées (trait noir) tandis que les autres connexions sortant de P ou entrant dans P' sont affaiblies (trait pointillé).

Tel que nous l'avons calculé, le schéma $Z[P \rightarrow P']$ est donc applicable à n'importe quel couple (P, P') de parties du réseau R , dans la mesure où celles-ci n'ont pas de neurones en commun. Remarquons que cette condition est satisfaite dans la plupart des cas : qu'il s'agisse de processus d'activation désordonnés ou d'une propagation régulière, un neurone sera rarement actif en deux instants consécutifs. Si toutefois il se trouve que $P \cap P' \neq \emptyset$, l'aspect de Z sera un peu différent à cause de certaines variations ΔW_{ij} qui seront obligatoirement nulles au lieu de valoir $\approx +\alpha$, et cette restriction permettra alors aux autres connexions d'être un peu plus grandes.

Remarquons par ailleurs que l'hypothèse $n, n' \ll N$ est indépendante de la formulation de Z , qui serait également valable pour des valeurs d'activité grandes. Cependant, comme on l'a montré plus haut, le fait d'adopter $\gamma = 1/2N$ n'a d'intérêt que si n et n' restent petits devant N . En effet, si on avait par exemple $n=n'=N/2$, les quatre valeurs de la matrice seraient : $0, -\alpha/2, -\alpha/2$ et 0 , ce qui voudrait dire que les connexions de L_0 sont maintenues au même niveau que celles de L_+ , tandis que celles de L_- et L'_- sont réprimées (schéma "en creux" à rejeter). Donc $Z[P(t) \rightarrow P(t+1)]$ ne sera vraiment constructif que si $n(t) \ll N$.

Il reste que l'état des connexions W qui a servi de base au calcul était assez particulier dans la mesure où ses sommes entrantes et sortantes étaient *exactement* égales à s_0 . Cependant, nous pouvons affirmer que si W ne vérifiait pas exactement mais seulement approximativement ces sommes, le calcul de la variation ΔW donnerait quand-même un résultat très voisin de la matrice de liaison $Z[P \rightarrow P']$. Donc, en réalité, la relation de récurrence écrite plus haut n'est pas tout à fait exacte, mais reste très proche de la vraie formule, qui est :

$$W(t+1) = W(t) + Z[P(t) \rightarrow P(t+1)] + \epsilon(t)$$

où ϵ est un terme correctif, négligeable devant Z : $\|\epsilon\| \ll \|Z\|$. Plus précisément, il est facile de vérifier que ce terme correctif est proportionnel aux "résidus" des sommes de W , c'est à dire :

$$\epsilon_{ij}(t) = -2\gamma(\xi_i(t) + \xi_j'(t))$$

où :

$$S_i(W(t)) = s_0 + \xi_i(t) \text{ et } S_j'(W(t)) = s_0 + \xi_j'(t)$$

En outre, en calculant la relation de récurrence de $\epsilon(t)$ à $\epsilon(t+1)$ on peut montrer que l'amplitude de ϵ ne diverge pas au cours du temps ce qui signifie que le processus de correction des sommes par le terme B_{ij} est stable, et le calcul montre que la condition de cette stabilité est : $\gamma \leq 1/2N$. Cependant, même dans le cas où γ est un peu supérieur à $1/2N$, la forte non-linéarité introduite par la troncation systématique des valeurs W_{ij} qui sortent de l'intervalle $[0, 1]$ fait que l'ordre de grandeur de ϵ reste de toutes façons petit. Quoi qu'il en soit, on doit avoir dans tous les cas :

$$\gamma N \sim 1 \text{ et } \gamma n \ll 1$$

pour que les connexions de $P \times P'$ augmentent.

Voici à présent l'ordre de grandeur des composantes de Z dans le cas $\beta > 0$:

- $(i,j) \in L_+ \Rightarrow Z_{ij}[P \rightarrow P'] \approx \alpha + 4\gamma N \beta$
- $(i,j) \in L_- \Rightarrow Z_{ij}[P \rightarrow P'] \approx -2\gamma n' \alpha + (2\gamma N - 1)\beta$
- $(i,j) \in L_- \Rightarrow Z_{ij}[P \rightarrow P'] \approx -2\gamma n \alpha + (2\gamma N - 1)\beta$
- $(i,j) \in L_0 \Rightarrow Z_{ij}[P \rightarrow P'] \approx 2\gamma(n+n')\beta$

(dans cette expression simplifiée, on a négligé tous les termes d'ordre γn devant 1 et devant γN , à l'intérieur des facteurs respectifs de α et β). On voit donc que si $\gamma = 1/2N$, les termes principaux de cette matrice sont : $\alpha + 2\beta$, 0, 0 et 0. Dans ce cas, à cause de la réaction du terme B_{ij} , le paramètre β qui devait agir en décrétement sur les connexions de L_- et L_-' se retrouve finalement comme incrément sur les connexions de L_+ . Ainsi, avec $\beta > 0$, la croissance des connexions L_+ est accélérée, tandis que l'écart se creuse entre les connexions de L_- et L_-' et celles de L_0 .

En conclusion, la variation des connexions est toujours calquée sur le même modèle : $\Delta W \approx Z$, dans la mesure où l'état des connexions W auquel cette matrice vient s'ajouter respecte approximativement la conservation des sommes : elle garantit alors la perpétuation de cette conservation, moyennant la petite correction ε que nous avons évoquée ci-dessus. Ainsi, la formule de $Z[P(t) \rightarrow P(t+1)]$ est applicable en pratique à toutes les époques, quel que soit l'état d'hétérogénéité des connexions ou des probabilités d'activation. C'est pourquoi cette matrice agit véritablement comme une "empreinte" : à chaque itération, elle vient en surimpression des valeurs courantes des connexions, un peu à la manière d'un tampon faisant apparaître un motif imprimé. L'emplacement de cette configuration stéréotypée dépendra du comportement des activités, car elle s'accumulera systématiquement entre les deux derniers groupes de neurones activés, P et P' . En d'autres termes, on peut dire que l'empreinte de liaison se dépose dans le "sillage" de la propagation.

4.3 Connexions seules sous un régime forcé

Après avoir calculé une étape élémentaire de l'évolution des connexions, nous l'utilisons à présent pour décrire les phénomènes d'auto-organisation susceptibles d'apparaître sur le réseau. Dans un premier temps, au §4.3, on se limitera à une étude partielle du modèle, mais sous les conditions inverses de celles du paragraphe §4.1 : au lieu de fixer les connexions et d'étudier les activités, on forcera cette fois le processus d'activation pour étudier ses conséquences sur les connexions. Cette situation artificielle servira de préambule aux paragraphes §4.4 et §4.5, où sera finalement abordée l'étude du réseau dans des conditions complètes de fonctionnement. Ainsi, le présent paragraphe §4.3 est-il destiné à mieux comprendre les conditions d'apparition d'une structure de chaîne.

On rappelle que : $\Delta W_{ij}(t+1) = Z_{ij}[P(t) \rightarrow P(t+1)] + \varepsilon_{ij}(t)$, avec $\|\varepsilon\| \ll \|Z\|$, et qu'il faut également tenir compte des bornes $[0, 1]$ qui encadrent à chaque instant les connexions W_{ij} . Donc, en réalité, la véritable variation n'est pas ΔW , mais une matrice tronquée ΔW^* qui vérifie : $|\Delta W_{ij}^*| \leq |\Delta W_{ij}|$. Cependant, tant que la valeur d'une connexion W_{ij} est suffisamment éloignée des bornes 0 et 1, l'effet de la troncature n'apparaît pas et ΔW_{ij}^* est identique à ΔW_{ij} . En partant à l'instant 0 de l'état homogène de connexions W^0 , l'évolution générale des connexions peut alors s'écrire (en ne retenant que la partie principale Z) :

$$\begin{aligned} W(0) &= W^0 \\ W(1) &\approx W^0 + Z[P(0) \rightarrow P(1)] \\ W(2) &\approx W^0 + Z[P(0) \rightarrow P(1)] + Z[P(1) \rightarrow P(2)] \\ &\dots \\ W(t) &\approx W^0 + Z[P(0) \rightarrow P(1)] + Z[P(1) \rightarrow P(2)] + \dots + Z[P(t-1) \rightarrow P(t)] \end{aligned}$$

Les empreintes de liaison sont ajoutées les unes après les autres, en suivant le déplacement du groupe d'activation $P(t)$, chaque ajout étant conditionné par une éventuelle saturation des valeurs.

Quelle structure de connexions peut-on alors espérer voir apparaître ?

Il est clair que l'organisation du réseau dépendra entièrement du comportement des activités au cours du temps, car c'est justement à l'endroit des transmissions réussies de potentiels $i \rightarrow j$ que les connexions seront augmentées. Donc nous ne pouvons rien déduire de la formule ci-dessus quant à une éventuelle organisation des connexions, tant que nous ne connaissons pas par ailleurs, réciproquement, la composition des groupes actifs $P(t)$. Ainsi, le problème revient à déterminer l'évolution des activités $x(t)$, et à comprendre comment les processus de décharge, au départ totalement aléatoires, seront progressivement influencés et canalisés *en retour* par la focalisation de l'innervation qu'ils auront eux-mêmes provoquée. C'est justement cette *dépendance mutuelle immédiate* entre corrélations et connexions, et sa capacité à rendre compte d'une auto-organisation cohérente du réseau à partir d'un état aléatoire informe qui constituera le cœur du modèle.

Mais pour l'instant, avant d'étudier le comportement réel des activités évoluant conjointement avec les connexions, nous déplacerons la question précédente dans le domaine conditionnel : Si les activations suivaient telle ou telle loi $P(t)$, quelle structure verrait-on alors apparaître ? Ainsi, au lieu de suivre le réseau pas à pas, en inférant W à partir de x , puis x à partir de W à chaque itération, il sera plus instructif, pour commencer, de projeter directement l'état global $(x(t), W(t))$ dans une situation hypothétique, caractérisée par un certain comportement de $x(t)$, imposé a priori, et donc une certaine évolution de $W(t)$. C'est ainsi que nous verrons deux cas de figure très différents, caractérisés par des processus de décharge radicalement opposés : au §4.3.a on envisagera un cas de propagation totalement aléatoire, tandis qu'au §4.3.b on supposera au contraire que la propagation est ordonnée, et repasse régulièrement par les mêmes groupes de neurones.

4.3.a Propagation aléatoire : recoupements et destruction

Le régime d'activités étudié ici sera semblable à celui qui s'était développé spontanément au §4.1 : le réseau est caractérisé par des états $x(t)$ aléatoires, dont l'activité totale $n(t)$ fluctue autour d'une valeur moyenne stable $\langle n(t) \rangle \approx n^*$ petite devant N . Pour simplifier, nous assimilerons les probabilités d'activation à une constante n^*/N . Ce mode aléatoire sera noté $x^a(t)$ et vérifie en conséquence :

$$\forall t, \forall j \in R, \text{proba}[x_j^a(t)=1] = n^*/N$$

On posera alors a priori : $\forall t, x(t) = x^a(t)$. Nous examinons à présent la façon dont ces activités aléatoires vont agir sur les connexions W , et pour cela, on reprendra les paramètres adoptés au §4.2, à savoir : $\alpha > 0$, $\beta = 0$, et $\gamma = 1/2N$.

Sous ces conditions de propagation, intéressons-nous au destin d'une connexion particulière W_{ij} , et plus précisément à son espérance mathématique $\langle W_{ij}(t) \rangle$ à un instant t . De façon générale, cette valeur est le résultat d'un calcul statistique portant sur tous les événements d'activations qui sont survenus dans le réseau depuis l'instant initial 0 jusqu'à t , soit : $x(0), x(1), \dots, x(t)$, pondérés par leur probabilité conjointe d'apparition. En clair, il s'agit de la moyenne :

$$\langle W_{ij}(t) \rangle = \sum_{x_0, x_1, \dots, x_t} \text{proba}[x(0)=x_0, x(1)=x_1, \dots, x(t)=x_t] W_{ij}(t)$$

où les vecteurs x_0, x_1, \dots, x_t parcourent tous les états possibles du domaine discret $\{0,1\}^N$ ($W_{ij}(t)$ étant considérée comme une fonction des $t+1$ états en question).

Il est alors possible de calculer cette quantité *par récurrence*, grâce à la relation entre $W_{ij}(t+1)$ et $W_{ij}(t)$ exposée plus haut. Dans ce but, on négligera pour l'instant le terme correctif ϵ , et d'autre part, comme les activités ne dépendent pas des connexions dans ce paragraphe, on pourra renvoyer la troncature à la fin des calculs : cette économie est donc tolérable dans la mesure où elle ne modifie pas le comportement des activations, comme c'est le cas ici. Ainsi, en notant plus simplement $p[x_0, x_1, \dots, x_t]$ la probabilité conjointe à $t+1$ états, on obtient :

$$\begin{aligned} \langle W_{ij}(t) \rangle &= \sum_{x_0, \dots, x_t} p[x_0, \dots, x_t] (W_{ij}(t-1) + Z_{ij}[x_{t-1} \rightarrow x_t]) \\ &= \sum_{x_0, \dots, x_{t-1}} p[x_0, \dots, x_{t-1}] W_{ij}(t-1) + \sum_{x_{t-1}, x_t} p[x_{t-1}, x_t] Z_{ij}[x_{t-1} \rightarrow x_t] \\ &= \langle W_{ij}(t-1) \rangle + \langle Z_{ij}[x(t-1) \rightarrow x(t)] \rangle. \end{aligned}$$

Dans cette décomposition, le terme Z_{ij} ne dépend par définition que du dernier événement d'activation : $x(t-1) \rightarrow x(t)$, donc son espérance est calculée sur la loi marginale $p[x(t-1), x(t)]$. Ainsi, sachant que l'état initial $W_{ij}(0)$ vaut W_{ij}^0 , l'espérance de $W_{ij}(t)$ sera donnée par :

$$\langle W_{ij}(t) \rangle = W_{ij}^0 + \langle Z_{ij}[x(0) \rightarrow x(1)] \rangle + \dots + \langle Z_{ij}[x(t-1) \rightarrow x(t)] \rangle.$$

Remarquons que cette formule est valable de façon générale pour n'importe quel processus $x(t)$ a priori non-stationnaire.

On rappelle d'autre part que la variation $Z_{ij}(t)$ est déterminée par le succès de la transmission qui a lieu du neurone i vers le neurone j , c'est à dire par l'événement $x_i(t-1) \rightarrow x_j(t)$: s'il s'agit de $i|1 \rightarrow j|1$ (événement "réussi") la connexion W_{ij} sera renforcée, tandis qu'après $i|1 \rightarrow j|0$ ou $i|0 \rightarrow j|1$ (événements "ratés") elle se trouvera légèrement affaiblie. D'après les calculs effectués au §4.2, les quantités associées à ces modifications dépendent des nombres $n(t-1)$ et $n(t)$, et on a donc en général :

$$\begin{aligned} \langle Z_{ij}[x(t-1) \rightarrow x(t)] \rangle = & \text{proba}[x_i(t-1)=1, x_j(t)=1] (+\alpha(1-(n(t-1)+n(t))/N)) \\ & + \text{proba}[x_i(t-1)=1, x_j(t)=0] (-\alpha n(t)/N) \\ & + \text{proba}[x_i(t-1)=0, x_j(t)=1] (-\alpha n(t-1)/N) \end{aligned}$$

Dans le cas aléatoire uniforme $x(t) = x^a(t)$ qui nous occupe à présent (processus stationnaire), les activités sont *décorrélées* et le nombre de neurones actif vaut à tout instant n^* : l'espérance de Z_{ij} sera donc constante et uniforme sur le réseau :

$$\begin{aligned} \langle Z_{ij}[x^a(t-1) \rightarrow x^a(t)] \rangle = & (n^*/N)(n^*/N) (+\alpha(1-2n^*/N)) \\ & + (n^*/N)(1-n^*/N) (-\alpha n^*/N) \\ & + (1-n^*/N)(n^*/N) (-\alpha n^*/N) = -\alpha(n^*/N)^2 \end{aligned}$$

Par conséquent, l'espérance globale de W_{ij} à l'instant t s'écrira :

$$\langle W_{ij}(t) \rangle = W_{ij}^0 - \alpha(n^*/N)^2 t$$

Ceci voudrait dire que, globalement, l'ensemble des connexions a tendance à être affaibli, et chute finalement vers 0 au bout d'un certain nombre d'itérations : il devrait alors stationner, ne pouvant descendre vers les nombres négatifs à cause de la troncation (par exemple, avec $n^*=5$, $N=100$ et $\alpha=0,1$, $\alpha(n^*/N)^2$ vaut $2,5 \cdot 10^{-4}$, c'est à dire 400 fois moins que la valeur initiale des connexions $w_0=0,1$).

Cependant, ce résultat n'est pas exact car nous avons négligé le terme correctif ϵ_{ij} : un calcul plus précis qui tiendrait compte de ce terme fournirait en effet une variation moyenne *nulle* : $\langle Z_{ij} + \epsilon_{ij} \rangle = 0$ (ce calcul n'est pas présenté ici). Donc, en réalité, l'espérance de W_{ij} ne décroît pas mais stationne à tout instant à sa valeur initiale :

$$\forall t, \langle W_{ij}(t) \rangle = W_{ij}^0$$

(ce qui montre au passage que ϵ_{ij} est important pour maintenir les sommes proches de s_0).

Quoiqu'il en soit, que l'on ait fait l'approximation $\Delta W = Z$, ou que l'on ait gardé rigoureusement $\Delta W = Z + \epsilon$, on voit que la variation moyenne des connexions sous un régime d'activations totalement décorréliées vérifie, à peu de choses près, les trois propriétés suivantes : elle est identique sur tout le réseau,

invariable dans le temps, et proche de 0. Comme c'était prévisible, on peut donc conclure que *rien ne se construit sous un régime d'activations aléatoires* : l'organisation d'une éventuelle structure est perpétuellement remise en cause par l'augmentation anarchique, c'est à dire *non concertée*, des connexions W_{ij} . A chaque instant, celles-ci ont toutes la même probabilité de transmettre un événement $i|1 \rightarrow j|1$, donc d'être renforcées, et elles ont également la même probabilité de transmettre un événement $i|1 \rightarrow j|0$ ou $i|0 \rightarrow j|1$, donc d'être affaiblies. Au total, toutes les connexions ont la *même espérance de variation*, et de cette homogénéité aucune structure ne peut évidemment surgir.

Il est également possible de décrire en termes de *chaînes* ce que le calcul statistique global a révélé sur la variation des connexions individuelles, le terme "chaîne" faisant référence à une succession de quelques groupes disjoints P, P', P'', \dots reliés par des connexions légèrement plus fortes que la moyenne : $Z[P \rightarrow P'] + Z[P' \rightarrow P''] + \dots$. On peut donc considérer que la propagation aléatoire ne cesse d'esquisser des morceaux de chaîne qu'elle s'empresse aussitôt d'effacer. En effet, dans un premier temps, tout groupe $P(t)$ sera en général suivi par l'activation de quelques autres groupes, $P(t+1), P(t+2), \dots, P(t+m)$, qui n'ont aucun neurone en commun avec $P(t)$, et cette succession provoquera le renforcement de toutes les connexions qui vont de l'un à l'autre : on peut donc voir apparaître une amorce de chaîne. Mais celle-ci sera ensuite rapidement brouillée par la poursuite des activations et finira par disparaître : à partir de $t+m+1$, quelques neurones appartenant à ces paquets seront réactivés au hasard, en même temps que d'autres neurones du reste du réseau, ce qui aura pour conséquence de faire croître de nouvelles connexions, et donc, à travers la compétition, d'affaiblir les connexions qui avaient commencé à se former dans la chaîne $t, \dots, t+m$. Ainsi, les éventuelles ébauches de chaînes ne sont jamais confirmées dans leur croissance, donc ne peuvent pas achever leur maturation, car la probabilité pour que les activations repassent ultérieurement par les mêmes paquets (ne serait-ce que partiellement) dans l'ordre qu'ils avaient initialement, est infiniment faible. Bien au contraire, ces ébauches sont détruites par les tentatives d'autres chaînes concurrentes, qui prennent racine sur des regroupements de neurones différents.

On voit donc que la propagation change incessamment de "direction" : les neurones appartenant aux derniers groupes activés sont sans cesse mélangés et redistribués dans de nouveaux groupes. Dans ce jeu à somme nulle, tous les liens veulent augmenter à la fois, mais comme ils sont entravés par la compétition, aucun n'augmente vraiment : les chances sont trop également réparties sur le réseau, il y a trop de symétrie, donc pas d'organisation particulière.

En conclusion, ce régime aléatoire n'a pas besoin d'être imposé a priori : il peut se poursuivre tout seul sous la double dynamique complète, c'est à dire avec une dépendance normale des activations neuronales en fonction des potentiels d'action. En effet, les variations des connexions étant globalement homogènes et faibles, les

processus de décharge aléatoires continueront à peu près dans les mêmes conditions (avec cependant plus de fluctuations), ce qui contribuera rétro-activement à maintenir les connexions dans leur état de neutralité moyenne, et ainsi de suite. L'état du réseau caractérisé par des activités aléatoires et des connexions homogènes est donc un *état stable* de la double dynamique.

4.3.b Propagation régulière : création d'une chaîne

La conclusion précédente nous fait comprendre la nécessité d'une *régulation* de la propagation des activités, que celle-ci soit imposée par un système extérieur, ou, ce qui est évidemment bien mieux, qu'elle apparaisse spontanément. C'est ainsi que, dans cette présente partie, nous verrons comment on peut facilement créer une chaîne de connexions grâce à une propagation régulière forcée, puis plus tard, aux paragraphes 4.4 et 4.5, comment cette propagation peut en réalité s'"auto-entretenir" à travers la chaîne qu'elle est elle-même en train de créer, sans la présente contrainte (les chaînes sont également des états stables de la double dynamique). Dans l'immédiat, il sera donc utile de définir une chaîne, en examinant également les conditions idéales de sa construction et de sa entretien, avant de voir plus tard comment ces conditions seront simultanément créées par la chaîne elle-même.

Dans le contexte simplifié où les délais de transmission sont tous égaux, les structures de connexions recherchées sont les "*synfire-chains*" (Abeles, 1982). Ce sont les cas particuliers les plus simples de "tresses", c'est à dire de focalisations permettant de soutenir des CST linéaires (voir les définitions de ces termes dans l'introduction de ce chapitre 4). De telles chaînes sont simplement constituées de *groupes de neurones reliés en série* : les connexions vont toujours d'un groupe au groupe suivant, sans sauter d'étapes et sans relier les neurones à l'intérieur d'un même groupe (figure 5).

De façon générale, on définira une synfire-chain R_c comme étant une succession de $m+1$ groupes de neurones $\{P_0, P_1, \dots, P_m\}$ disjoints deux à deux, de tailles à peu près identiques $n_0 \approx n_1 \approx \dots \approx n_m$, et par un état spécial de connexions W^c qui relie ces groupes de proche en proche de la façon suivante :

$$\forall k = 0, \dots, m-1,$$

- $i \in P_k$ et $j \in P_{k+1} \Rightarrow W_{ij}^c \approx 1$
- $i \in P_k$ et $j \notin P_{k+1} \Rightarrow W_{ij}^c \approx 0$
- $i \notin P_k$ et $j \in P_{k+1} \Rightarrow W_{ij}^c \approx 0$

Ainsi, les connexions qui relient deux groupes consécutifs, c'est à dire $P_k \times P_{k+1}$, seront prédominantes, tandis que les connexions qui relient ces groupes d'une autre manière, c'est à dire $P_k \times P_{k' \neq k+1}$, ou bien qui relient les groupes et l'extérieur de la chaîne, $P_k \times (R \setminus R_c)$, seront très faibles (quant aux connexions entièrement

extérieures à la chaîne, $(R \setminus R_c) \times (R \setminus R_c)$, leurs valeurs pourront être quelconques). Il s'agit donc d'une chaîne complètement connectée, qui contient en tout et pour tout : $n_0 n_1 + n_1 n_2 + \dots + n_{m-1} n_m$ connexions fortes. Pour simplifier, on supposera dans la suite que tous les n_k sont égaux et on notera n_c leur valeur commune (n_c est la "largeur" de la chaîne). Remarquons que pour respecter la conservation des sommes ($S_i(W^c) = S_j(W^c) = s_0$), le nombre n_c doit être justement du même ordre de grandeur que s_0 (le rapport entre n_c et s_0 et ses conséquences sur les valeurs des connexions sera commenté plus longuement au §4.4. et §4.5).

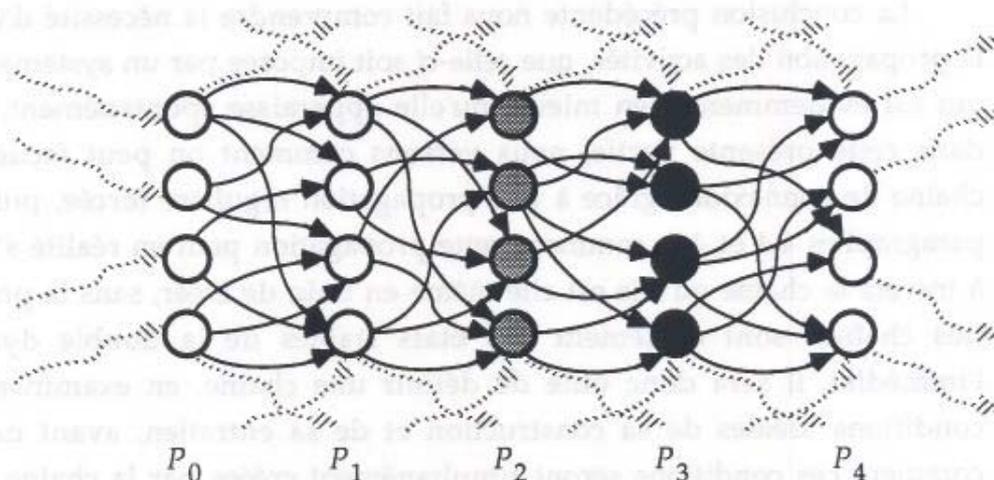


Figure 5 : Illustration d'une "synfire-chain". Les connexions d'un groupe P_k au groupe P_{k+1} sont fortes, tandis que les connexions sortant ou entrant de la chaîne sont faibles (les connexions ne sont pas toutes dessinées ici). Par ailleurs, sur ce schéma, une "onde d'activation" (voir texte) est en train de se propager sur la chaîne; à l'instant t représenté, on a : $P(t) = P_3$ (neurones en noir) et $P(t-1) = P_2$ (neurones en gris).

Il est alors clair que le processus organisateur du présent modèle, qui agit par additions d'empreintes de liaison Z , est parfaitement adapté à la construction et à l'entretien d'une chaîne comme celle-ci, car les variations apportées par les matrices Z prises entre deux paquets consécutifs P_k et P_{k+1} vont précisément dans le sens des valeurs finales de W^c :

$$\forall k = 0, \dots, m-1,$$

- $i \in P_k$ et $j \in P_{k+1} \Rightarrow Z_{ij}[P_k \rightarrow P_{k+1}] = +\alpha(1-2n_c/N) > 0$
- $i \in P_k$ et $j \notin P_{k+1} \Rightarrow Z_{ij}[P_k \rightarrow P_{k+1}] = -\alpha n_c/N < 0$
- $i \notin P_k$ et $j \in P_{k+1} \Rightarrow Z_{ij}[P_k \rightarrow P_{k+1}] = -\alpha n_c/N < 0$

Donc, pour obtenir W^c on voit qu'il suffirait d'appliquer ces empreintes *de façon répétée, entre tous les groupes consécutifs de la série* (P_0, P_1, \dots, P_m) . Par exemple, la première étape de cette construction idéale serait :

$$W = W^0 + Z[P_0 \rightarrow P_1] + Z[P_1 \rightarrow P_2] + \dots + Z[P_{m-1} \rightarrow P_m]$$

(comparer la figure 4 à la figure 5), ce qui commencerait à mettre en évidence les connexions de la chaîne en les augmentant une première fois de $\approx +\alpha$. Puis il

faudrait réitérer plusieurs fois cet enchaînement d'additions, et, finalement, grâce à l'accumulation de $\approx +\alpha$, dans les limites de la troncation, on obtiendrait assez rapidement des connexions $P_k \times P_{k+1}$ fortes, tandis que l'accumulation des décréments $-\alpha n/N$ provoquerait l'affaiblissement des autres connexions jusqu'à 0.

Ainsi, dans notre modèle, cette chaîne pourrait être l'aboutissement de l'évolution récurrente $W(t+1) = W(t) + Z[P(t) \rightarrow P(t+1)]$, à condition que celle-ci soit guidée par une propagation d'activités assez particulière qui devrait reproduire la *succession ordonnée* des groupes $\{P_0, P_1, \dots, P_m\}$. En d'autres termes, nous pouvons écrire : $W^c = \lim_{t \rightarrow +\infty} W(t)$, avec une loi de propagation qui vérifierait la propriété suivante :

$$\forall k=0 \dots m-1, P(t) = P_k \Rightarrow P(t+1) = P_{k+1}$$

c'est à dire que les activations devraient se faire exclusivement d'un groupe au groupe suivant dans la série. Il faudrait aussi *plusieurs* enchaînements de ce type pour faire croître les connexions et entretenir leur stabilité, donc on devrait relancer cette propagation en posant $P(t) = P_0$ en des instants t suffisamment fréquents :

$$\exists t_1, t_2, t_3, \dots ; (t_{i+1} - t_i) \leq \Delta t_{\max} \text{ et } P(t_i) = P_0$$

Mais cela ne signifie pas que P_0 doit être réactivé à intervalles réguliers, car après la dernière activation $P(t) = P_m$, le réseau peut rester silencieux pendant une durée indéterminée sans affecter les valeurs des connexions déjà acquises. Finalement, il suffirait simplement que les deux propriétés ci-dessus soient réalisées pour produire la chaîne désirée, W^c , entre les groupes $\{P_0, P_1, \dots, P_m\}$: la première porte sur la régularité de la succession et la deuxième sur sa répétition systématique (mais pas forcément périodique).

Ce processus de décharge consistera donc à réitérer la propagation de ce que nous appellerons une *onde d'activation* sur la chaîne (figure 5). On notera $P^c(t)$ cette loi de propagation, et $x^c(t)$ l'évolution correspondante des états du réseau. Dans ces conditions, on peut reprendre le calcul statistique du §4.3.a en remplaçant le régime aléatoire x^a par le régime régulier x^c . Cette fois-ci, on a :

$$\begin{aligned} \langle Z_{ij}[x^c(t) \rightarrow x^c(t+1)] \rangle = & \text{proba}[x_i(t)=1, x_j(t+1)=1] (+\alpha(1-2n_c/N)) \\ & + \text{proba}[x_i(t)=1, x_j(t+1)=0] (-\alpha n_c/N) \\ & + \text{proba}[x_i(t)=0, x_j(t+1)=1] (-\alpha n_c/N) \end{aligned}$$

Par conséquent, avec $x(t) = x^c(t)$, il faudra distinguer quatre cas suivant l'appartenance ou non de i et j à des paquets consécutifs de la chaîne. Les probabilités ci-dessus sont équivalentes à des *fréquences* d'apparition des événements $i|1 \rightarrow j|1$, $i|1 \rightarrow j|0$ ou $i|0 \rightarrow j|1$. Pour les calculer plus précisément, on supposera que les ondes qui parcourent la chaîne à tour de rôle sont séparées par des intervalles de temps en

moyenne égaux à $m' \geq m$.

(a) Pour $(i,j) \in P_k \times P_{k+1}$, la fréquence moyenne de $i|1 \rightarrow j|1$ est $1/m'$, tandis que $i|1 \rightarrow j|0$ ou $i|0 \rightarrow j|1$ ne se produisent jamais ($i|0 \rightarrow j|0$ est le plus fréquent avec une probabilité de $(m'-1)/m'$, mais il n'a pas de conséquence sur W_{ij}). Donc il est clair que les connexions $P_k \times P_{k+1}$ *augmentent* globalement : elles ne changent pas pendant $m'-1$ itérations, puis elles sont brusquement augmentées de $\approx +\alpha$ à l'instant où la vague des activités les traverse. Par conséquent, leur incrément "moyen" au cours du temps vaut $\approx +\alpha/m'$.

(b) En revanche, les connexions concurrentes subissent surtout des événements ratés dans un sens comme dans l'autre et *diminuent* globalement. Par exemple, sur les connexions $(i,j) \in P_k \times P_{k' \neq k+1}$, les événements $i|1 \rightarrow j|1$ ne se produisent jamais. Par contre, ces connexions sont porteuses d'une fausse transmission de chaque type à chaque fois qu'une onde se propage sur la chaîne : une fois sur m' , elles diminuent de $-\alpha n_c/N$ à cause de $i|1 \rightarrow j|0$, et encore une fois à cause de $i|0 \rightarrow j|1$, ce qui donne un décrement "moyen" de $-2\alpha n_c/Nm'$. Quant aux autres connexions extérieures à la chaîne, c'est à dire $P_k \times (R \setminus R_c)$ (connexions sortant des paquets vers l'extérieur) et $(R \setminus R_c) \times P_k$ (connexions entrant sur les paquets en provenance de l'extérieur), elles sont le siège d'un seul événement raté à chaque propagation ($i|1 \rightarrow j|0$ pour les premières et $i|0 \rightarrow j|1$ pour les deuxièmes).

Finalement, les connexions (a) augmenteront jusqu'à atteindre 1, et les connexions (b) diminueront jusqu'à atteindre 0 : donc la propagation $P^c(t)$ permet de construire rapidement la chaîne W^c .

En résumé, une chaîne ne peut émerger du réseau et s'y maintenir de façon durable que si la croissance de ses connexions s'appuie sur un mode de propagation $P(t)$ *régulier et répétitif*. Autrement dit, la chaîne doit être soutenue par une *forte corrélation* entre les activités des neurones, ce terme faisant référence à la fréquence élevée des événements réussis $i|1 \rightarrow j|1$ par rapport aux autres événements, $i|1 \rightarrow j|0$, $i|0 \rightarrow j|1$, et $i|0 \rightarrow j|0$. En clair, on retrouve ici le postulat hebbien : pour qu'une connexion (i,j) soit renforcée et devienne prédominante, elle doit être le siège de transmissions $i|1 \rightarrow j|1$ *statistiquement plus nombreuses* que dans une situation aléatoire, c'est à dire que l'activation du neurone i doit entraîner presque systématiquement celle du neurone j et, de plus, cet événement doit se produire suffisamment souvent. Donc les conditions forcées de propagation que nous avons définies ici réalisent idéalement ces propriétés, puisque, pour tout $i \in P_k$ et tout $j \in P_{k+1}$, elles posent : $x_i(t) = 1 \Rightarrow x_j(t+1) = 1$, ainsi que $x_i(t') = 1$ en de nombreux instants t' .

Or on verra au §4.4 que ces conditions artificielles peuvent être en réalité produites tout naturellement par la chaîne elle-même, et, que de la sorte, cette chaîne peut assurer son *auto-entretien*. Donc, comme pour l'état aléatoire $x^a(t)$ analysé plus haut, il n'était pas nécessaire que la propagation $x^c(t)$ soit forcée, car la chaîne de connexions est justement le support parfait pour une telle propagation :

les activités sont transmises d'un paquet à l'autre à travers des connexions fortes, créant ainsi des potentiels d'actions qui provoquent automatiquement l'activation des neurones du paquet suivant.

De plus, on verra au §4.5 que ces ondes d'activations (spontanées) ont non seulement un effet conservateur sur la chaîne, mais aussi un effet véritablement "organisateur" : l'activation réitérée du dernier paquet de la chaîne P_m finira par provoquer l'agrégation d'un paquet supplémentaire P_{m+1} , et la chaîne se développera en longueur par empilement de maillons successifs. Ainsi, il sera montré dans les deux paragraphes suivants que la chaîne peut assurer à la fois sa subsistance et sa croissance, moyennant le déclenchement réitéré d'ondes de propagation.

En conclusion, deux phénomènes extrêmes de développement ont été évoqués ici : dans un cas, les activations repassent toujours par les mêmes groupes, et on obtient rapidement une chaîne de connexions fortes (§4.3.b), dans un autre cas, les activations ne repassent jamais par les mêmes groupes, et le réseau reste homogène (§4.3.a). Nous avons donc insisté sur le fait que la propagation des activités était déterminante dans l'organisation du réseau : selon sa "direction", la loi générale d'évolution des connexions, qui consiste à additionner des empreintes de liaison, aura des conséquences radicalement opposées. Tout dépend en effet des sites d'accumulation de ces empreintes : si elles se dispersent sur toutes les parties du réseau, rien n'apparaît, tandis que si elles se concentrent sur les mêmes sous-ensembles, une chaîne apparaît.

Nous avons également donné le point de vue équivalent des connexions individuelles : il est clair que l'évolution de W_{ij} dépend de la fréquence relative des événements réussis de la transmission $i \rightarrow j$, c'est à dire du *degré de corrélation* entre l'activation de i et celle de j . Ainsi, une propagation aléatoire est caractérisée par des corrélations uniformément nulles entre $x_i(t)$ et $x_j(t+1)$, tandis qu'une propagation régulière et répétée est caractérisée par des corrélations positives sur les connexions de la chaîne et négatives sur les connexions concurrentes.

Comme nous l'annoncions, il reste maintenant à compléter l'analyse de ce modèle en faisant interagir la propagation avec la structure de connexions qu'elle trace progressivement sur le réseau.

4.4 Conditions de stabilité des chaînes

Nous avons vu plus haut (§4.2 et §4.3) comment les connexions évoluaient en fonction des activités, à travers un module Z qui ne dépendait que des corrélations entre les deux derniers états de neurones. Nous nous proposons maintenant de reprendre l'étude inverse, c'est à dire l'évolution des activités x en fonction de l'état des connexions. Le paragraphe 4.1 avait traité le cas particulier d'un réseau homogène, et avait montré en quoi le processus aléatoire pouvait se stabiliser autour d'une activité globale moyenne n^* . A présent, on se place dans un état W non-homogène, et l'étude des activations doit tenir compte des changements intervenus dans les connexions, sous forme des accroissements locaux $Z[P \rightarrow P']$ entre divers groupes de neurones P et P' . Les connexions varient conjointement avec les activations, et on verra comment cette double dynamique peut rendre compte de la stabilité des chaînes (§4.4) et de leur construction progressive (§4.5).

Remarquons dès à présent que nous ne pourrions calculer, au mieux, que des probabilités d'activation, et non pas des états particuliers x . C'est pourquoi, dans toute la suite, il sera plus commode d'introduire la notation $q_j(t)$ pour désigner les probabilités correspondant aux variables d'état $x_j(t)$:

$$\forall t, \forall i \in R, \text{proba}[x_j(t)=1] = q_j(t)$$

où l'on a : $q_j(t) = \sigma_T(V_j(t) - \theta_j)$ (cf. §3.0). Ainsi, nous nous contenterons de caractériser le réseau de neurones par son état moyen $q(t)$ faute de pouvoir déterminer son état réel $x(t)$. Cependant, la connaissance de q suffira à mettre en évidence des phénomènes d'organisation du réseau, car ceux-ci sont justement caractérisés par une différenciation progressive des probabilités d'activation (par exemple, les neurones faisant partie d'une chaîne ont dans l'ensemble des probabilités d'activation plus fortes que celle des autres neurones). L'organisation du réseau sera donc révélée par la transition d'un paysage de probabilité q uniformément plat à un paysage contrasté dans lequel certaines valeurs q_j seront proches de 1, et d'autres proches de 0. Bien sûr, il est clair que $q(t)$ n'est pas une caractéristique stable ou lentement variable comme $W(t)$: $q(t)$ ne cesse de changer au cours de la propagation, et c'est donc plutôt une *succession reproductible* de transitions $q(t), q(t+1), q(t+2), \dots$ qui sera caractéristique d'une organisation en chaîne (remarquons que le calcul de l'évolution de $q(t)$ est compliqué par le fait qu'il dépend de façon cruciale des détails de l'état précédent $x(t-1)$, auquel nous n'avons justement pas accès : par conséquent, on ne pourra pas fournir de relation récurrente entre $q(t)$ et $q(t-1)$, mais seulement entre les quantités moyennes $\langle q(t) \rangle$, et ce, à travers le même type d'approximation que celle du §4.1 où l'on avait remplacé les distributions par leur espérance).

Dans ce paragraphe 4.4, nous commencerons par étudier des chaînes *déjà formées*, avant d'aborder leur construction effective au §4.5. Il s'agira

essentiellement d'examiner leur stabilité sous la double dynamique des activations neuronales et des modulations synaptiques. On a vu au §4.3 que les activités devaient repasser par les mêmes paquets, dans un ordre déterminé, pour assurer la robustesse des connexions qui vont de l'un à l'autre : cette propagation spéciale a été appelée *onde d'activation*. On verra ici comment une onde d'activation peut être naturellement guidée par la focalisation en chaîne.

Tout d'abord, au §4.4.a, nous donnerons une définition moins stricte d'une onde à l'aide des taux d'activations dans chaque paquet, le but étant de calculer la probabilité d'activation des neurones d'un groupe P_{k+1} induites par le groupe précédent P_k à travers les connexions fortes $P_k \rightarrow P_{k+1}$. L'étude de cette transmission élémentaire sera faite au §4.4.b. Puis on passera *par récurrence* à la chaîne tout entière au §4.4.c, en réitérant m fois cette transmission, ce qui nous amènera à discuter les conditions de convergence des taux d'activité. En résumé : pour qu'une onde d'activation soit canalisée sur une chaîne du premier au dernier paquet, sans dégénérer vers l'extinction ou vers l'explosion, les paramètres devront obéir à certaines relations (seuils des neurones, largeur de la chaîne, valeur des connexions fortes, et température). De plus, cette propagation dépendra de façon précise des conditions de déclenchement sur le premier paquet P_0 . Ainsi, au §4.4.d, on tentera de réaliser la réactivation de P_0 à travers le réseau extérieur à la chaîne, qui sera supposé homogène avec des processus de décharge aléatoires (comme au §4.1) : on montrera alors que ces conditions ne sont pas suffisantes pour déclencher la propagation d'une onde, et on conclura à la nécessité d'un germe de régime imposé, sous forme d'une pulsation générique dans P_0 . Cette pulsation et ses conséquences importantes sur l'auto-organisation des chaînes seront examinées en détail au §4.5.

4.4.a Entretien d'une chaîne par ondes d'activation

Nous reprenons ici la définition formelle d'une chaîne donnée au §4.3 : des groupes de neurones disjoints $\{P_0, P_1, \dots, P_m\}$ (de même taille $n_c \ll N$, pour simplifier) sont reliés en série par des connexions plus fortes que la moyenne du réseau. Cette organisation, notée W^c , sera en général caractérisée par les valeurs suivantes :

$$\forall k = 0, \dots, m-1,$$

- $i \in P_k$ et $j \in P_{k+1} \Rightarrow W_{ij}^c \approx w_+$
- $i \in P_k$ et $j \notin P_{k+1} \Rightarrow W_{ij}^c \approx w_-$
- $i \notin P_k$ et $j \in P_{k+1} \Rightarrow W_{ij}^c \approx w_-$

où w_+ représente la valeur moyenne des connexions internes à la chaîne, et w_- celle des connexions qui relient la chaîne R_c au reste du réseau $R \setminus R_c$ (les valeurs exactes fluctuent légèrement autour de w_+ et w_-). w_+ est proche de 1 et w_- est proche de 0, mais, plus précisément, leurs valeurs seront déterminées par la contrainte de

conservation des sommes. Ainsi, en supposant que $S_i(W^c)$ et $S_j(W^c)$ sont approximativement égales à s_0 pour tout $i \in P_k$ et tout $j \in P_{k+1}$, on obtient l'équation suivante :

$$n_c w_+ + (N - n_c) w_- = s_0$$

(car la chaîne est complètement connectée). Sachant d'autre part que w_+ et w_- sont comprises dans l'intervalle $[0, 1]$, il faut envisager deux cas :

- si $n_c < s_0$ (chaîne étroite), alors : $w_+ \approx 1$ et $w_- \approx (s_0 - n_c)/N$
- si $n_c \geq s_0$ (chaîne large), alors : $w_+ \approx s_0/n_c < 1$ et $w_- \approx 0$

En particulier, pour $n_c = s_0$, on aura : $w_+ \approx 1$ et $w_- \approx 0$. Par ailleurs, on peut décrire la chaîne W^c comme étant une somme de *maillons* que nous noterons $Y[P_k \rightarrow P_{k+1}]$, c'est à dire de structures locales de connexions reliant les paquets consécutifs :

$$\forall i, j \in R_c, W_{ij}^c = Y_{ij}[P_0 \rightarrow P_1] + Y_{ij}[P_1 \rightarrow P_2] + \dots + Y_{ij}[P_{m-1} \rightarrow P_m]$$

Autrement dit : $i \in P_k$ ou $j \in P_{k+1} \Leftrightarrow Y_{ij}[P_k \rightarrow P_{k+1}] = W_{ij}^c$ (et $Y_{ij} = 0$ sinon). L'aspect de Y est donc le même que celui de Z (connexions fortes d'un paquet vers l'autre et connexions faibles en direction et en provenance de l'extérieur; cf. figure 4), et chaque Y sera le résultat de l'accumulation de plusieurs Z . Remarquons que la définition de la chaîne est limitée aux connexions dont l'origine ou l'extrémité sont dans R_c . Le reste du réseau, c'est à dire les (W_{ij}) tels que $i \notin R_c$ et $j \notin R_c$, sont a priori quelconques (dans le respect de la conservation des sommes), c'est à dire non nécessairement homogènes, et pourront éventuellement contenir d'autres chaînes.

Etant donné cette chaîne W^c , nous nous intéressons à présent aux événements de propagation stable qu'elle peut porter : il s'agit des ondes d'activation présentées au §4.3.b, et que nous allons redéfinir ici de façon moins stricte. On rappelle que ces ondes sont très importantes pour l'entretien de la chaîne, car chaque passage d'une vague d'activités sur un maillon a pour effet de renforcer les connexions internes et d'affaiblir davantage les connexions externes concurrentes, grâce à l'addition d'une nouvelle empreinte Z . La répétition de ce processus fait donc tendre w_+ et w_- vers les valeurs limites, qui sont dictées à la fois par la conservation des sommes à s_0 et par la saturation individuelle à 1, comme on l'a calculé plus haut.

La définition idéale d'une onde que nous avons donnée au 4.3.b était :

$$\forall k = 0, \dots, m, P(t_0 + k) = P_k$$

où t_0 marque l'instant de départ sur la racine P_0 . Cependant, il est possible d'en donner une définition moins stricte : au lieu d'exiger une coïncidence parfaite entre $P(t_0 + k)$ et P_k , on peut se contenter d'un certain taux de recouvrement entre ces deux

ensembles, c'est à dire $P(t_0+k) \approx P_k$ (onde "bruitée"). Pour pouvoir étudier les conditions de stabilité d'une onde bruitée, il sera alors plus commode de la définir à l'aide des *taux d'activation locaux* des paquets de la chaîne. Ainsi, on peut imaginer que le passage d'une onde est détecté par des "électrodes" fictives qui mesurent l'activité locale collective $n_{P_k}(t)$ à l'intérieur de chaque groupe P_k (en réalité ce groupe n'est pas localisé dans l'espace : il est constitué de neurones répartis aléatoirement dans le réseau). La définition d'une propagation bruitée peut alors s'écrire :

$$\forall k = 0, \dots, m, n_{P_k}(t_0+k) \approx n_c$$

Toutefois, cette condition n'est pas suffisante, car elle est satisfaite en particulier par des états $x(t)$ qui n'ont rien à voir avec une propagation stable (par exemple, la catastrophe $n(t)=N$ vérifie trivialement cette condition). Donc, pour caractériser véritablement une onde, il manque les conditions complémentaires suivantes :

$$\forall \tau \neq 0, \forall k = 0, \dots, m, n_{P_k}(t_0+\tau+k) = 0$$

Il s'agit de la propagation de ce que nous appellerons des "ondes de non-activation", en des instants décalés d'un délai τ , antérieurement ou postérieurement par rapport à t_0 .

Ainsi, une onde d'activation solitaire le long d'une chaîne de longueur $m+1$, n'est pas simplement définie par l'activation de P_k à l'instant t_0+k , mais également par la "non-activation" de ce même P_k à tous les instants $t_0+k+\tau$, pour $\tau \neq 0$ variant de $-m$ à $+m$.

4.4.b Transmission des activations sur un maillon

Après avoir rappelé les conditions de propagation qui permettent d'entretenir une chaîne (sous une forme "bruitée"), nous abordons maintenant l'étude inverse, pour montrer comment les activations sont naturellement transmises sur la chaîne à travers la suite des maillons, par la simple règle du neurone à seuil. Pour cela, on commencera par calculer au §4.4.b les probabilités de transmission sur un maillon quelconque $Y[P \rightarrow P']$ puis on utilisera cette étape élémentaire pour constituer une loi de récurrence sur toute la chaîne au §4.4.c.

L'empreinte de liaison, Z , qui vient sans cesse s'ajouter aux connexions en suivant la trace des activités, est l'élément de base de toutes les constructions pouvant apparaître sur le réseau. Or, réciproquement, elle joue aussi un rôle actif en infléchissant le cours de la propagation : en effet, si les connexions qui vont d'un groupe P vers un groupe P' sont supérieures à la moyenne des connexions qui vont de P à l'extérieur de P' (comme c'est le cas après une ou plusieurs applications de $Z[P \rightarrow P']$), et si de plus une fraction suffisante de P est réactivée à un instant t , alors les neurones de P' auront une probabilité plus forte que le reste du réseau d'être

activés à l'instant $t+1$. On voit donc que le module Z sert de *guide* à la propagation : son intervention entre P et P' orientera ultérieurement la décision d'activation en faveur de P' , lorsque les neurones de P seront à nouveau actifs ensemble. Les connexions $P \rightarrow P'$ forment donc une "déviation" vers P' , efficace seulement à partir de P . En un mot : là où les connexions ont été renforcées, les activités auront tendance à repasser.

Nous allons maintenant formaliser cette constatation, en calculant le paysage des probabilités $q(t+1)$, à partir d'un état d'activité $x(t)$ donné : de façon générale, on s'intéressera à l'action d'un maillon $Y[P \rightarrow P']$ (équivalent à plusieurs empreintes Z), caractérisé, comme on le rappelle, par des connexions $P \times P'$ fortes :

- $i \in P$ et $j \in P' \Rightarrow Y_{ij}[P \rightarrow P'] \approx w_+$
- $i \in P$ et $j \notin P' \Rightarrow Y_{ij}[P \rightarrow P'] \approx w_-$
- $i \notin P$ et $j \in P' \Rightarrow Y_{ij}[P \rightarrow P'] \approx w_-$

Etant donné cette configuration locale de connexions, la question est alors la suivante : soit un état d'activités $x(t)$ a priori quelconque, caractérisé par $n(t)$ neurones actifs, dont $n_P(t)$ se trouvent dans P et $n_{R \setminus P}(t)$ en dehors de P , quel sera le nombre moyen de neurones actifs dans P' à l'instant $t+1$, $\langle n_{P'}(t+1) \rangle$?

On rappelle que la probabilité d'activation d'un neurone j à l'instant $t+1$ vaut :

$$q_j(t+1) = \frac{1}{1 + e^{-(V_j(t+1) - \theta_j)/T}}$$

Or pour un neurone j appartenant au paquet P' , le potentiel $V_j(t+1)$ est donné par :

$$\begin{aligned} V_j(t+1) &\approx \sum_{i \in P} w_+ x_i(t) + \sum_{i \notin P} w_- x_i(t) \\ &\approx w_+ n_P(t) + w_- n_{R \setminus P}(t) \end{aligned}$$

C'est donc le même potentiel pour tous les $j \in P'$. Par conséquent, le nombre moyen de neurones actifs dans le paquet P' à l'instant $t+1$ vaut n fois la probabilité q_j (on suppose que P et P' sont de même taille n) :

$$\langle n_{P'}(t+1) \rangle \approx \frac{n}{1 + \exp\left(-\frac{1}{T} (w_+ n_P(t) + w_- n_{R \setminus P}(t) - \Theta_0)\right)}$$

Ainsi, $n_{P'}$ ne dépend pas seulement de n_P à l'instant précédent, mais aussi de $n_{R \setminus P}$ par l'intermédiaire des connexions entrantes w_- . Donc, par récurrence, pour traiter la transmission sur le maillon suivant $P' \times P''$, il faudrait calculer non seulement $n_{P'}(t+1)$, mais aussi $n_{R \setminus P'}(t+1)$. Cependant, le problème est qu'on ne connaît pas la force des connexions extérieures $(R \setminus P) \times (R \setminus P')$, ce qui nous empêche alors de

calculer $n_{R \setminus P}(t+1)$. On voit donc qu'il existe une relation croisée qui fait dépendre $n_P(t+1)$ et $n_{R \setminus P}(t+1)$ de $n_P(t)$ et $n_{R \setminus P}(t)$, c'est à dire une influence non seulement interne à $P \times P'$, mais aussi sortant de P vers l'extérieur, et entrant de l'extérieur vers P' .

Donc, pour simplifier, nous supposons que ce croisement est "coupé", c'est à dire que $w_- = 0$. Dans ce cas, la chaîne ne communique plus avec l'extérieur, et la relation précédente est restreinte à une simple récurrence entre $n_P(t+1)$ et $n_P(t)$:

$$\langle n_P(t+1) \rangle \approx \frac{n}{1 + \exp\left(-\frac{1}{T}(w_+ n_P(t) - \Theta_0)\right)}$$

La propagation des activités à l'intérieur la chaîne sera donc indifférente au comportement du reste du réseau. On verra au §4.5 et dans les simulations du chapitre 5 que cette situation est parfaitement plausible : en effet, les chaînes moins larges que s_0 (qui communiquent donc avec l'extérieur à travers des connexions w_- non-nulles) ne sont en réalité pas stables car les paquets ont toujours tendance à "s'étaler" le plus possible, c'est à dire à recruter de nouveaux neurones jusqu'à atteindre un nombre limite imposé par la compétition et la saturation (ce nombre étant égal à la taille du paquet précédent si celui-ci est plus grand que s_0 , ou bien à s_0 s'il est plus petit). Ainsi, sans anticiper davantage, retenons simplement que n_c est en général supérieur à s_0 et donc que $w_- \approx 0$ et $w_+ \approx s_0/n_c$ (cas des chaînes larges).

4.4.c Conditions de propagation d'une onde

En utilisant les probabilités de transmission qui viennent d'être calculées sur un maillon, nous allons maintenant voir comment la chaîne entière peut porter une onde d'activation telle qu'elle a été définie au §4.4.a : cette structure a en effet la propriété de pouvoir soutenir la propagation d'une onde avec grande robustesse, et nous allons en définir plus précisément les conditions de stabilité.

Il est apparu plus haut que pour caractériser une onde d'activation nous étions conduits à nous intéresser à *plusieurs suites* de taux d'activités à la fois, paramétrées par τ : $\{n_{P_k}(t_0 + \tau + k)/n_c\}_{k=0, \dots, m}$ avec $\tau = -m, \dots, +m$. Ces suites doivent toutes valoir uniformément 0, sauf la suite correspondant à $\tau = 0$ qui doit au contraire stationner à 1. Pour cela, on verra que les paramètres s_0 , Θ_0 et T doivent appartenir à un certain domaine de valeurs, en dehors duquel la propagation dégénère soit vers un état d'"embrasement" (tous les neurones de la chaîne sont actifs, i.e. toutes les suites convergent vers 1), soit vers le silence total (toutes les suites convergent vers 0). Dans ces conditions, il existe alors un *seuil de déclenchement* de l'onde, correspondant à un certain pourcentage de neurones activés dans P_0 à l'instant t_0 . Ce seuil sera déterminé à partir d'une relation de récurrence approximative entre les nombres moyens de neurones actifs dans P_k et P_{k+1} .

Prenons l'une quelconque des suites ci-dessus, caractérisée par un retard τ , et tentons d'en déterminer la loi de récurrence. Pour plus de commodité, on notera désormais $t = t_0 + \tau$ l'instant de départ en P_0 : on cherche donc à relier les nombres $n_{P_k}(t+k)$ et $n_{P_{k+1}}(t+k+1)$. Soit $x(t)$ l'état du réseau à l'instant t , a priori quelconque, comprenant en tout $n(t)$ neurones actifs, parmi lesquels $n_{P_0}(t)$ appartiennent à P_0 . A l'instant suivant, l'espérance du nombre de neurones actifs dans le paquet suivant, P_1 , est alors donnée par :

$$\langle n_{P_1}(t+1) \rangle = \frac{n_c}{1 + \exp\left(-\frac{1}{T}(w_+ n_{P_0}(t) - \Theta_0)\right)}$$

Ainsi, connaissant le nombre exact $n_{P_0}(t)$, on peut calculer le nombre moyen $\langle n_{P_1}(t+1) \rangle$. Bien sûr, il en va de même dans les itérations ultérieures pour les paquets suivants : pour tout $k=0, \dots, m$, $\langle n_{P_{k+1}}(t+k+1) \rangle$ dépendra de $n_{P_k}(t+k)$ à travers la même formule. Cependant, il ne s'agit que d'une espérance conditionnelle, connaissant l'état précédent à $t+k$: donc, si l'on voulait obtenir l'espérance globale de $n_{P_{k+1}}(t+k+1)$ intégrée sur tous les paquets antérieurs depuis t jusqu'à $t+k$, il faudrait d'abord calculer (par récurrence) la distribution de probabilité de $n_{P_k}(t+k)$, puis une combinaison de la sigmoïde par ces distributions.

Nous sommes donc confrontés au même problème que celui du §4.1, et nous adopterons la même approximation : ainsi, on supposera que le nombre actif moyen du paquet P_k , $\langle n_{P_k}(t+k) \rangle$, est suffisamment représentatif de la distribution de $n_{P_k}(t+k)$, pour servir seul au calcul du nombre moyen suivant, $\langle n_{P_{k+1}}(t+k+1) \rangle$. De la sorte, on obtient une relation de récurrence approximative entre quantités moyennes :

$$\langle n_{P_{k+1}}(t+k+1) \rangle = \frac{n_c}{1 + \exp\left(-\frac{1}{T}(w_+ \langle n_{P_k}(t+k) \rangle - \Theta_0)\right)}$$

Cette loi de récurrence est donc tout à fait similaire à celle que nous avons trouvée au §4.1 : il s'agit d'ailleurs de la même fonction σ , appliquée ici sur une échelle " n_c " au lieu d'une échelle " N ". En effet, comme on l'a montré au §4.4.a, les connexions fortes de la chaîne valent : $w_+ \approx s_0/n_c$; or, à l'époque où le réseau était homogène, les valeurs initiales de connexions w_0 vérifiaient : $w_0 \approx s_0/N$; donc les premières se déduisent des deuxièmes par : $w_+ \approx w_0(N/n_c)$. Ceci veut bien dire que la chaîne est une focalisation de la connectivité s_0 , qui était initialement répartie entre N neurones, et qui relie à présent des paquets contenant seulement n_c neurones.

Par conséquent, si on note n_t^* le point fixe de cette relation (il s'agit ici d'une récurrence sur k , paramétrée par $t=t_0+\tau$), celui-ci vérifiera :

$$n_t^*(N/n_c) = \sigma(n_t^*(N/n_c)).$$

Or on rappelle que les intentions concernant la convergence des différentes suites sont doubles : d'une part, nous voulons une convergence vers $n_t^* = n_c$ pour $t = t_0$,

mais d'autre part une convergence vers $n_t^* = 0$ pour $t \neq t_0$. Ici, il faut donc qu'une même relation de récurrence puisse servir à deux comportements opposés : ceci ne pourra être réalisé qu'à travers des conditions initiales différentes $n_{p0}(t)$, et à condition que la relation en question offre la possibilité d'une telle alternative. Plus précisément, les paramètres de σ seront tels qu'il existe un *seuil de déclenchement* pour $n_{p0}(t)$, au-delà duquel la suite $n_{pk}(t+k)$ croît vers n_c , et en-deçà duquel elle décroît vers 0. L'existence de ce seuil est à nouveau l'objet d'une étude graphique (figure 6), qui doit aussi être en accord avec les conclusions du §4.1 sur la stabilité de l'activité totale du réseau initial.

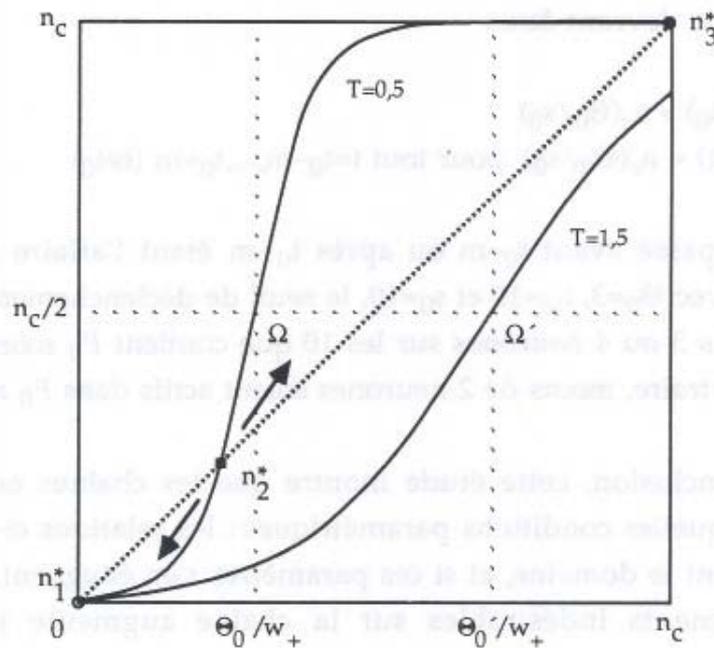


Figure 6 : Condition de stabilité des ondes d'activation sur la chaîne. Le principe de l'étude graphique est le même que dans la figure 1 (§4.1), à la différence que la norme du carré est $n_c=10$ au lieu de $N=100$: donc pour $s_0=10$, les connexions valent $w_+=1$ au lieu de $w_0=0,1$. Il est important de remarquer que ce schéma est directement lié à celui du §4.1 par l'application d'une homothétie de rapport N/n_c , et ceci veut dire que les conditions paramétriques sont *communes* aux deux processus (elles doivent à la fois réaliser une faible activation dans le réseau initialement homogène, et permettre la propagation d'une onde sur le réseau structuré en chaîne). On a représenté ici deux sigmoïdes de pentes différentes, $T=0,5$ et $T=1,5$, et de centres de symétrie Ω et Ω' également différents, qui correspondent respectivement à deux valeurs de seuil $\Theta_0=3$ et $\Theta_0'=7$. Du point de vue du réseau homogène (étude du §4.1), ces courbes sont toutes les deux satisfaisantes car elles assurent une convergence vers $n=0$. Cependant, du point de vue de la chaîne, seule la première offre le point de bifurcation n_2^* nécessaire à la propagation d'une onde, c'est à dire un "seuil de déclenchement" : selon que la valeur initiale $k=0$ de la suite $(t+k)$ se trouvera au-dessus ou en-dessous du point instable n_2^* , celle-ci va croître vers n_c ou au contraire décroître vers 0 (nous voulons le premier cas pour $t=t_0$, et le deuxième pour $t \neq t_0$).

Finalement, l'étude graphique montre que la sigmoïde doit impérativement couper la bissectrice en 3 points. Donc, sachant que les coordonnées du centre de la sigmoïde sont : $\Omega (\Theta_0/w_+; n_c/2)$, et sachant que la pente en ce point vaut : $n_c w_+ / 4T$, nous choisirons dans toute la suite : $\Theta_0/w_+ \leq n_c/2$ (Ω à gauche de la bissectrice), et il faudra par conséquent que la température vérifie : $n_c w_+ / 4T > (n_c/2) / (\Theta_0/w_+)$ (pente suffisamment raide pour couper la diagonale en n_2^*). Ces relations se résument à :

$$\Theta_0 \leq s_0/2 \quad \text{et} \quad T < \Theta_0/2$$

c'est à dire, par exemple : $s_0=10$, $\Theta_0=3$ et $T=0,5$ (remarquons qu'il serait également possible de choisir $\Theta_0 \geq s_0/2$ (point Ω du côté droit), auquel cas il faudrait alors : $T < (s_0 - \Theta_0)/2$).

Dans ces conditions, on obtient un seuil de déclenchement pour $n_{P_0}(t)$ proche de Θ_0/w_+ (voir la position de n_2^* sur la figure 6) qui vaut également : $n_c(\Theta_0/s_0)$ (nous sommes toujours dans le cas où $n_c \geq s_0$). Par conséquent, les $2m+1$ conditions initiales des suites permettant le déclenchement d'une onde de propagation stable sur la chaîne devront être :

- $n_{P_0}(t_0) > n_c(\Theta_0/s_0)$
- $n_{P_0}(t) < n_c(\Theta_0/s_0)$ pour tout $t=t_0-m, \dots, t_0+m$ ($t \neq t_0$)

(ce qui se passe avant t_0-m ou après t_0+m étant l'affaire d'une autre onde). Par exemple, avec $\Theta_0=3$, $n_c=10$ et $s_0=10$, le seuil de déclenchement vaut 3 : il faudra donc qu'au moins 3 ou 4 neurones sur les 10 que contient P_0 soient actifs à l'instant t_0 , et que, au contraire, moins de 2 neurones soient actifs dans P_0 aux autres instants.

En conclusion, cette étude montre que les chaînes ne sont pas stables dans n'importe quelles conditions paramétriques : les relations ci-dessus entre Θ_0 , s_0 et T en délimitent le domaine, et si ces paramètres s'en éloignent, le risque de provoquer des événements indésirables sur la chaîne augmente significativement, donc également le risque de détruire la chaîne par des corrélations d'activités parasites ("embrasement" à cause de seuils trop forts, "silence" à cause de seuil trop faibles, ou bruit uniforme à cause d'une température trop forte). Elle montre aussi que la notion de *seuil de déclenchement* sur la racine P_0 est inhérente à la définition d'une propagation d'onde : sur la même chaîne, il est indispensable de pouvoir propager aussi bien une onde d'activation qu'une onde de "non-activation".

4.4.d Conditions de déclenchement d'une onde

La question qui vient est alors celle-ci : Quel processus de déclenchement permet de réaliser les $2m+1$ conditions initiales définies ci-dessus ? Il s'agit donc de voir comment la racine de la chaîne peut être réactivée ponctuellement à un instant t_0 , et rester par ailleurs silencieuse à tous les instants qui précèdent ou qui suivent t_0 dans un intervalle $[t_0-m, t_0+m]$, autrement dit : comment P_0 peut spontanément "pulser" des potentiels dans P_1 avec une fréquence inférieure à $1/m$ (pour ne pas provoquer de "double-propagation") mais non-nulle.

Les neurones de P_0 sont naturellement activés par l'intermédiaire de leurs afférences : ces connexions n'appartiennent pas à la chaîne (voir la figure 5), et sont a

priori quelconques. Nous ferons alors l'hypothèse que la chaîne est plongée dans un *réseau homogène* : ainsi, les connexions extérieures, qui incluent en particulier les connexions entrantes de P_0 et les connexions sortantes de P_m , auront à peu près même valeur. La définition de W^c est alors complétée de la façon suivante :

$$\bullet (i \notin R_c \text{ ou } i \in P_m) \text{ et } (j \in R_c \text{ ou } j \in P_0) \Rightarrow W_{ij}^c \approx w_0'$$

où w_0' est déterminée par la conservation des sommes sur le réseau extérieur, c'est à dire, sachant que la chaîne contient en tout $N_c = (m+1)n_c$ neurones : $w_0' = s_0 / (N - N_c)$. Cette valeur est donc légèrement supérieure à $w_0 = s_0 / N$, mais on peut raisonnablement supposer que $w_0' \approx w_0$, dans la mesure où N est grand devant N_c .

Ainsi, on constate au passage que la formation d'une chaîne provoque l'augmentation globale des connexions du reste du réseau : ce phénomène, qui peut sembler paradoxal au premier abord, est dû au fait que les neurones de la chaîne privent les neurones extérieurs à la chaîne d'une certaine quantité de connexions afférentes (à cause de la focalisation exclusive vers les neurones du paquet voisin), et que, par conséquent, les neurones extérieurs doivent chercher "entre eux" les afférences qui leur manquent. Donc, comme ces neurones sont moins nombreux qu'avant ($N - N_c$ au lieu de N), le niveau moyen du réseau homogène qu'ils forment est nécessairement plus élevé. En réalité, on voit qu'il s'agit d'un effet de *taille finie*, car pour N très grand devant N_c , w_0' est égal à w_0 : autrement dit, un réseau très étendu ne serait pas perturbé par la structuration d'une chaîne, et pourrait "absorber" les variations globales qu'elle provoque, grâce à une répartition diffuse sur un très grand nombre de connexions. Nous supposons en effet que la chaîne est encore loin d'avoir épuisé toutes les réserves du réseau.

L'activité de l'environnement extérieur est donc quasi-indépendant de celle de la chaîne, puisque les liaisons qui vont de l'un à l'autre sont coupées ($w_- \approx 0$), et les seuls points de communication sont les paquets P_0 et P_m . On peut imaginer que la chaîne forme un "pont" au-dessus de la masse homogène du réseau : ce pont repose sur deux bases P_0 et P_m , qui constituent ses seules attaches avec le réseau. Entre ces extrémités, la succession des tronçons $P_0 \rightarrow P_1 \rightarrow \dots \rightarrow P_{m-1} \rightarrow P_m$ n'offre aucune dérivation vers l'extérieur, puisque $w_- \approx 0$ (comme si le pont était délimité par une "bordure" infranchissable; les connexions latérales de la figure 5 sont nulles). La circulation des activités est donc confinée à l'intérieur des limites de ce pont, et se fait toujours dans le même sens : une fois engagées sur P_0 , les activités sont canalisées à travers toutes les étapes P_k , et ne peuvent ressortir qu'en P_m à partir duquel elles se répandent dans le réseau extérieur.

Etant donné cet état de connexions W^c et un certain état d'activités initial $x(t)$, quel sera alors le comportement des activations ? Comme la propagation des activités dans la chaîne dépend entièrement des conditions initiales sur P_0 , on est ramené à l'étude du réseau aléatoire homogène $R \setminus R_c$ et à ses conséquences sur P_0 (on supposera que ce réseau était totalement silencieux pendant une période de longueur au moins égale à m précédant t : ceci garantit qu'aucune onde n'est en train de se propager sur la chaîne à cet instant). Au bout de quelques itérations, et

sous les bonnes conditions paramétriques, l'espérance du nombre de neurones actifs dans le réseau aléatoire augmente un peu et se stabilise à une valeur n^* petite devant N (n^* est donné approximativement par l'étude du point fixe faite au §4.1).

Peut-on alors espérer déclencher une onde d'activation à partir de P_0 sous ces conditions ? A chaque itération, environ n^* neurones parmi les $N - N_c$ que contient le réseau homogène sont activés au hasard : autrement dit, la probabilité de décharge d'un neurone de $R \setminus R_c$ est : $q \approx n^*/(N - N_c) \approx n^*/N$. Par conséquent, il en est de même pour les neurones du premier paquet P_0 , puisque leurs afférences homogènes proviennent directement de ce réseau, et la probabilité pour que n neurones soient actifs simultanément dans P_0 (qui en contient n_c) est alors une loi binomiale :

$$\text{proba}[n_{P_0}=n] = C(n_c, n) q^n (1-q)^{(n_c-n)}$$

dont l'espérance vaut :

$$\langle n_{P_0} \rangle = n_c q \approx n_c (n^*/N)$$

Or, dans la mesure où n^* est petit devant N , $\langle n_{P_0} \rangle$ sera également petit devant n_c . Ainsi, grâce à cette faible activité moyenne du réseau extérieur, les $2m$ conditions initiales qui exigent le silence de P_0 (de $t=t_0-m$ à $t=t_0+m$, sauf $t=t_0$) seront tout naturellement réalisées.

En revanche, il sera beaucoup plus difficile pour n_{P_0} de satisfaire la condition principale de déclenchement en $t=t_0$, c'est à dire de franchir le seuil $n_c(\Theta_0/s_0)$, car avec ce régime aléatoire, la probabilité pour qu'au moins n neurones de P_0 soient actifs simultanément est très faible. Autrement dit, la réactivation de la racine de la chaîne sous ces conditions constitue un événement improbable ou très rare. Pour que n_{P_0} ait des chances significatives de dépasser occasionnellement le seuil n , il faudrait que les paramètres soient modifiés pour produire une activité moyenne n^* sensiblement plus forte, mais dans ce cas, on met alors en péril la stabilité de la chaîne : si n_{P_0} dépassait n trop souvent, il déclencherait des ondes à intervalles trop rapprochés, ce qui provoquerait l'augmentation de connexions parasites entre paquets non-consécutifs. Ce serait également un danger pour le réseau extérieur, car n'oublions pas que les connexions continuent à varier sous l'effet des corrélations fortuites : tant que n^* est faible, l'équilibre entre activations aléatoires et réseau homogène est stable (conjugaison du 4.1 et du 4.3.a) mais si n^* prend de l'ampleur, la situation peut dégénérer vers une augmentation incontrôlée des connexions, et donc une explosion.

Finalement, on voit que le déclenchement d'une onde d'activation à travers un processus de décharge aléatoire sur un réseau extérieur homogène est une entreprise sinon impossible du moins très délicate : il faut parier sur les fluctuations de ce processus, c'est à dire sur les probabilités pour que l'agitation désordonnée de $R \setminus R_c$ se concentre accidentellement sur P_0 et produise une vague qui traverse le "pont" de P_0 à P_m puis se dilue à nouveau dans $R \setminus R_c$. On s'expose alors à de grands

risques : soit un embrasement général qui détruirait rapidement la chaîne, soit un silence prolongé qui l'effacerait progressivement en raison de l'augmentation aléatoire de connexions concurrentes.

Tout le problème est donc de réaliser à la fois l'exigence d'un taux élevé occasionnel sur P_0 (n/n_c) et l'exigence d'un taux faible permanent sur $R \setminus R_c$ (n^*/N), ce qui n'est pas possible si le niveau d'activité moyen de P_0 est le même que celui de $R \setminus R_c$. En conclusion, les afférences de P_0 ne doivent pas appartenir à un réseau homogène : il faut envisager au départ l'introduction d'une inhomogénéité dans le réseau, sous la forme de connexions afférentes plus fortes que la moyenne, provenant d'un "neurone-source" ou "neurone-horloge" auquel l'activation des neurones de P_0 sera subordonnée. Nous développerons cet nouvel aspect au §4.5, et nous montrerons que, grâce à ce "germe" initial, la décharge simultanée et répétée d'un groupe particulier de neurones sera non seulement nécessaire à l'entretien de la chaîne, mais aussi et surtout à sa construction par *croissance agrégative* ("cristallisation").

4.5 Auto-organisation des chaînes par pulsations locales

Le paragraphe précédent a montré que l'entretien d'une chaîne déjà formée se faisait naturellement à travers la propagation d'ondes d'activation sur cette chaîne. Ainsi, pour repasser régulièrement par la même succession de paquets $P_0 \rightarrow P_1 \rightarrow P_2 \rightarrow \dots \rightarrow P_{m-1} \rightarrow P_m$, il suffit de tirer profit des propriétés de cette structure de connexions : une focalisation en chaîne de ce type *canalise* les activités et crée précisément l'onde d'activation qui permet de la renforcer. On a montré aussi qu'une onde ne pouvait pas apparaître dans n'importe quel contexte paramétrique : il faut que les seuils des neurones soient proches de $s_0/2$, et que la température reste relativement basse. De plus, il est apparu que pour provoquer les conditions initiales de déclenchement d'une onde, il n'était pas possible de parier sur des processus de décharge aléatoires. Ces conditions nécessitent une concentration exceptionnelle d'activités sur la racine de la chaîne P_0 , à un instant particulier t_0 , et, au contraire, une relative inactivité sur ce même paquet avant et après t_0 . Un réseau homogène ne permet donc pas de réaliser cette simultanéité ponctuelle des décharges de façon fiable, et il faut alors envisager des connexions localement plus fortes que la moyenne.

Dans ce paragraphe, on verra comment la propagation des ondes permet non seulement de stabiliser une chaîne déjà constituée mais aussi de la faire *croître* en longueur : la décharge simultanée des neurones du dernier paquet P_m , qui est l'aboutissement de la cascade des décharges ayant traversé les paquets précédents,

provoquera la focalisation des afférences de P_m vers un nouveau paquet P_{m+1} , en même temps que la synchronisation des décharges des neurones de P_{m+1} . Autrement dit, un nouveau groupe de neurones synchrones sera agrégé tandis qu'un maillon supplémentaire de connexions $P_m \rightarrow P_{m+1}$ viendra s'ajouter en bout de chaîne : ce double phénomène constitue le mécanisme fondamental de l'auto-organisation des chaînes.

Pour l'analyser, on commencera par revenir à la source de la chaîne, au §4.5.a, en localisant un groupe spécial de neurones ayant la faculté de décharger simultanément, et ce de façon répétée : ce groupe sera appelé *groupe pulsatile*. Puis, au §4.5.b et au §4.5.c, on examinera les conséquences des deux premières pulsations de ce groupe sur le réseau initialement homogène, en mettant en évidence un début de différenciation dans le paysage des probabilités (en résumé : les neurones qui se sont trouvés actifs juste après une pulsation auront plus de chances d'être réactivés à la pulsation suivante). La discussion du §4.5.d montrera alors comment l'insistance du groupe pulsatile finit par sélectionner certains neurones du réseau, grâce au renforcement mutuel des corrélations post-pulsatiles et de la focalisation. Ces neurones forment alors un nouveau groupe pulsatile et finiront par provoquer à leur tour l'agrégation d'un paquet par la focalisation de leurs efférences, et ainsi de suite : ceci constitue donc le mécanisme récurrent de croissance de la chaîne, qui sera finalement commenté au §4.5.e.

L'objet de ce paragraphe 4.5 est donc de montrer comment un "germe" local (inhomogénéité de connexions) est capable de provoquer la "croissance cristalline" d'une structure de chaîne.

4.5.a Neurones pulsatiles

Dans le réseau initialement homogène $W = W^0$ (tel que $W_{ij}^0 = w_0$), nous supposons à présent qu'il existe un neurone i_0 possédant des efférences beaucoup plus fortes que la moyenne, c'est à dire exceptionnellement supérieures à 1 (figure 7) Ce neurone i_0 innerve un petit groupe de neurones Q de telle sorte qu'il en provoque systématiquement la décharge lorsqu'il est activé (pour cela, il suffit que les connexions $i_0 \rightarrow Q$ soit plus grandes que les seuils); d'autre part, il reçoit normalement des afférences en provenance des autres neurones. Ainsi, les neurones de Q auront la faculté de décharger toujours *simultanément*, après chaque activation fortuite de i_0 : nous les appellerons "*neurones pulsatiles*".

On voit donc que la singularité introduite par i_0 a pour conséquence de constituer précisément le groupe P_0 qui sera nécessaire au déclenchement des ondes d'activation (ici, $P_0 = Q$) : la propriété la plus importante des neurones pulsatiles est donc d'être *synchrones*, et de plus, cette décharge simultanée devra se produire relativement fréquemment pour être en mesure d'entretenir une chaîne, et également pour provoquer l'agrégation d'un maillon supplémentaire à l'extrémité de celle-ci. Un modèle simple de pulsation est donné par le processus suivant :

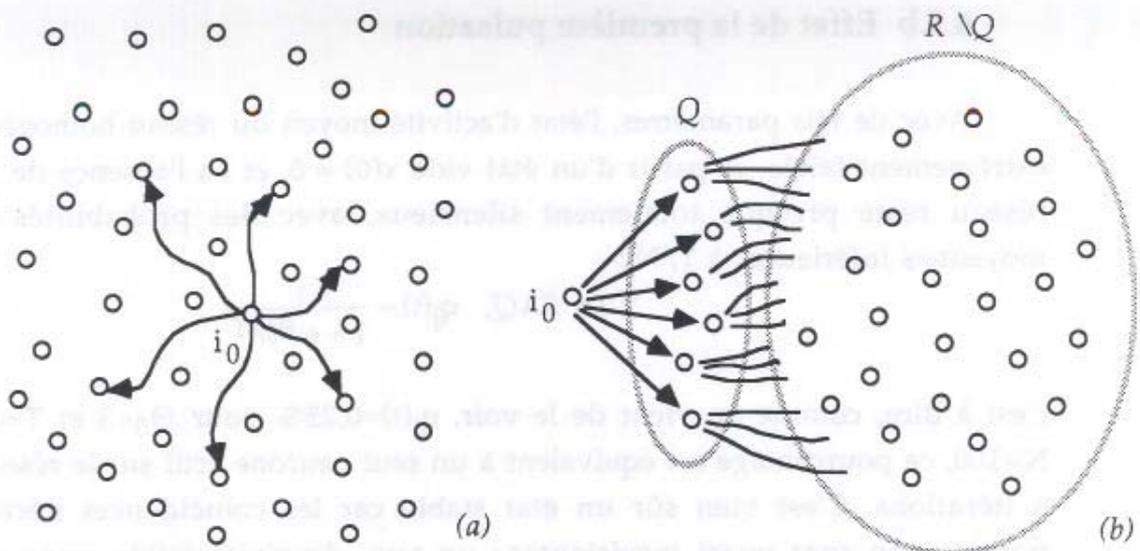


Figure 7 : Inhomogénéité locale de connexions. (a) Un neurone i_0 innerve quelques neurones à travers des efférences particulièrement fortes, ce qui provoque l'activation simultanée de ceux-ci à chaque fois que i_0 est actif. (b) Vue équivalente du réseau, où l'on a mis en évidence le groupe Q des neurones innervé par i_0 ("groupe pulsatile"). Ce groupe innerve à son tour le reste du réseau $R \setminus Q$.

$$\forall i \in Q, \forall t, x_i(t) = x_{i_0}(t-1), \text{ avec : } \text{proba}[x_{i_0}(t-1)=1] = 1/r.$$

ce qui veut dire que les neurones de Q reproduiront systématiquement l'état d'activité du neurone i_0 une unité de temps plus tard, et que cette loi probabiliste est uniforme, caractérisée par une fréquence moyenne de décharge $1/r$ ($r \geq 1$ représente la période moyenne entre deux décharges). Remarquons que cette pulsation n'a pas besoin d'être parfaitement périodique et que, bien au contraire, une pulsation très régulière peut s'avérer nuisible à long terme pour la stabilité de la chaîne, à cause des fortes corrélations croisées qu'elle pourrait provoquer (une pulsation régulière entraînera des propagations multiples sur la chaîne dès que $r \leq m$). Cette loi permet donc une certaine variation de l'intervalle de séparation des pulsations autour de la valeur moyenne r .

Munis des neurones pulsatiles, nous reprenons ici les événements depuis l'instant initial t_0 : cette singularité mise à part, le réseau se trouve dans un état de connexions à peu près homogène, $W_{ij}^0 \approx w_0$ avec $w_0 \approx s_0/N$, et les processus de décharge sont rares et aléatoires. Nous allons maintenant montrer comment une chaîne peut se construire par l'action du groupe Q : elle prendra racine sur ce groupe et se développera en longueur grâce à aux pulsations répétées de Q . En notant n_0 le nombre de neurones pulsatiles, on choisira $n_0 \geq s_0$: il faudra alors s'attendre à une chaîne de largeur $n_c = n_0$, comme on le montrera dans la discussion du §4.5.d. Les paramètres seront ici : $s_0=10$ et $n_0=10$, donc, d'après les conclusions du §4.4.c, on peut choisir par exemple : $\Theta_0=3$ et $T=0,5$ afin d'assurer la stabilité de la future chaîne (le seuil de déclenchement est de $3/10$).

4.5.b Effet de la première pulsation

Avec de tels paramètres, l'état d'activité moyen du réseau homogène initial est extrêmement faible. A partir d'un état vide $x(0) = 0$, et en l'absence de pulsation, le réseau reste presque totalement silencieux, avec des probabilités d'activation moyennes inférieures à $1/N$:

$$\forall j \in R \setminus Q, \quad q_j(t) \approx \frac{1}{1 + e^{\Theta_0/T}}$$

c'est à dire, comme on vient de le voir, $q_j(t) \approx 0,25\%$ pour $\Theta_0=3$ et $T=0,5$: donc si $N=100$, ce pourcentage est équivalent à un seul neurone actif sur le réseau toutes les 4 itérations. C'est bien sûr un état stable car les coïncidences nécessaires à la potentiation sont quasi inexistantes : un taux d'activité faible comme celui-ci ne modifie pas W_{ij} , qui reste homogène et perpétue cette inactivité générale.

C'est dans de telles conditions que les n_0 neurones du groupe pulsatile Q déchargent une première fois à un instant t : $\forall i \in Q, x_i(t) = 1$. Cet événement change alors brusquement l'ensemble des probabilités d'activations de l'instant $t+1$: celles-ci sont toujours uniformes, mais elles ont un peu augmenté :

$$\forall j \in R \setminus Q, \quad q_j(t+1) \approx \frac{1}{1 + e^{-(w_0 n_0 - \Theta_0)/T}}$$

Avec $w_0=0,1$ et $n_0=10$, elles valent : $q_j(t+1) \approx 1,8\%$. Donc pour $N=100$, on peut espérer à l'instant $t+1$ l'activation de $n=1, 2$ ou 3 neurones dont nous noterons P le groupe.

Aussitôt, les corrélations fortuites apparues entre les neurones pulsatiles Q les neurones de P vont provoquer l'augmentation de toutes les connexions qui vont de Q à P , soit : $W(t+1) = W^0 + Z[Q \rightarrow P]$, avec :

- $i \in Q$ et $j \in P \Rightarrow Z_{ij}[Q \rightarrow P] = +\alpha(1 - (n_0 + n)/N)$
- $i \in Q$ et $j \notin P \Rightarrow Z_{ij}[Q \rightarrow P] = -\alpha n/N$
- $i \notin Q$ et $j \in P \Rightarrow Z_{ij}[Q \rightarrow P] = -\alpha n_0/N$
- $i \notin Q$ et $j \notin P \Rightarrow Z_{ij}[Q \rightarrow P] = 0$

Pour plus de commodité dans la suite des calculs, on ne retiendra que la partie principale de ces valeurs, c'est à dire : $+\alpha$ entre Q et P , et 0 partout ailleurs, avec par exemple, $\alpha=0,1$ (on négligera aussi le terme correctif ϵ , quasiment nul dans ces conditions; par ailleurs, la troncation n'aura pas à intervenir sur des valeurs de connexions encore faibles comme celles-ci).

L'état $x(t+1)$ est alors insuffisant pour provoquer à son tour l'activation d'autres neurones à l'instant $t+2$: il diffuse des potentiels de l'ordre de nw_0 , qui sont trop faibles pour maintenir des probabilités supérieures à $1/N$ (par exemple, en supposant dans le meilleur des cas que $n=2$, on retombe vers une probabilité moyenne de l'ordre de $q_j(t+2) \approx 0,37\%$). Ainsi, pour le moment, les conséquences de

la décharge simultanée de n_0 neurones à l'instant t se résument à l'activation de $n < n_0$ neurones à $t+1$, et le réseau est à nouveau silencieux à partir de $t+2$.

Il faut donc attendre la prochaine décharge de Q pour voir apparaître une nouvelle phase d'activité sur le réseau. On note t' l'instant de la deuxième pulsation : celle-ci se produira environ une période r après la première, c'est à dire $t' \approx t+r$ (avec par exemple $r=20$). Comme c'était le cas à $t+1$, le niveau moyen des probabilités est alors nettement plus grand à $t'+1$.

Cependant, il existe maintenant une différence entre les neurones de P et les autres neurones : en effet, la probabilité pour que les premiers soient réactivés est désormais *plus forte* que la moyenne du réseau, puisqu'ils reçoivent un potentiel plus important grâce à l'augmentation précédente des connexions $Q \rightarrow P$. Ces connexions valent $\approx w_0 + \alpha = 0,2$ depuis l'itération $[t, t+1]$, tandis que les connexions $Q \rightarrow R \setminus P$ sont encore à $\approx w_0 = 0,1$ (en réalité toutes ces connexions ont sans doute légèrement diminué à cause d'événements fortuits du type $i|0 \rightarrow j|1$ ou $i|1 \rightarrow j|0$ survenus entre-temps). Ainsi, le paysage de probabilités à $t'+1$ n'est déjà plus uniforme, et favorise P : les connexions $Q \rightarrow P$ forment une ébauche très faible de "maillon", avec des tailles de paquets inégales ($n_0=10, n=2$). Donc, pour tout $j \in R \setminus Q$ on a :

$$\begin{aligned} \bullet j \in P &\Rightarrow q_j(t'+1) = \frac{1}{1 + e^{-((w_0 + \alpha) n_0 - \Theta_0)/T}} \\ \bullet j \notin P &\Rightarrow q_j(t'+1) = \frac{1}{1 + e^{-(w_0 n_0 - \Theta_0)/T}} \end{aligned}$$

Avec les paramètres habituels, on trouve alors une probabilité $q_j(t'+1)=12\%$ dans le premier cas, contre $q_j(t'+1)=1,8\%$ seulement dans le deuxième cas. On voit donc qu'un simple renforcement de $+\alpha$ des connexions $Q \rightarrow P$ peut modifier notablement les chances de réactivation de P , consécutivement à l'activation de Q .

Ainsi, cette toute première étape montre bien que les débuts de focalisation (empreintes Z) ont un pouvoir "directionnel" sur la propagation, mais que celui-ci ne peut s'exprimer que lorsque le groupe de départ est suffisamment réactivé. Un seul accroissement des connexions $Q \rightarrow P$ peut donc perturber le paysage des probabilités en faveur des neurones P à condition que, auparavant, les neurones de Q aient été réactivés *ensemble*. En d'autres termes, les neurones de Q doivent *coopérer* pour maximiser les chances de réactiver ceux qu'ils ont contactés une première fois, donc de renforcer davantage ces contacts, et ainsi de suite. Tout l'intérêt des neurones pulsatiles réside précisément dans la simultanéité de leur activation, qui est une forme d'action concertée. Bien sûr, il est encore trop tôt ici pour en apprécier pleinement les effets, et il faudra attendre encore plusieurs pulsations avant de voir émerger un premier paquet de neurones liés à Q par des connexions fortes qui garantissent leur réactivation post-pulsatile systématique. Cependant, on voit déjà ici le principe du mécanisme d'auto-organisation.

4.5.c Effet de la deuxième pulsation

Nous en sommes donc au point où les quelques neurones de P ont théoriquement plus de chances que le reste du réseau d'être activés à l'instant $t'+1$, c'est à dire après la 2^{ème} pulsation. Mais les valeurs numériques proposées en exemple montrent que ces probabilités sont encore assez faibles : même si la majorité des neurones n'a que 2 chances sur 100, on voit que ceux de P ont en tous cas moins d'une chance sur 8. Donc, en réalité, il est vraisemblable que les neurones de P ne seront *pas* réactivés après cette 2^{ème} pulsation, mais que celle-ci provoquera en revanche l'activation de $n'=2$ autres neurones de $R \setminus Q$ formant un petit ensemble P' . Dans ce cas, les nouvelles corrélations apparues cette fois-ci entre Q et P' vont provoquer l'augmentation de $+\alpha$ des connexions $Q \rightarrow P'$: celles-ci vaudront maintenant $\approx w_0 + \alpha$, tandis que les connexions $Q \rightarrow P$ qui valaient $\approx w_0 + \alpha$ rediminueront légèrement à cause de la compétition entre efférences de Q , mais de façon tout à fait négligeable. Par conséquent, si on note t'' l'instant de la 3^{ème} décharge du groupe pulsatile Q , on aura :

$$\begin{aligned} \bullet j \in P \cup P' &\Rightarrow q_j(t''+1) = \frac{1}{1 + e^{-((w_0 + \alpha)n_0 - \theta_0)/T}} \\ \bullet j \notin P \cup P' &\Rightarrow q_j(t''+1) = \frac{1}{1 + e^{-(w_0 n_0 - \theta_0)/T}} \end{aligned}$$

Ainsi, on constate que les neurones de P et de P' se trouvent maintenant à égalité de chances pour leur éventuelle réactivation après la 3^{ème} pulsation (comme précédemment, cette probabilité est de 12%, contre 1,8% pour le reste du réseau, à la différence qu'elle concerne maintenant environ 4 neurones au lieu de 2).

Cependant, il se pourrait également que la 2^{ème} pulsation provoque un événement moins probable que celui qui vient d'être envisagé, événement qui incluerait la réactivation d'au moins un neurone de P (avec $n=2$, et $q(t'+1) \approx 12\%$, cet événement a 1 chance sur 5 de se produire). Or, dans ce cas de figure, le futur paysage de probabilités à $t''+1$ n'aura pas la même forme que dans la situation ci-dessus. En effet, si un neurone j de P est à nouveau actif juste après la 3^{ème} pulsation, alors les connexions $Q \rightarrow j$ augmenteront une nouvelle fois de $+\alpha$, et, en conséquence, les chances de réactivation après la 3^{ème} pulsation seront nettement plus grandes que celles qu'il aurait eu normalement (cf. calcul ci-dessous). Donc, il faut ici distinguer *trois* groupes de neurones, et non plus seulement deux, pour décrire le futur paysage de probabilités à $t''+1$ consécutif à la 3^{ème} pulsation : (a) ceux qui étaient actifs à la fois après la 1^{ère} et après la 2^{ème} pulsation; (b) ceux qui étaient actifs soit après la 1^{ère} soit après la 2^{ème} pulsation; (c) ceux qui n'étaient actifs ni après la 1^{ère} ni après la 2^{ème} pulsation. Ces trois cas correspondent respectivement à :

$$\begin{aligned} \bullet j \in P \cap P' &\Rightarrow q_j(t''+1) \approx \frac{1}{1 + e^{-((w_0+2\alpha)n_0 - \Theta_0)/T}} \\ \bullet j \in P \Delta P' &\Rightarrow q_j(t''+1) \approx \frac{1}{1 + e^{-((w_0+\alpha)n_0 - \Theta_0)/T}} \\ \bullet j \in P \cup P' &\Rightarrow q_j(t''+1) \approx \frac{1}{1 + e^{-((w_0)n_0 - \Theta_0)/T}} \end{aligned}$$

(où $P \Delta P' = (P \cup P') \setminus (P \cap P')$ représente la différence symétrique de P et P'), et les valeurs numériques sont : 50%, 12% et 1,8% (figure 8).

On voit donc que, à l'instant $t'+1$ suivant la 2^{ème} pulsation, il est important de savoir si certains neurones de P feront partie des neurones réactivés ou non pour calculer les futures probabilités à $t'+1$. Cet exemple montre que, pour continuer la description des phénomènes d'auto-organisation, on ne peut plus se contenter de calculer le nombre moyen de neurones actifs après chaque pulsation (comme c'était le cas après la 1^{ère}), mais que l'on a aussi besoin de connaître avec précision les détails de l'historique des activités $x(t)$. Nous arrêtons ici cet examen détaillé, et nous entamons à présent une description globale des phénomènes d'organisation causés par les pulsations.

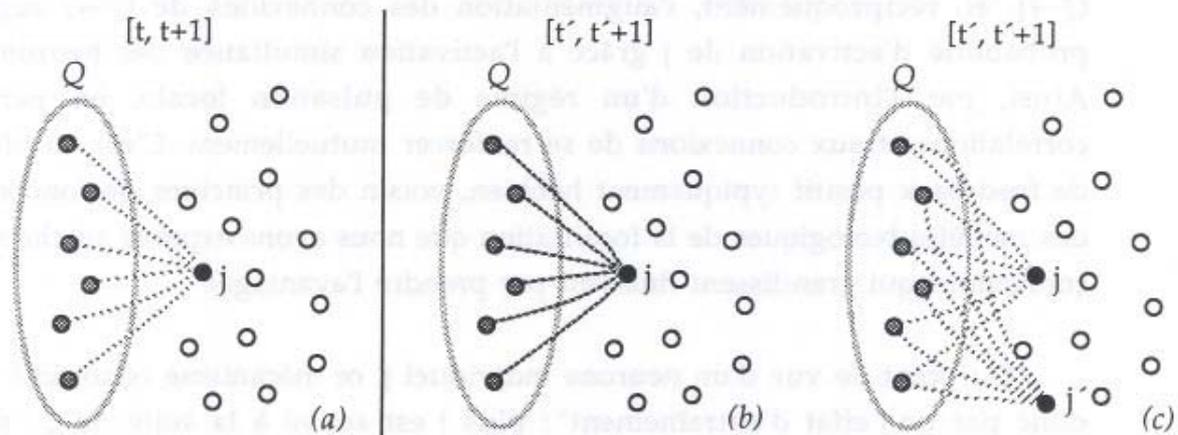


Figure 8 : Illustration de l'effet des deux premières pulsations. (a) Le groupe pulsatile Q est actif à l'instant t et provoque l'activation d'un neurone j à l'instant $t+1$: toutes les connexions $Q \rightarrow j$ augmentent alors de $+0,1$. La probabilité pour j d'être réactivé à la pulsation suivante est désormais plus importante que celle des autres neurones (elle passe de 1,8% à 12%). La pulsation suivante a lieu à l'instant t' , et peut être suivie (b) soit par l'activation de j , (c) soit par l'activation d'un autre neurone j' . Dans le premier cas, j renforce davantage ses afférences en provenance de Q et s'assure une probabilité de future réactivation de 50%. Dans le deuxième cas, j et j' sont à égalité avec 12%.

4.5.d Pulsations suivantes et genèse du premier maillon

De la discussion qui précède, il ressort clairement que les neurones commencent à se différencier d'une pulsation à l'autre en fonction du nombre de fois où ils ont été actifs consécutivement aux pulsations précédentes, c'est à dire du degré de leur *activation "post-pulsatile" passée*. Ainsi, à la 3^{ème} pulsation de Q , la probabilité de réactivation d'un neurone j est nettement différente selon que j a déjà été activé 2 fois après chacune des 2 premières pulsations, ou une seule fois après

une seule d'entre elles, ou encore aucune fois (le calcul ci-dessus a donné respectivement : 50%, 12% et 1,8%). Donc, de façon générale, l'avenir de j dépendra du nombre d'événements de type $Q|1 \rightarrow j|1$ qui ont eu lieu sur $Q \rightarrow j$, c'est à dire de *l'historique de ses corrélations avec Q* : c'est la règle de potentiation hebbienne qui mémorise cet historique dans les connexions $Q \rightarrow j$. A chaque fois que j est réactivé après Q , ces connexions augmentent de $\approx +\alpha$ (pondéré par un coefficient légèrement inférieur à 1, dû à la compétition), tandis que les événements $Q|1 \rightarrow j|0$ et $Q|0 \rightarrow j|1$ les font légèrement décroître.

Or, comme on l'a montré plus haut, ce ne sont pas seulement les connexions $Q \rightarrow j$ qui varient en fonction de l'historique des corrélations, mais c'est aussi la *probabilité d'activation* de j lui-même, grâce à la *décharge simultanée* des neurones pulsatiles à travers des connexions plus fortes. En effet, il ne suffit pas que les connexions $Q \rightarrow j$ soient renforcées pour assurer une plus grande probabilité d'activation à j , mais encore faut-il que cette focalisation soit utilisée par une *coopération* d'activations à la racine : donc pour maximiser les chances d'avoir $x_j(t+1)=1$, il est important que $x_i(t)$ soit égal à 1 là où $W_{ij}(t)$ est plus fort.

En résumé : l'activation de j consécutivement à Q augmente les connexions $Q \rightarrow j$, et, réciproquement, l'augmentation des connexions de $Q \rightarrow j$ augmente la probabilité d'activation de j grâce à l'activation simultanée des neurones de Q . Ainsi, par l'introduction d'un régime de pulsation locale, on permet aux corrélations et aux connexions de se renforcer mutuellement. C'est un phénomène de feed-back positif typiquement hebbien, voisin des principes de fonctionnement des modèles biologiques de la focalisation que nous avons exposés au chapitre 2 (les connexions qui grandissent finissent par prendre l'avantage).

Du point de vue d'un neurone individuel j , ce mécanisme rétro-actif se traduit donc par un "effet d'entraînement" : plus j est activé à la suite de Q , plus il est encouragé à recommencer. Nous avons vu que sa probabilité q_j était une fonction sigmoïdale d'un multiple de $+\alpha$, et on avait obtenu, dans l'ordre, les valeurs suivantes : 1,8% pour $+0.\alpha$, 12% pour $+1.\alpha$, et 50% pour $+2.\alpha$. Donc, en poursuivant ce raisonnement, on obtiendra 88% pour $+3.\alpha$: par conséquent, s'il se trouve que j a été activé trois fois, après chacune des 3 premières pulsations, il est désormais presque certain que j sera à nouveau actif après la 4^{ème} pulsation. Ceci entraînera alors une augmentation encore plus grande de ses connexions, donc de sa probabilité : avec un ajout supplémentaire de $+\alpha$, on atteint alors $+4.\alpha$ et q_j passe à $\approx 98\%$. On peut donc considérer que j est désormais définitivement lié à Q : les pulsations suivantes entraînent systématiquement l'activation de j , et quelques unes d'entre elles suffisent à faire grandir les connexions $Q \rightarrow j$ jusqu'à la valeur limite $w_+ \approx s_0/n_0$. La configuration locale de connexions qui en résulte est alors extrêmement robuste, et seuls des accidents hautement improbables tels que la non-activation répétée de j seraient susceptibles de l'altérer. Désormais, ces connexions garantissent à j un potentiel post-pulsatile très fort, donc son activation

systematique après Q , donc se garantissent à elles-mêmes un auto-renforcement permanent.

Bien sûr, rien n'indique qu'un neurone j qui a déjà 3 activations post-pulsatiles à son compte, et qui a donc environ 50% de chances d'en avoir une quatrième, sera réactivé justement après la 4^{ème} pulsation : il se peut très bien que j ne se décide qu'après la 5^{ème} ou la 6^{ème} pulsation ou même plus tard. Ce neurone dispose en effet d'un "crédit" de connexions $Q \rightarrow j$ (donc de potentiel post-pulsatile), qui lui donne un certain avantage sur les autres, et dont il peut encore profiter longtemps au cours des itérations ultérieures. Il en est d'ailleurs de même pour des neurones moins engagés : le fait que j ait vécu 2 activations post-pulsatiles ne signifie pas qu'elles se soient forcément produites après la 1^{ère} et la 2^{ème} décharge de Q . L'augmentation de $+2.\alpha$ qui en est résultée a très bien pu intervenir plus tardivement (par exemple : après la 7^{ème} et après la 12^{ème} pulsation).

Cependant, de tels neurones ne doivent pas non plus trop attendre, car chaque transmission ratée $Q | 1 \rightarrow j | 0$, due à des absences d'activations post-pulsatiles, fait légèrement décroître leurs connexions en provenance de Q , et fait reculer leurs chances de réactivation future. Nous avons jusqu'à présent négligé ces petits décrets dans les calculs précédents, car ils intervenaient trop peu pour être réellement nuisibles : cependant, ils pourraient sensiblement amoindrir les connexions s'ils devaient s'appliquer trop souvent, sans aucune contre-partie d'augmentation. Donc si j restait silencieux trop longtemps, la lente diminution de ses afférences $Q \rightarrow j$ pourrait finir par ramener celles-ci vers la valeur initiale w_0 , en raison des activations post-pulsatiles d'autres neurones, car l'avance de ceux-ci se fait nécessairement *au détriment* de j à cause de la compétition (voir la figure 14 au chapitre 5). Dans ce cas, j aurait perdu tout le crédit qu'il avait gagné auparavant, et, donc, toutes ses chances de réactivation. Par contre, il suffit que ce neurone j soit réactivé une fois après une pulsation pour gagner encore en connexions entrantes, et donc en probabilité.

On voit donc qu'un neurone déjà contacté par le groupe pulsatile a le choix entre deux comportements : soit une inactivité durable qui fait progressivement décroître ses chances de réactivation, et qui tend par conséquent à prolonger cette inactivité; soit au contraire quelques activations supplémentaires consécutives aux pulsations qui le poussent très rapidement vers un état d'activité post-pulsatile systématique se renforçant lui-même. Un tel neurone est à la frontière d'un équilibre instable : trop d'hésitations le font redescendre, tandis que quelques décisions le propulsent vers le haut. Cette constatation rejoint la remarque suivante : comme nous l'avons dit, l'augmentation des connexions $Q \rightarrow j$ fait augmenter la probabilité de réactivation post-pulsatile de j , mais ceci n'implique pas sa réactivation effective. Or seule cette dernière peut faire augmenter davantage les connexions : il existe donc une "barrière de décision" à franchir (qui se compte en nombre d'activations post-pulsatiles) pour être durablement lié au groupe pulsatile.

Dans les exemples numériques que nous avons donnés, le dilemme est particulièrement net pour un neurone j n'ayant vécu que 2 activations post-pulsatiles : en effet, se trouvant alors à $q_j \approx 12\%$, il a environ 1 chance sur 10 de faire croître ses connexions $Q \rightarrow j$ de la quantité $+\alpha$, alors qu'il a 9 chances sur 10 de les faire décroître d'une quantité $-\alpha n/N$, dont l'ordre de grandeur est de quelques centièmes de α . Ainsi, l'espérance de gain fournie par les chances d'augmentation est grosso-modo contrebalancée par l'espérance de perte issue des risques de diminution (avec toutefois un léger avantage pour l'augmentation). En revanche, à $q_j \approx 50\%$, la situation est fortement déséquilibrée en faveur d'une augmentation, et, en réalité, le neurone j peut être considéré comme déjà lié à Q : il faudrait en effet de très nombreux décrets pour le faire reculer, alors qu'un seul incrément (qui a également 1 chance sur 2 d'être appliqué, tout comme les décrets) suffit à le porter vers $q_j \approx 88\%$. Remarquons que le risque de recul est néanmoins important si d'autres neurones ont réussi *avant j*, à franchir le seuil critique de probabilité et à se lier fortement avec Q : ces derniers sont alors indétronables, et, à travers la compétition, freinent notablement les chances des neurones encore "indécis" comme j . Mais si ce n'est pas encore le cas, on voit que la barrière de décision que j doit franchir peut être évaluée en pondérant les éventuels gains et pertes par leurs probabilités respectives, et en considérant le cas où cette combinaison est nulle (ce calcul est bien sûr approximatif).

A travers ces quelques remarques, on discerne mieux le mécanisme de genèse du premier maillon qui prend racine sur le groupe pulsatile Q : ce maillon est un ensemble de connexions fortes qui s'établissent progressivement entre Q et un premier groupe de neurones caractérisés par une corrélation post-pulsatile forte, les connexions soutenant les corrélations et réciproquement. Ce groupe constitue le *premier paquet de la chaîne*, et il sera noté P_1 . Sa définition est donc la suivante : P_1 est l'ensemble final des neurones dont les afférences $Q \rightarrow j$ sont suffisamment élevées pour garantir leur activation post-pulsatile systématique. Ceci suppose par ailleurs que Q contienne un *nombre minimal* de neurones, n_0 : pour fournir des potentiels forts, ce nombre doit être en effet supérieur au seuil Θ_0 . Il existe donc une "*masse critique*" de Q provoquant l'agrégation du groupe P_1 .

Après quelques fluctuations (dont nous reparlons plus bas), la composition du groupe P_1 finit par se stabiliser et atteindre une taille n_1 . Il est clair que n_1 est limité par la compétition, et on peut montrer qu'il sera égal à n_0 . En effet, les connexions fortes du maillon $Q \rightarrow P_1$ convergent vers une *même* valeur d'équilibre w_+ , puisque les neurones de P_1 sont tous activés simultanément après chaque décharge de Q , et qu'en conséquence leurs afférences augmentent toutes à la fois. Donc la compétition de ces connexions pour maintenir une somme égale à s_0 provoquera leur nivelage à w_+ , dont la valeur est donnée par les conditions *entrantes sur P_1* : comme on le sait, on aura $w_+ \approx s_0/n_0$, dans le cas où $n_0 \geq s_0$. Or c'est justement w_+ qui déterminera la taille limite de P_1 , et ce à travers les conditions *sortantes de Q* : en effet, ces sommes sortantes valent $n_1 w_+$ et doivent également stationner autour de s_0 . Finalement, on obtient : $n_1 \approx s_0/w_+$, c'est à dire $n_1 \approx n_0$. Remarquons que si on avait un nombre plus petit de neurones pulsatiles, $n_0 < s_0$, les connexions w_+ pourraient alors converger vers la valeur maximale 1, et cette fois-ci, n_1 serait limité par s_0 lui-même. En

conclusion, on a dans tous les cas : $n_1 = \max \{n_0, s_0\}$ (figure 9).

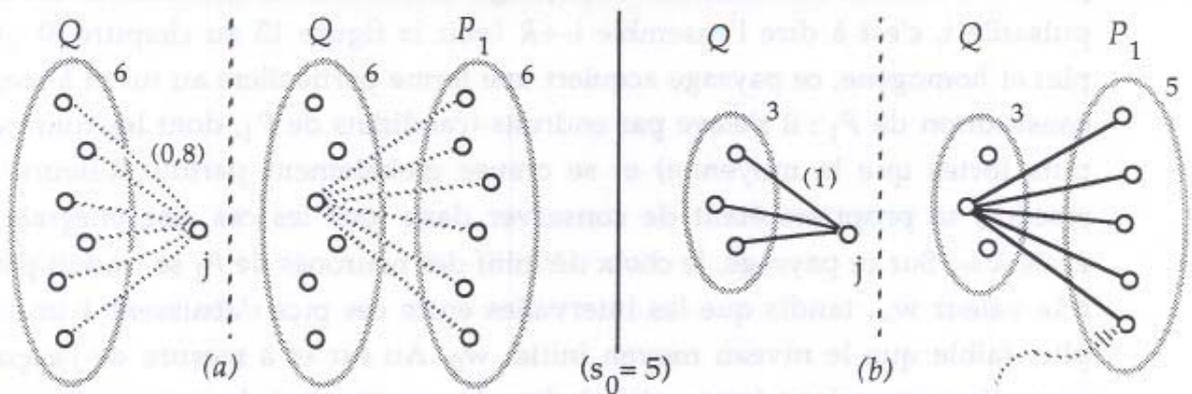


Figure 9 : Taille-limite n_1 du premier paquet post-pulsatile. Pour cette illustration, la valeur normative des sommes est $s_0=5$. (a) Si $n_0=6>s_0$ alors w_+ ne pourra pas excéder $5/6=0,8$ du point de vue des sommes entrantes de j : par conséquent, du point de vue des sommes sortantes de Q , n_1 pourra atteindre $5/0,8=6$, c'est à dire $n_1=n_0$. (b) Si $n_0=3<s_0$ alors w_+ pourra atteindre 1, donc n_1 sera égal à $5/1=5$, c'est à dire $n_1=s_0$ (remarquons que, dans ce cas, les sommes entrantes de j ne sont pas saturées, et ne le seront jamais).

Le groupe P_1 n'est évidemment pas formé en un bloc, mais plutôt par étapes successives, aussi bien dans le temps que dans le recrutement des neurones. Au départ, certains neurones j se distinguent de la masse du réseau par des afférences $Q \rightarrow j$ un peu plus élevées que la moyenne, grâce au fait qu'ils se sont trouvés actifs après une ou deux pulsations : on dira que ces neurones sont "candidats" à l'intégration du premier paquet de la chaîne. Puis le hasard fait que certains d'entre eux gagnent encore en connexions afférentes, donc en potentiel, et, par l'effet d'"avalanche ascendante" dont nous avons parlé plus haut, se hissent rapidement vers un état de liaison robuste avec Q . Du statut de candidats, ces neurones ont donc accompli la transition vers celui de neurones "recrutés" : le hasard les a sélectionnés pour qu'ils fassent désormais partie intégrante du premier paquet de la chaîne, P_1 . Si aucun événement spécial ne vient perturber ce développement (événement comportant de fortes corrélations parasites, dues à des pulsations étrangères à Q par exemple), il est alors impossible de déloger un neurone recruté de sa position extrêmement stable, et ce sont de tels neurones qui formeront le premier groupe P_1 .

Cependant, ils n'y accéderont pas tous en même temps : certains s'installeront avant d'autres, et les premiers arrivés feront alors obstacle aux autres candidats à travers la compétition. En effet, comme on l'a montré plus haut, la taille de P_1 est limitée par le fait que les connexions fortes des neurones recrutés (entretenues par leur activation post-pulsatile systématique) monopolisent une partie des sommes sortantes de Q , et restreignent le nombre de neurones qu'il reste à sélectionner, donc les chances de transition de candidat à recruté.

Ainsi, en raison de la contrainte de conservation des sommes, la quantité totale d'afférences, s_0 , qui est initialement répartie de façon diffuse sur toutes les connexions du réseau, doit ensuite être partagée par les quelques connexions fortes

qui émergent de Q vers le premier paquet : c'est la focalisation. Cette organisation peut être examinée à travers le "paysage" des connexions sortantes d'un neurone pulsatile i , c'est à dire l'ensemble $i \rightarrow R$ (voir la figure 13 au chapitre 5). Au départ plat et homogène, ce paysage acquiert une forme particulière au fur et à mesure de la constitution de P_1 : il s'élève par endroits (candidats de P_1 , dont les connexions sont plus fortes que la moyenne) et se creuse globalement partout ailleurs (reste du réseau), sa propriété étant de conserver dans tous les cas une *intégrale constante* égale à s_0 . Sur ce paysage, le choix définitif des neurones de P_1 se traduit par des pics à la valeur w_+ , tandis que les intervalles entre ces pics s'abaissent à un niveau w_- plus faible que le niveau moyen initial w_0 . Au fur et à mesure de l'apparition de nouvelles connexions fortes, c'est à dire du recrutement de nouveaux neurones, w_- diminue et la compétition devient de plus en plus serrée entre les neurones candidats restants : seul le hasard fera basculer la décision vers les uns ou vers les autres. Finalement, l'intégration du dernier neurone dans P_1 achève de puiser dans les réserves sortantes disponibles des neurones de Q (donc $w_- \approx 0$), et l'accès est définitivement fermé aux autres candidats dont les connexions retombent rapidement à 0.

En résumé, l'agrégation du premier paquet est le résultat d'une lutte entre neurones initialement indistincts pour devenir candidats, puis d'une lutte entre candidats pour être recrutés. Par le hasard de leurs activations, les neurones commencent à se lier au paquet pulsatile (ils deviennent candidats), puis il leur faut franchir un seuil critique de réactivation pour finir l'ascension vers l'appartenance définitive à P_1 (ils sont recrutés). Il est intéressant de suivre l'évolution du groupe des candidats, c'est à dire l'historique de tous les neurones qui ont été liés de près ou de loin à Q parce qu'ils se sont trouvés actifs au moins une fois après l'une de ses pulsations. Cette évolution est une longue série de transformations, qui comporte de nombreux arrivées de candidats et de nombreuses expulsions (on verra dans les simulations qu'elle peut s'étendre sur plusieurs dizaines de pulsations, avec les conditions paramétriques standard) : elle débute avec les premières activations post-pulsatiles (instants $t+1$, $t'+1$, ... du §4.5.b), et s'achève avec la clôture définitive du paquet P_1 évoquée ci-dessus, c'est à dire l'intégration du n_1 -ème neurone.

On peut la diviser en deux grandes périodes : une période d'"expansion", et une période d'"épuration". Au départ, le groupe des candidats a tendance à s'étaler, car de nombreux neurones ont été activés une ou deux fois après le groupe pulsatile mais n'ont jamais confirmé leur engagement par une troisième réactivation. Or, le groupe pulsatile ne peut pas se permettre d'innervé un trop grand nombre de candidats indécis, en raison de la quantité limitée de ses efférences, et c'est pourquoi cette expansion homogène doit nécessairement s'arrêter. Entre-temps, certains candidats ont déjà franchi le seuil critique et ont été recrutés : ils attirent alors vers eux la focalisation des efférences de Q , au détriment des autres candidats, qui subissent un recul général. Donc, tant que le paquet pulsatile a encore des efférences

disponibles, de nouveaux neurones peuvent être sélectionnés les uns après les autres, et au fur et à mesure de leur intégration dans P_1 , le champ de l'innervation de Q se rétrécit, ce qui fait chuter de plus en plus de candidats déjà affaiblis par un silence prolongé. Avec ces intégrations, le groupe des neurones liés à Q finit par diminuer progressivement de volume, et se stabilise au nombre final $n_1 = \max \{n_0, s_0\}$ avec l'intégration d'un dernier neurone. Dans cette course pour le recrutement, la température joue le rôle de *diversification* en proposant des candidats, et, de leur côté, les phénomènes de coopération et de compétition opèrent une *sélection* parmi ces candidats.

4.5.e Genèse d'une chaîne par empilement de maillons

La description détaillée du §4.5.b, §4.5.c et §4.5.d a montré comment l'agrégation d'un premier paquet de neurones P_1 pouvait être provoquée par les pulsations répétées d'un groupe particulier Q . La décharge simultanée des neurones de Q induit des potentiels postsynaptiques élevés, et lorsqu'un neurone répond à ces potentiels en déchargeant à son tour, il attire immédiatement vers lui les connexions de Q : de la sorte, on assiste à l'*auto-organisation* du premier maillon $Q \rightarrow P_1$. Or, il est clair que ce processus d'agrégation et de focalisation peut maintenant se reproduire à partir de P_1 : en effet, les neurones de P_1 sont désormais *synchrones* 1 unité de temps après ceux de Q . Par conséquent, P_1 constitue un nouveau groupe pulsatile, dont les efférences ne sont pas saturées et peuvent se focaliser sur un futur deuxième paquet P_2 , exactement de la même manière que celles de Q l'avaient fait sur P_1 .

On voit donc que le mécanisme de formation du premier maillon est le stéréotype de la formation de tous les autres maillons de la chaîne : il peut se reproduire à chaque étape en adoptant le dernier paquet formé comme nouveau point de départ. Celui-ci fait office de nouveau groupe pulsatile, et provoque l'agrégation d'un paquet supplémentaire. Plus précisément, supposons *par récurrence* que la structuration du réseau est arrivée au point où il existe une chaîne de longueur $m+1$ prenant racine sur Q , soit : $Q \rightarrow P_1 \rightarrow P_2 \dots \rightarrow P_m$. Chaque décharge de Q déclenche alors une *onde d'activation* qui se propage tout le long de cette chaîne, de paquet en paquet, et qui se termine par l'activation simultanée des neurones de P_m : ainsi, P_m répercutera systématiquement la pulsation de Q avec un retard m , et ces activations feront de lui un groupe pulsatile au même titre que Q . La différence, cependant, est que les connexions sortantes de Q se sont déjà focalisées sur P_1 , et ne peuvent plus innover d'autres neurones, tandis que celles de P_m sont encore dans un état diffus et homogène, donc disponibles pour contacter de nouveaux candidats. Les pulsations de P_m entraînent alors la focalisation de ses efférences vers un autre ensemble de neurones-candidats, qui se stabilise finalement à un groupe P_{m+1} : la chaîne s'est allongée d'un nouveau maillon $P_m \rightarrow P_{m+1}$ et contient désormais les paquets $Q \rightarrow P_1 \rightarrow P_2 \dots \rightarrow P_m \rightarrow P_{m+1}$. En résumé : une chaîne peut pousser à partir d'un

premier groupe de neurones synchrones par empilements successifs de nouveaux groupes synchrones, chacun étant lié au précédent à travers une focalisation de connexions fortes.

Bien sûr, ce mécanisme d'empilement n'est pas rigoureusement itératif, et en réalité, le paquet P_{m+1} n'attend pas l'achèvement du paquet P_m pour commencer à se constituer. Autrement dit, la longue période de formation de P_{m+1} (évoquée plus haut à propos de P_1), qui comprend une expansion du groupe des candidats suivie d'une contraction autour des candidats recrutés, est en *recouvrement* notable avec celles de P_m et de P_{m+2} . En effet, avec $\Theta_0=3$, dès que P_m contient 3 ou 4 neurones recrutés (i.e. des neurones dont l'activation post- P_{m-1} est systématique), il atteint la "masse critique" suffisante pour former un groupuscule pulsatile capable à son tour d'innover "en aval" des candidats à l'intégration dans P_{m+1} . Ainsi, les premiers candidats de P_{m+1} sont sollicités par les efférences d'un groupuscule P_m qui n'est pas encore parvenu à sa pleine maturité (c'est à dire qui n'a pas encore saturé les efférences de P_{m-1}). Par conséquent, on voit que la situation récurrente est plus compliquée que la situation initiale partant de Q , car le paquet P_m qui fait office de groupe pulsatile est *lui-même en construction* au moment où il commence à innover un autre paquet P_{m+1} . Donc le maillon $P_m \rightarrow P_{m+1}$ se forme par adjonctions successives de neurones non seulement du côté de P_{m+1} , mais aussi du côté de P_m : certains neurones rejoindront P_m (i.e. seront recrutés par P_{m-1}) *après* que d'autres aient rejoint P_{m+1} (i.e. aient été recrutés par P_m). Cependant, ces permutations dans l'arrivée des neurones ne changeront pas la convergence finale vers le nombre maximal de neurones dans chaque paquet, qui est donné par n_0 .

Les neurones sont donc intégrés tantôt dans P_m , tantôt dans P_{m+1} avec cependant des probabilités d'intégration plus grande pour P_m , qui sera de ce fait achevé avant P_{m+1} : en effet, les pulsations qui organisent P_m proviennent de P_{m-1} , tandis que les pulsations qui organisent P_{m+1} proviennent de P_m , donc si on suppose par récurrence que P_{m-1} est achevé avant P_m (i.e. est plus gros que celui-ci au cours de son évolution), on en conclut que P_m sera également achevé avant P_{m+1} , puisque des groupes pulsatiles plus gros induisent des potentiels plus élevés dans le réseau, donc des probabilités plus fortes d'activation et de focalisation.

Ainsi, la chaîne en construction présente un profil typiquement "en biseau", c'est à dire des largeurs de paquets graduellement décroissantes (figure 10) : des paquets non achevés, mais déjà reliés les uns aux autres dans un ordre hiérarchique sont en compétition pour innover de futurs candidats, la lutte étant déséquilibrée en faveur des paquets plus gros, qui se trouvent "en amont" (vers la racine). L'insistance des pulsations en provenance de Q et les ondes d'activation répétées qu'elles provoquent finissent par élargir les anciens paquets jusqu'à saturation de toutes les efférences, tandis que de nouveaux paquets continuent à s'empiler. En résumé, la chaîne s'allonge avant d'atteindre sa pleine largeur, ce qui lui donne un aspect effilé en son extrémité.

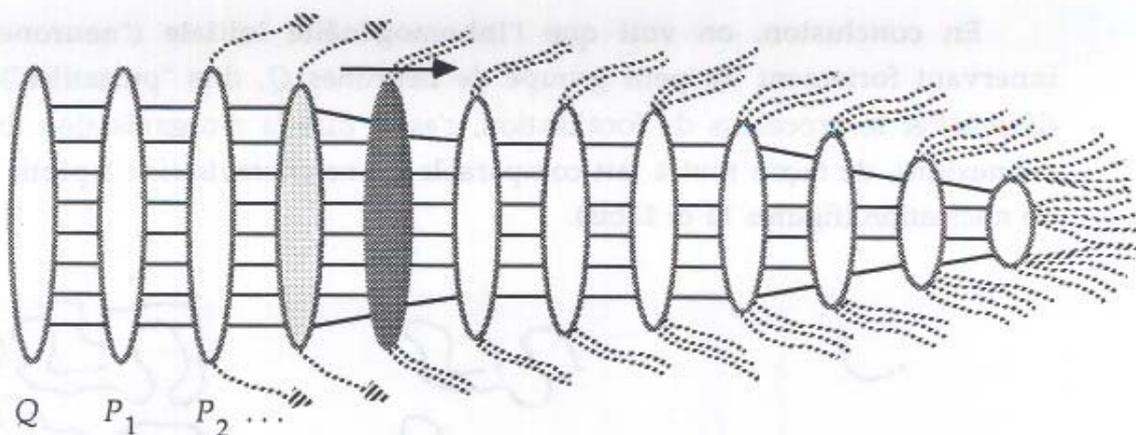


Figure 10 : Profil de la chaîne en construction. Les ellipses allongées symbolisent des groupes de neurones synchrones (leur hauteur est proportionnelle au nombre de neurones), et les segments horizontaux entre ces groupes symbolisent les connexions fortes de la chaîne (qui relie en réalité tous les neurones d'un groupe à tous les neurones du groupe suivant; mais pour ne pas surcharger le dessin, les segments "croisés" entre paquets n'ont pas été représentés). On a également représenté une onde se propageant sur la chaîne (à l'instant considéré, le paquet P_4 est activé juste après P_3). La largeur de la chaîne est graduellement décroissante, car la période de formation complète d'un paquet est en large recouvrement avec celles de ses voisins. Ainsi, la chaîne s'allonge avant de s'élargir. En conséquence, les paquets disposent d'une certaine quantité d'efférences (symbolisées par les traits en pointillés) encore disponibles pour recruter de nouveaux candidats vers l'aval (les neurones recrutés par P_k feront partie de P_{k+1} , par définition). Cette quantité efférente est plus grande pour les petits groupes (situés en aval), mais en revanche, du point de vue d'un éventuel candidat extérieur, les potentiels induits par un groupe plus gros (situé en amont) seront plus grands. La chaîne est composée de deux parties : un "tronc" stable, de largeur n_0 (ici, les trois premiers paquets), et une extrémité en forme de pointe, sorte de "tête chercheuse", qui s'allonge et s'épaissit progressivement en recrutant des candidats à la fois vers l'avant et sur les côtés. Tout ceci est possible grâce à la conjonction de deux éléments : les ondes de propagation qui parcourent la chaîne et les activations aléatoires des neurones extérieurs à la chaîne, qui provoque leur agrégation sur celle-ci (la chaîne est "nourrie" par le bruit extérieur).

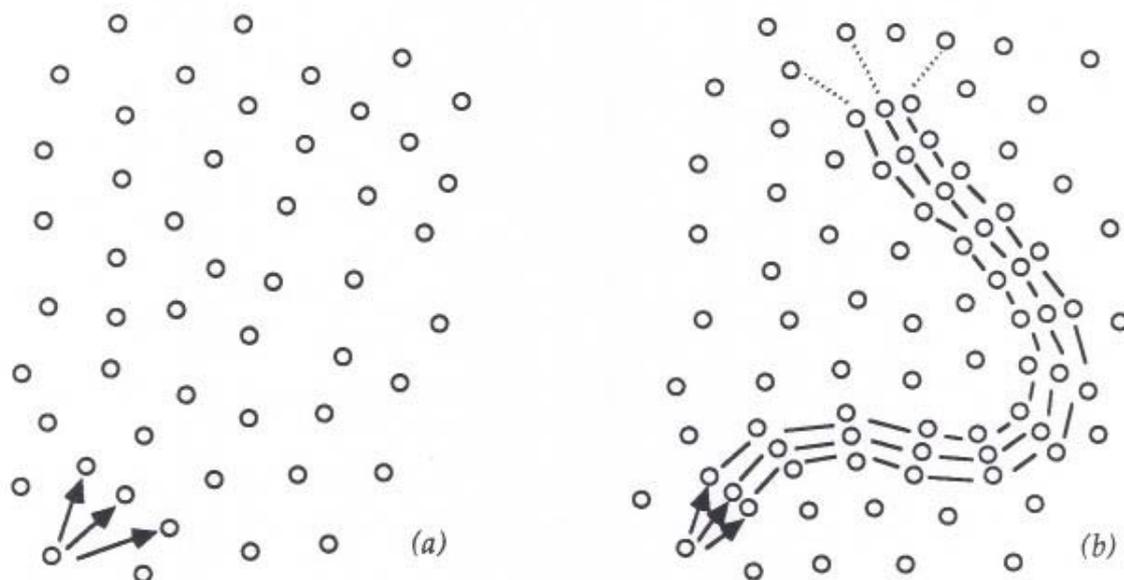


Figure 11 : Illustration de la croissance agrégative d'une chaîne. (a) Au départ, le réseau est homogène (connexions non-représentées) à l'exception des afférences d'un neurone particulier, plus fortes que la moyenne. (b) Cette singularité constitue un "germe" qui provoque la croissance "cristalline" d'une chaîne. Bien sûr, le réseau ne possède pas de topologie, et, en réalité, cette chaîne n'est pas visible (les neurones ne se déplacent pas pour former des groupes) : les illustrations que nous avons l'habitude de proposer sont donc des vues commodes, qui réorganisent spatialement le réseau pour mettre en évidence la structure linéaire de la focalisation (on aura une idée plus réaliste de l'apparence de cette focalisation en se reportant à la figure 19, chapitre 5).

En conclusion, on voit que l'inhomogénéité initiale ("neurone-source" i_0 innervant fortement un petit groupe de neurones Q , dits "pulsatiles") permet de déclencher le processus de focalisation, c'est à dire la réorganisation du réseau de connexions, de façon tout à fait comparable à une *crystallisation* à partir d'un germe de nucléation (figures 11 et 11bis).

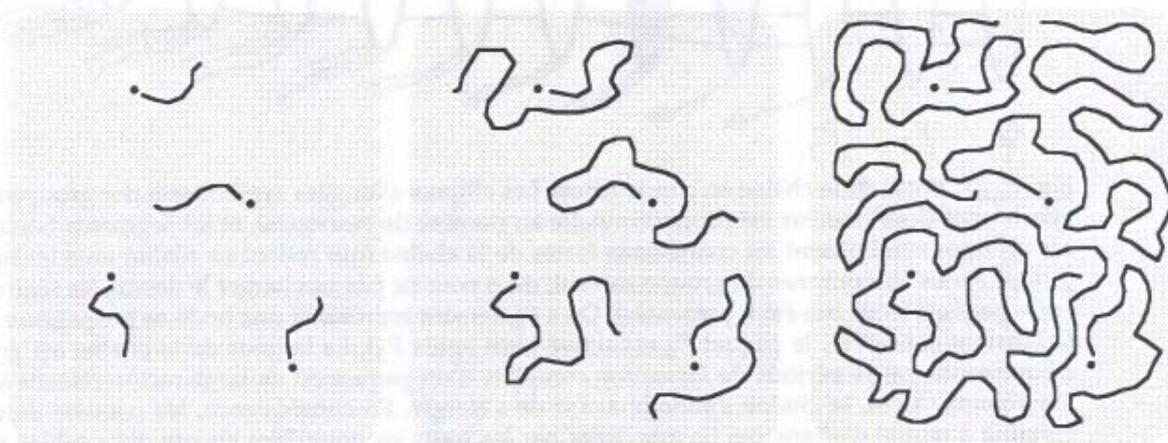


Figure 11bis : Croissance simultanée de plusieurs chaînes à partir de germes indépendants. Cette illustration montre comment plusieurs chaînes telles que celle de la figure 11 pourraient croître indépendamment grâce aux pulsations décorrélées de neurones-sources différents (points noirs). Nous anticipons ici sur le chapitre 6 qui traitera de la croissance de configurations spatio-temporelles complexes, issues de la composition de morceaux de chaînes élémentaires.



Figure 11 : Illustration de la croissance séquentielle d'une chaîne. (a) Au départ, le réseau est homogène et les connexions sont faibles. (b) L'activation d'un "neurone-source" entraîne la formation d'une "chaîne" qui se propage à travers le réseau, entraînant la formation d'une configuration spatio-temporelle complexe. Les connexions sont plus fortes dans les zones où la chaîne a été activée.

5 SIMULATIONS NUMÉRIQUES

5.1 Neurones excitateurs et délais uniformes

Dans ce premier paragraphe, nous reprendrons les conditions schématiques du chapitre 4 : le réseau ne contiendra pas d'inhibiteurs ($N_I=0$) et les délais de transmission seront uniformes ($\tau_{ij}=1$).

Pour plus de clarté dans les explications, le groupe pulsatile Q sera considéré comme *extérieur* au réseau R : ainsi les neurones de R seront numérotés de 1 à N , tandis que les neurones de Q seront numérotés de $N+1$ à $N+n_c$. Les neurones pulsatiles ne recevront pas d'afférences en provenance du reste du réseau, donc l'architecture complète contiendra $N(N-1)$ connexions $R \times R$ et $n_c N$ connexions $Q \times R$ (ainsi, en chaque neurone de R , il y aura $N-1$ efférences et $N+n_c-1$ afférences, mais ce léger surplus n'aura pas de conséquences notables sur l'équilibre des sommes à s_0).

5.1.a Formation d'un premier maillon

Nous commencerons ici par examiner le processus d'agrégation du premier paquet d'une chaîne, tel qu'il a été décrit formellement au §4.5 : dans cet exemple, le réseau contient $N=100$ neurones et le groupe pulsatile $n_c=10$ neurones, soit : $R = \{\#1, \dots, \#100\}$ et $Q = \{\#101, \dots, \#110\}$. Pour faciliter cet examen, les seuils et les connexions initiales seront uniformes : $\Theta_0=3$ et $w_0=0,1$, avec $s_0=10$. Les paramètres de la dynamique seront les suivants : $\alpha=0,1$, $\beta=0$, $\gamma=0,005$ et $T=0,5$. Les neurones pulsatiles déchargeront simultanément avec une probabilité constante $1/30$: donc il se produira environ 3 pulsations par périodes de 100 itérations.

Dans cet exemple, on constate que l'activité de Q provoque la focalisation de ses efférences vers un premier paquet de 10 neurones :

$$P_1 = \{\#9, \#35, \#40, \#46, \#55, \#64, \#77, \#81, \#88, \#100\}.$$

Les figures 13 et 14 montrent comment ces neurones sont progressivement sélectionnés, c'est à dire comment certains candidats se lient définitivement à Q , et comment d'autres candidats, initialement contactés, finissent par s'en détacher faute d'avoir été suffisamment réactivés : sur la figure 13 on peut suivre l'évolution de l'état des efférences du groupe pulsatile, et, réciproquement, sur la figure 14, l'évolution de ces connexions du point de vue de certains neurones post-pulsatiles (donc ces graphiques présentent respectivement le point de vue spatial et le point de vue temporel de la formation du premier paquet; cf. explications de la figure 12).

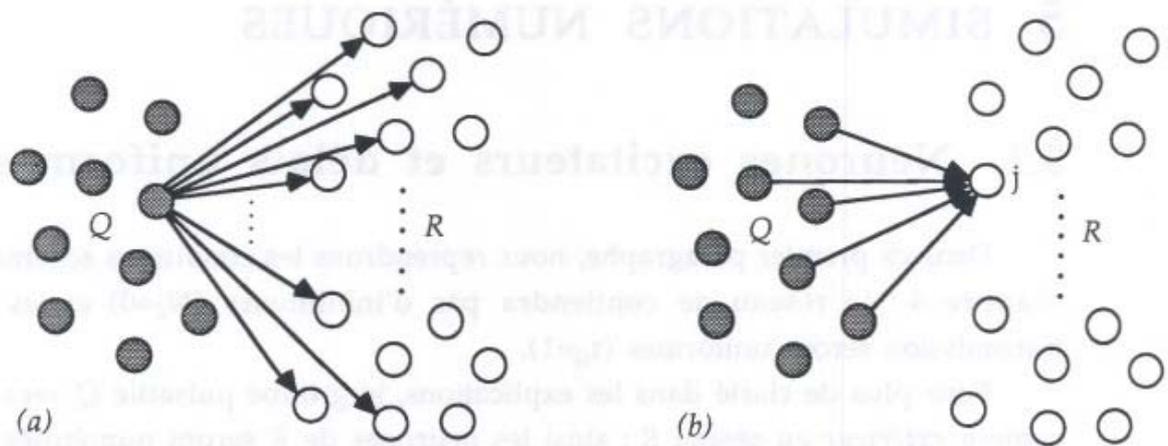


Figure 12 : Vue schématique du groupe pulsatile Q (en gris) et de l'ensemble du réseau R (en blanc). (a) Ce que représentera la figure 13 : paysage des connexions efférentes de l'un des neurones pulsatiles (un tableau de la figure 13 = un instant post-pulsatile différent). (b) Ce que représentera la figure 14 : évolution au cours du temps des afférences en provenance de Q d'un neurone j. Remarquons que ces connexions sont toujours identiques, donc la figure 14 montre l'évolution de leur valeur commune (un tableau de la figure 14 = un neurone j différent).

La figure 15 affiche l'activité instantanée du réseau au cours des 1000 premières itérations : au bout de cette période, le premier paquet P_1 est quasiment achevé, et un deuxième paquet, P_2 , a déjà commencé à se constituer.

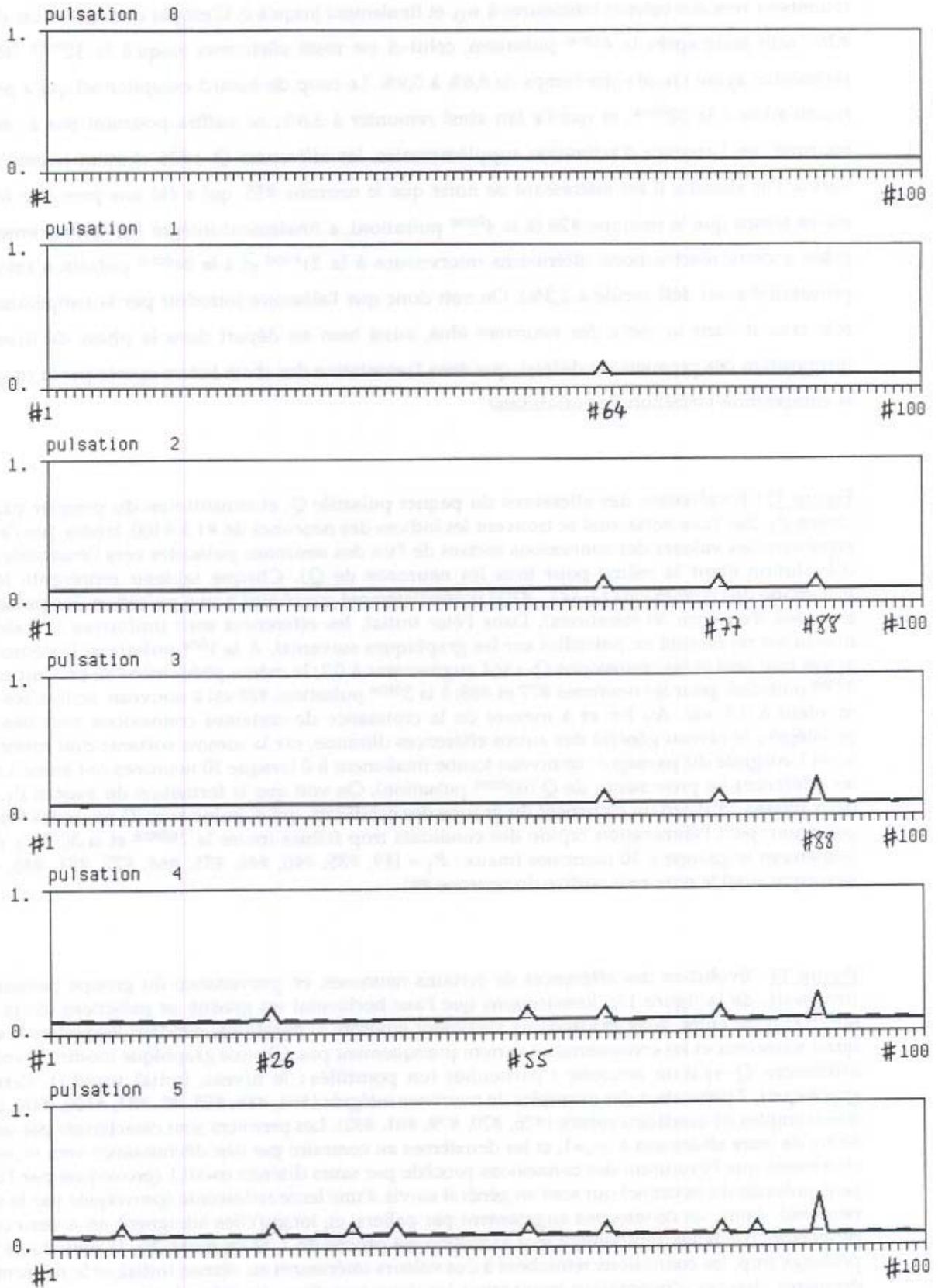
Description détaillée commune aux figure 13 et 14 :

Au départ, le réseau est quasi silencieux : partant d'un état vide, les probabilités d'activations valent toutes $\sigma_T(-\theta_0) = 0,25\%$, ce qui équivaut en moyenne à 1 seul neurone sur 100 activé toutes les 4 itérations. Juste après la 1^{ère} décharge du groupe Q, le niveau uniforme des probabilités augmente à 1,7%, et un premier neurone, le #64, se trouve activé : ses chances de réactivations post-pulsatiles deviennent alors 8,7%, ce qui le place en position avancée par rapport aux autres. Après la 2^{ème} pulsation, le #64 n'est pourtant pas réactivé, et deux autres le sont à sa place : #77 et #88. Les probabilités de future réactivation post-pulsatile sont alors : $q_{64}=q_{77}=q_{88}=8,7\%$ et $q=1,6\%$ sur le reste du réseau. A la 3^{ème} pulsation, le neurone #88 est justement réactivé, ce qui le propulse à 36%. Ce neurone gravit ensuite très rapidement les échelons : après un temps d'attente au moment de la 4^{ème} pulsation, il est à nouveau activé à la 5^{ème} et passe à 75%. Il est ensuite systématiquement réactivé et atteint définitivement 100% en 3 pulsations (à travers 95%, 98% et 99,8%). Le neurone #64 est beaucoup plus long à se décider : ce n'est qu'après la 20^{ème} pulsation qu'il est à nouveau activé, alors que le faible avantage qu'il possédait avait déjà beaucoup diminué : après cette longue période d'attente, il ne se trouvait plus qu'à 2,7% au lieu des 8,7% du début. C'est donc in extremis que ce neurone se rattrape, et fait remonter ses chances à 13,6%. Le hasard veut alors qu'il soit une nouvelle fois actif après la pulsation suivante (la 21^{ème}), ce qui le pousse encore vers 47,6%. A partir de cet instant, le neurone #64 peut être considéré comme définitivement récupéré, ce qui est confirmé par les événements ultérieurs : il ne cesse d'être réactivé à chaque pulsation et atteint 100% à la 25^{ème}. Quant au neurone #77, son destin est très semblable à celui du #64, à la différence qu'il décolle un peu plus tôt que ce dernier, i.e. après la 10^{ème} pulsation.

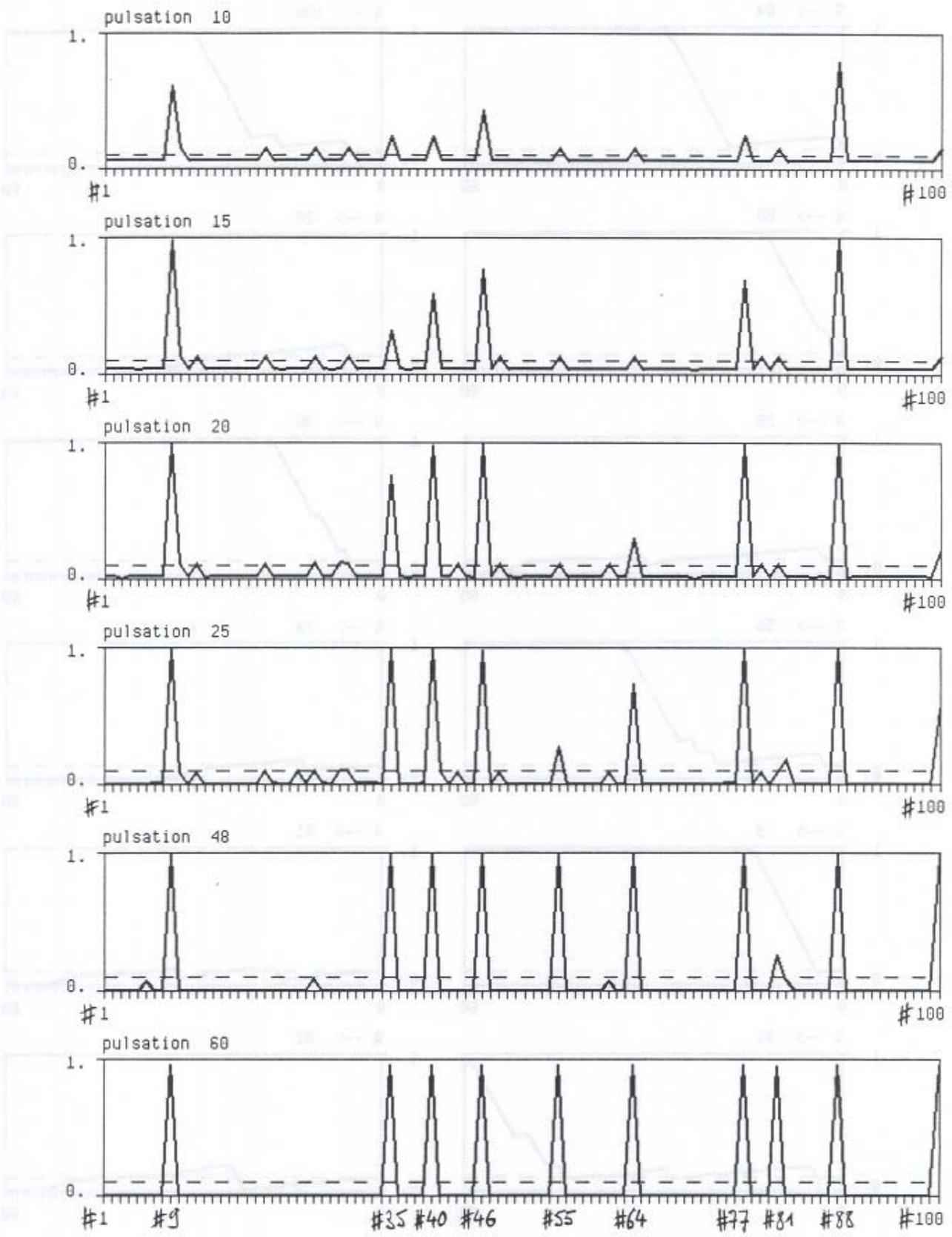
D'autres neurones, qui ont été activés après les toutes premières pulsations à l'instar du #64 ou du #88, n'ont pas eu la chance de ces derniers : faute de réactivations suffisantes, leurs afférences sont retombées vers des valeurs inférieures à w_0 , et finalement jusqu'à 0. C'est par exemple le cas du neurone #26 : actif juste après la 4^{ème} pulsation, celui-ci est resté silencieux jusqu'à la 32^{ème} itération, sa probabilité ayant chuté entre temps de 8,6% à 0,9%. Le coup de hasard exceptionnel qui a provoqué sa réactivation à la 32^{ème}, et qui l'a fait ainsi remonter à 3,6%, ne suffira pourtant pas à récupérer ce neurone : en l'absence d'activation supplémentaire, les afférences $Q \rightarrow \#26$ chutent irrémédiablement vers 0. Par ailleurs, il est intéressant de noter que le neurone #55, qui a été une première fois actif en même temps que le neurone #26 (à la 4^{ème} pulsation), a finalement intégré P_1 , contrairement au #26, grâce à deux réactivations ultérieures intervenues à la 21^{ème} et à la 24^{ème} pulsation (alors que sa probabilité avait déjà reculé à 2,3%). On voit donc que l'aléatoire introduit par la température joue un rôle crucial dans le choix des neurones élus, aussi bien au départ dans la phase de diversification (proposition des premiers candidats), que dans l'orientation des choix futurs opérés par la coopération et la compétition (sélection des candidats).

Figure 13 : Focalisation des efférences du paquet pulsatile Q , et constitution du premier paquet de la chaîne P_1 . Sur l'axe horizontal se trouvent les indices des neurones de #1 à #100, tandis que l'axe vertical représente les valeurs des connexions sortant de l'un des neurones pulsatiles vers l'ensemble du réseau (l'évolution étant la même pour tous les neurones de Q). Chaque tableau représente le paysage instantané des connexions $Q \rightarrow \#1... \#100$ immédiatement consécutif à une pulsation (les pulsations sont espacées d'environ 30 itérations). Dans l'état initial, les efférences sont uniformes et valent 0,1 (ce niveau est représenté en pointillés sur les graphiques suivants). A la 1^{ère} pulsation, le neurone #64 est activé tout seul et les connexions $Q \rightarrow \#64$ augmentent à 0,2; le même phénomène se produit ensuite à la 2^{ème} pulsation pour les neurones #77 et #88; à la 3^{ème} pulsation, #88 est à nouveau actif et ses afférences montent à 0,3; etc. Au fur et à mesure de la croissance de certaines connexions vers des neurones privilégiés, le niveau général des autres efférences diminue, car la somme sortante doit rester constante (c'est l'intégrale du paysage) : ce niveau tombe finalement à 0 lorsque 10 neurones ont attiré à eux toutes les efférences en provenance de Q (60^{ème} pulsation). On voit que la formation du paquet P_1 passe par deux phases : d'abord un étalement du groupe des candidats, qui culmine avec 23 neurones vers la 25^{ème} pulsation, puis l'élimination rapide des candidats trop faibles (entre la 25^{ème} et la 30^{ème}), qui réduit finalement le groupe à 10 neurones finaux : $P_1 = \{\#9, \#35, \#40, \#46, \#55, \#64, \#77, \#81, \#88, \#100\}$. On remarque aussi la remontée tardive du neurone #81.

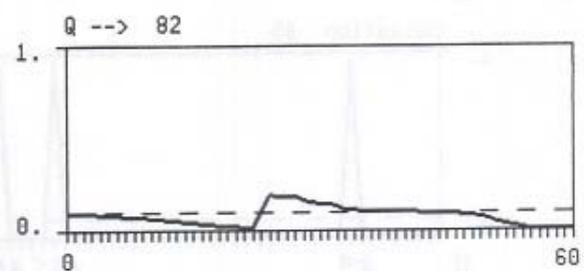
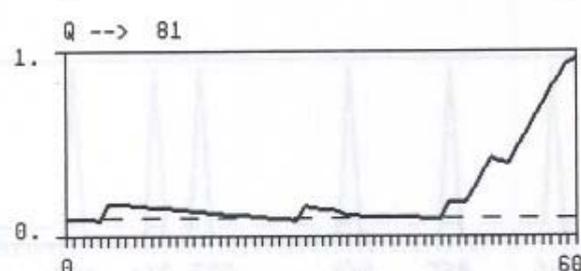
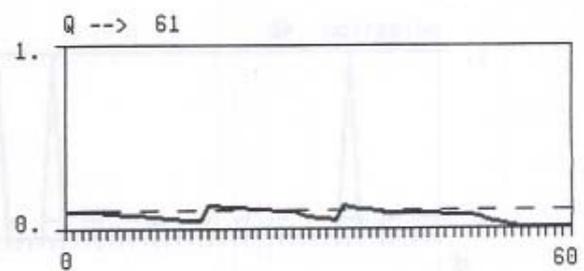
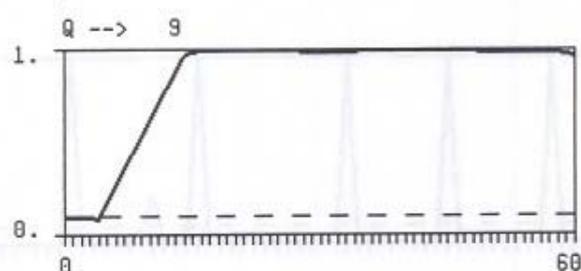
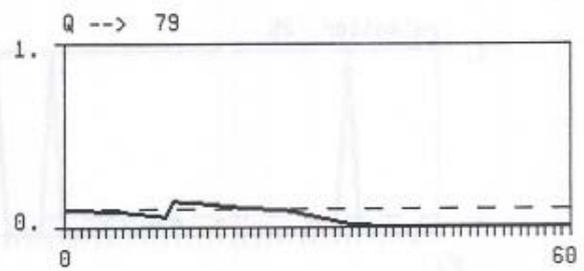
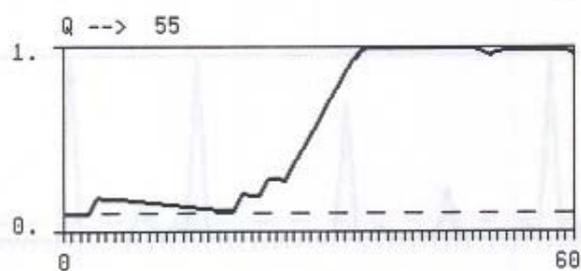
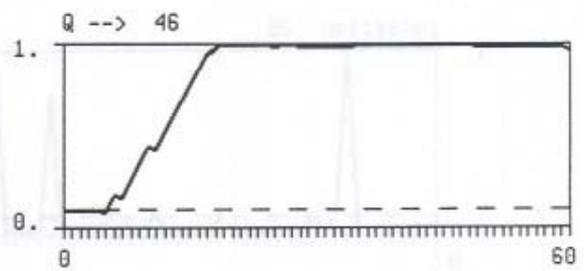
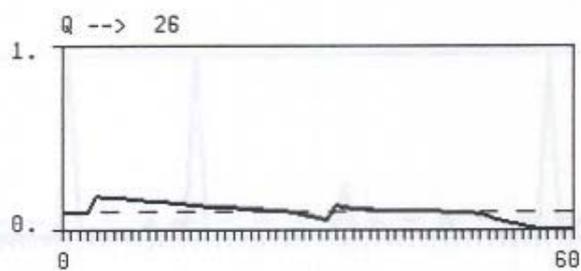
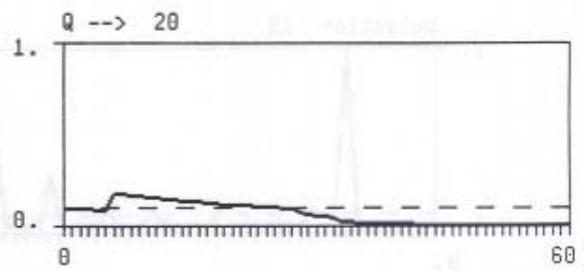
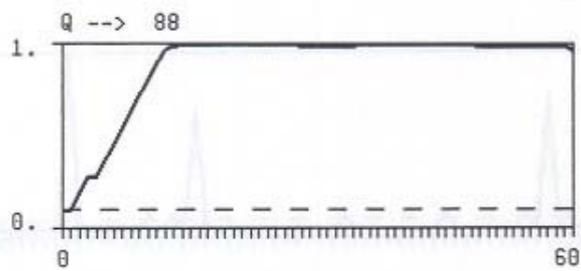
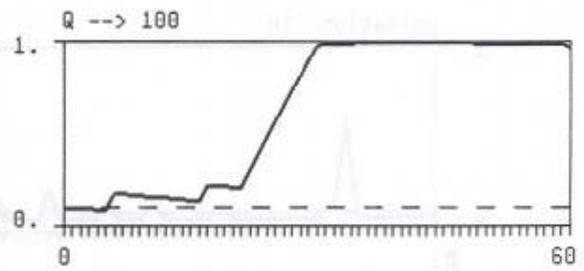
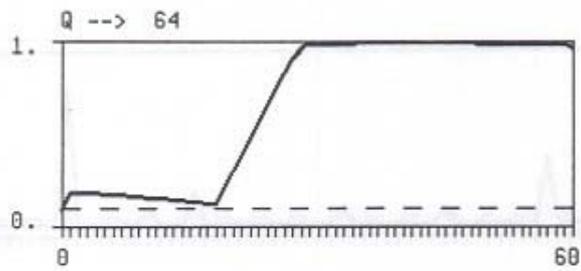
Figure 14 : Evolution des afférences de certains neurones, en provenance du groupe pulsatile (coupe temporelle de la figure 13). Remarquons que l'axe horizontal est gradué en pulsations (de la 0^{ème} à la 60^{ème}) : donc entre deux graduations s'écoulent environ 30 itérations, pendant lesquelles le réseau est quasi silencieux et les connexions ne varient pratiquement pas. Chaque graphique montre l'évolution des afférences $Q \rightarrow j$ d'un neurone j particulier (en pointillés : le niveau initial $w_0=0,1$). Parmi les 12 graphiques, 7 présentent des exemples de neurones intégrés (#64, #88, #55, #9, #81, #100, #46), et 5 autres des exemples de candidats rejetés (#26, #20, #79, #61, #82). Les premiers sont caractérisés par une montée finale de leurs afférences à $w_+=1$, et les deuxièmes au contraire par une décroissance vers $w_-=0$. On voit clairement que l'évolution des connexions procède par sauts discrets $\alpha=+0,1$ (provoqués par l'activation post-pulsatile du neurone) qui sont en général suivis d'une lente redescende (provoquée par le silence du neurone). Ainsi, les connexions augmentent par paliers, et, lorsqu'elles atteignent un niveau critique, la probabilité d'achever rapidement leur ascension est proche de 1. Si en revanche, la période de silence se prolonge trop, les connexions retombent à des valeurs inférieures au niveau initial, et le neurone perd ses dernières chances d'intégration (remarquer les deux tentatives avortées du neurone #61, ainsi que la double remontée exceptionnelle du neurone #82, à la 24^{ème} et 25^{ème} pulsation, qui a également échoué par la suite). Encore une fois, on peut apprécier la remontée tardive du neurone #81, qui, actif dès la 5^{ème} pulsation, est finalement intégré après une période d'"hésitation" s'étalant sur 45 pulsations.



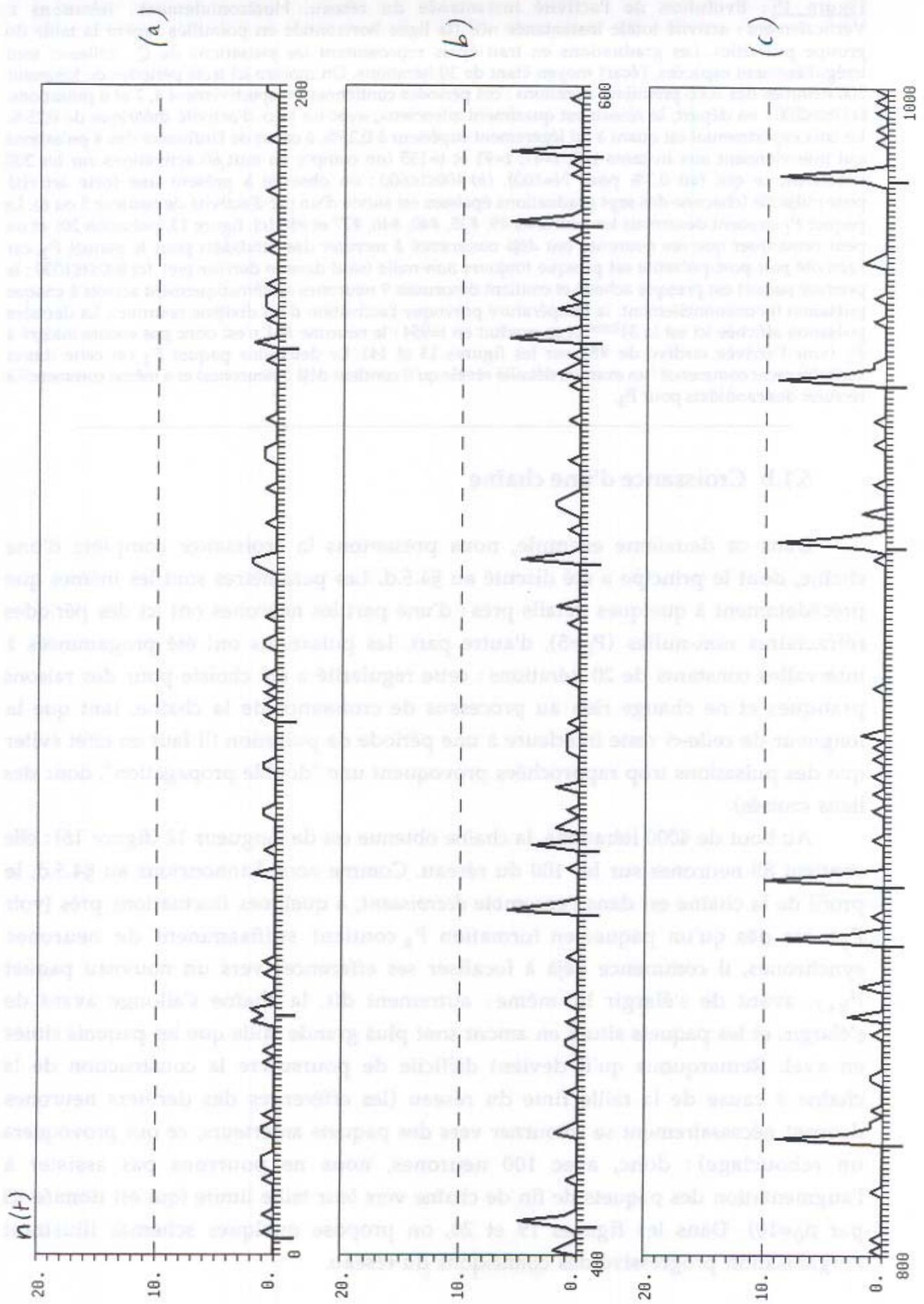
(figure 13, début)



(figure 13; fin)



(figure 14)



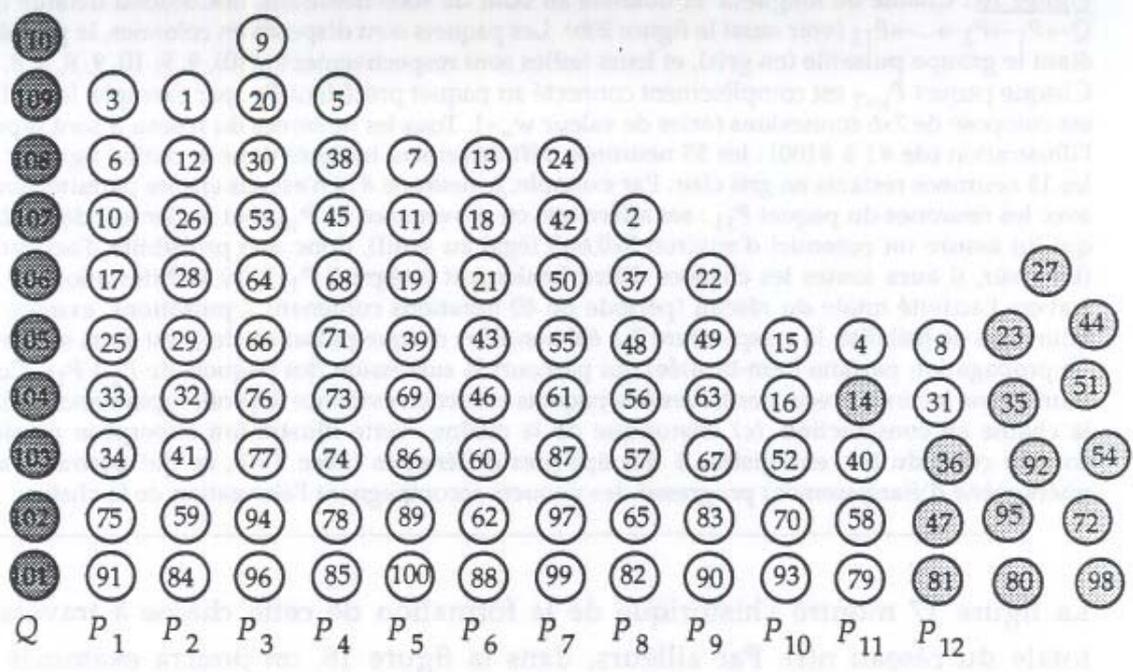
(figure 15)

Figure 15 : Evolution de l'activité instantanée du réseau. Horizontalement : itérations t . Verticalement : activité totale instantanée $n(t)$ (la ligne horizontale en pointillés repère la taille du groupe pulsatile). Les graduations en trait épais représentent les pulsations de Q : celles-ci sont irrégulièrement espacées, l'écart moyen étant de 30 itérations. On montre ici trois périodes de longueur 200 extraites des 1000 premières itérations : ces périodes contiennent respectivement 4, 7 et 6 pulsations. (a) $0 \leq t \leq 200$: au départ, le réseau est quasiment silencieux, avec un taux d'activité théorique de 0,25%. Le taux expérimental est quant à lui légèrement supérieur à 0,25%, à cause de l'influence des 4 pulsations qui interviennent aux instants $t=3$, $t=41$, $t=91$ et $t=155$ (on compte en tout 60 activations sur les 200 itérations, ce qui fait 0,3% pour $N=100$). (b) $400 \leq t \leq 600$: on observe à présent une forte activité post-pulsatile (chacune des sept graduations épaisses est suivie d'un pic d'activité de hauteur 5 ou 6). Le paquet P_1 contient désormais les neurones #9, #35, #40, #46, #77 et #88 (cf. figure 13, pulsation 20), et on peut remarquer que ces neurones ont déjà commencé à recruter des candidats pour le paquet P_2 , car l'activité post-post-pulsatile est presque toujours non-nulle (sauf dans le dernier pic). (c) $800 \leq t \leq 1000$: le premier paquet est presque achevé et contient désormais 9 neurones systématiquement activés à chaque pulsation (occasionnellement, la température provoque l'activation d'un dixième neurone). La dernière pulsation affichée ici est la 31^{ème} et se produit en $t=954$: le neurone #81 n'est donc pas encore intégré à P_1 (voir l'arrivée tardive de #81 sur les figures 13 et 14). Le deuxième paquet P_2 est cette fois-ci véritablement commencé (un examen détaillé révèle qu'il contient déjà 5 neurones) et a même commencé à recruter des candidats pour P_3 .

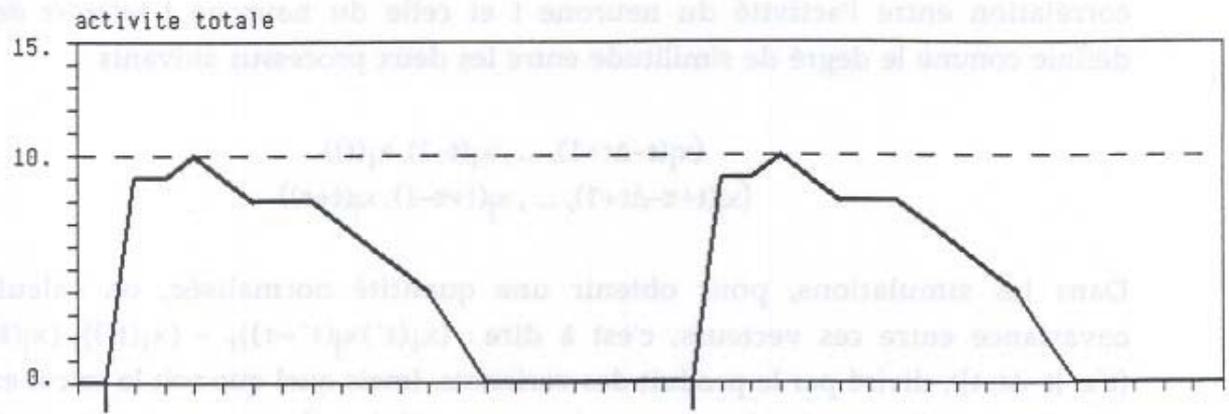
5.1.b Croissance d'une chaîne

Dans ce deuxième exemple, nous présentons la croissance complète d'une chaîne, dont le principe a été discuté au §4.5.d. Les paramètres sont les mêmes que précédemment à quelques détails près : d'une part les neurones ont ici des périodes réfractaires non-nulles ($P_0=5$), d'autre part, les pulsations ont été programmées à intervalles constants de 20 itérations : cette régularité a été choisie pour des raisons pratiques et ne change rien au processus de croissance de la chaîne, tant que la longueur de celle-ci reste inférieure à une période de pulsation (il faut en effet éviter que des pulsations trop rapprochées provoquent une "double propagation", donc des liens croisés).

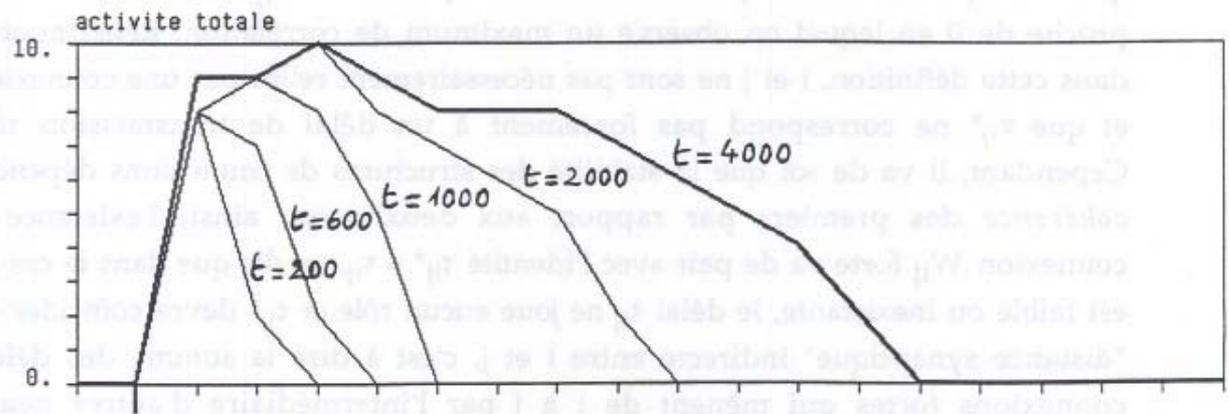
Au bout de 4000 itérations, la chaîne obtenue est de longueur 12 (figure 16) : elle contient 85 neurones sur les 100 du réseau. Comme nous l'annonçons au §4.5.d, le profil de la chaîne est dans l'ensemble *décroissant*, à quelques fluctuations près (voir P_3), car dès qu'un paquet en formation P_k contient suffisamment de neurones synchrones, il commence déjà à focaliser ses efférences vers un nouveau paquet P_{k+1} , avant de s'élargir lui-même : autrement dit, la chaîne s'allonge avant de s'élargir, et les paquets situés en amont sont plus grande taille que les paquets situés en aval. Remarquons qu'il devient difficile de poursuivre la construction de la chaîne à cause de la taille finie du réseau (les efférences des derniers neurones devront nécessairement se retourner vers des paquets antérieurs, ce qui provoquera un rebouclage) : donc, avec 100 neurones, nous ne pourrons pas assister à l'augmentation des paquets de fin de chaîne vers leur taille limite (qui est donnée ici par $n_0=10$). Dans les figures 19 et 20, on propose quelques schémas illustrant l'organisation progressive des connexions du réseau.



(a)



(b)



(c)

Figure 16 : Chaîne de longueur 12 obtenue au bout de 4000 itérations. (a) Contenu détaillé de la chaîne $Q \rightarrow P_1 \rightarrow P_2 \rightarrow \dots \rightarrow P_{12}$ (voir aussi la figure 20b). Les paquets sont disposés en colonnes, la première colonne étant le groupe pulsatile (en gris), et leurs tailles sont respectivement : (10), 9, 9, 10, 9, 8, 8, 8, 7, 6, 5, 4, 2. Chaque paquet P_{k+1} est complètement connecté au paquet précédent P_k : par exemple, le maillon $P_8 \rightarrow P_9$, est composé de 7×6 connexions fortes de valeur $w_{+}=1$. Tous les neurones du réseau R sont représentés sur l'illustration (de #1 à #100) : les 85 neurones définitivement intégrés dans la chaîne figurent en blanc, et les 15 neurones restants en gris clair. Par exemple, le neurone #14 n'est pas encore parfaitement synchrone avec les neurones du paquet P_{11} : ses afférences en provenance de P_{10} sont seulement de l'ordre de 0,6, ce qui lui assure un potentiel d'environ $5 \times 0,6 = 3$ (égal au seuil), donc une probabilité d'activation de 50% (bien sûr, il aura toutes les chances d'être finalement intégré à P_{11}). (b) Manifestation de la chaîne à travers l'activité totale du réseau (période de 40 itérations contenant 2 pulsations, extraite en $t=4000$). Pour plus de lisibilité, la température T a été annulée : chaque pulsation de Q est alors suivie d'une onde de propagation parfaite (non-bruitée) qui parcourt la succession des paquets de P_1 à P_{12} . Ce graphique fournit une lecture directe des tailles des paquets et met en évidence le profil typiquement décroissant de la chaîne en construction. (c) Historique de la chaîne : cette illustration superpose plusieurs profils comme celui du (b), enregistrés à des époques différentes (avec $T=0$), ce qui permet d'apprécier le phénomène d'élargissement progressif des paquets accompagnant l'élongation de la chaîne.

La figure 17 montre l'historique de la formation de cette chaîne à travers l'activité totale du réseau $n(t)$. Par ailleurs, dans la figure 18, on pourra examiner quelques *corrélogrammes* des activités neuronales : étant donné une période d'enregistrement $]t-\Delta t, t]$ et un certain délai τ (positif ou négatif, avec $|\tau| \ll \Delta t$), la corrélation entre l'activité du neurone i et celle du neurone j retardée de τ est définie comme le degré de similitude entre les deux processus suivants :

$$\frac{(x_i(t-\Delta t+1), \dots, x_i(t-1), x_i(t))}{(x_j(t+\tau-\Delta t+1), \dots, x_j(t+\tau-1), x_j(t+\tau))}$$

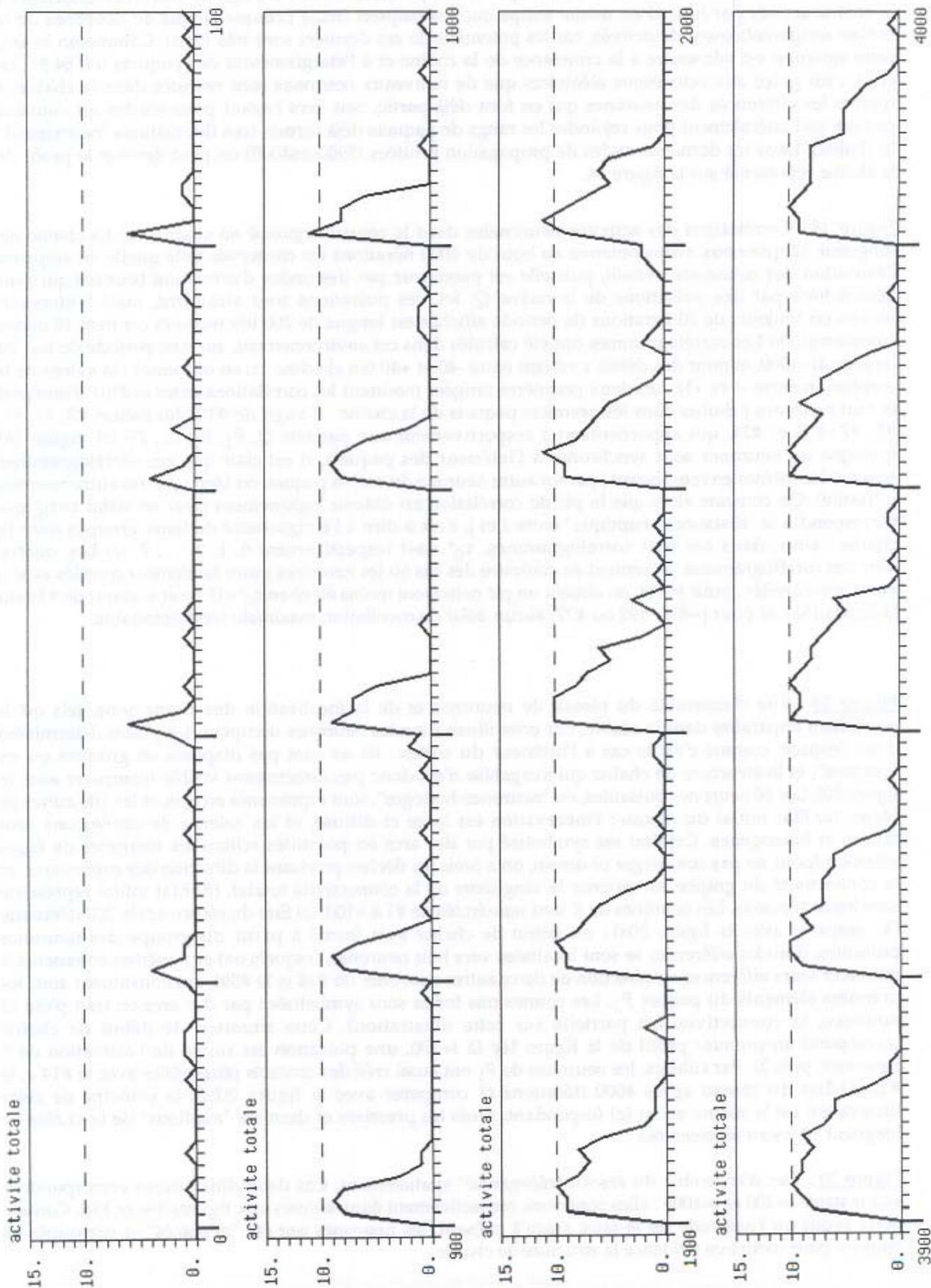
Dans les simulations, pour obtenir une quantité normalisée, on calculera la covariance entre ces vecteurs, c'est à dire : $\langle x_i(t')x_j(t'+\tau) \rangle_{t'} - \langle x_i(t') \rangle_{t'} \langle x_j(t'+\tau) \rangle_{t'}$ ($t' \in]t-\Delta t, t]$), divisé par le produit des variances. (mais quel que soit le calcul exact, le principe est de prendre le neurone i comme origine des temps, et de confronter systématiquement son activité à la réponse de j $|\tau|$ millisecondes plus tard ou plus tôt). On s'intéressera alors aux éventuels *pics* de ces corrélogrammes et aux délais qui correspondent à ces pics : on notera en particulier τ_{ij}^* le délai positif le plus proche de 0 en lequel on observe un maximum de corrélation. Remarquons que, dans cette définition, i et j ne sont pas nécessairement reliés par une connexion (i,j), et que τ_{ij}^* ne correspond pas forcément à un délai de transmission réel τ_{ij} . Cependant, il va de soi que la stabilité des structures de connexions dépend de la *cohérence* des premiers par rapport aux deuxièmes : ainsi, l'existence d'une connexion W_{ij} forte va de pair avec l'identité $\tau_{ij}^* = \tau_{ij}$, tandis que dans le cas où W_{ij} est faible ou inexistante, le délai τ_{ij} ne joue aucun rôle et τ_{ij}^* devra coïncider avec la "distance synaptique" indirecte entre i et j , c'est à dire la somme des délais des connexions fortes qui mènent de i à j par l'intermédiaire d'autres neurones : $\tau_{ij}^* = \tau_{ik} + \dots + \tau_{k'j}$. Ces propriétés particulières des CST, que nous aurons l'occasion de mieux apprécier avec des délais non-uniformes au §5.2, sont déjà observables dans la présente situation simple, où tous les τ_{ij} valent 1 (figure 18).

Figure 17 : Evolution de l'activité totale $n(t)$ du réseau R . Chaque tableau représente une période de 100 itérations (contenant 5 pulsations régulières). On voit clairement l'agrandissement de la chaîne d'une époque à l'autre, à travers des ondes de propagation de plus en plus longues. Les profils de ces ondes sont irréguliers car les activations sont bruitées par la température $T=0,5$: il s'agit de neurones extérieurs à la chaîne activés par hasard en même temps que les paquets (mais presque jamais de neurones de la chaîne accidentellement désactivés, car les potentiels de ces derniers sont très forts). Comme on le sait, cette agitation est nécessaire à la croissance de la chaîne et à l'élargissement des paquets (cf. §4.5) : en effet, c'est grâce aux activations aléatoires que de nouveaux neurones sont recrutés dans la chaîne, à travers les efférences des neurones qui en font déjà partie, soit vers l'avant pour fonder un nouveau paquet, soit latéralement pour rejoindre les rangs de paquets déjà formés (ces fluctuations "nourrissent" la chaîne). Dans les dernières ondes de propagation bruitées ($3900 \leq t \leq 4000$) on peut deviner le profil de la chaîne représenté sur la figure 16.

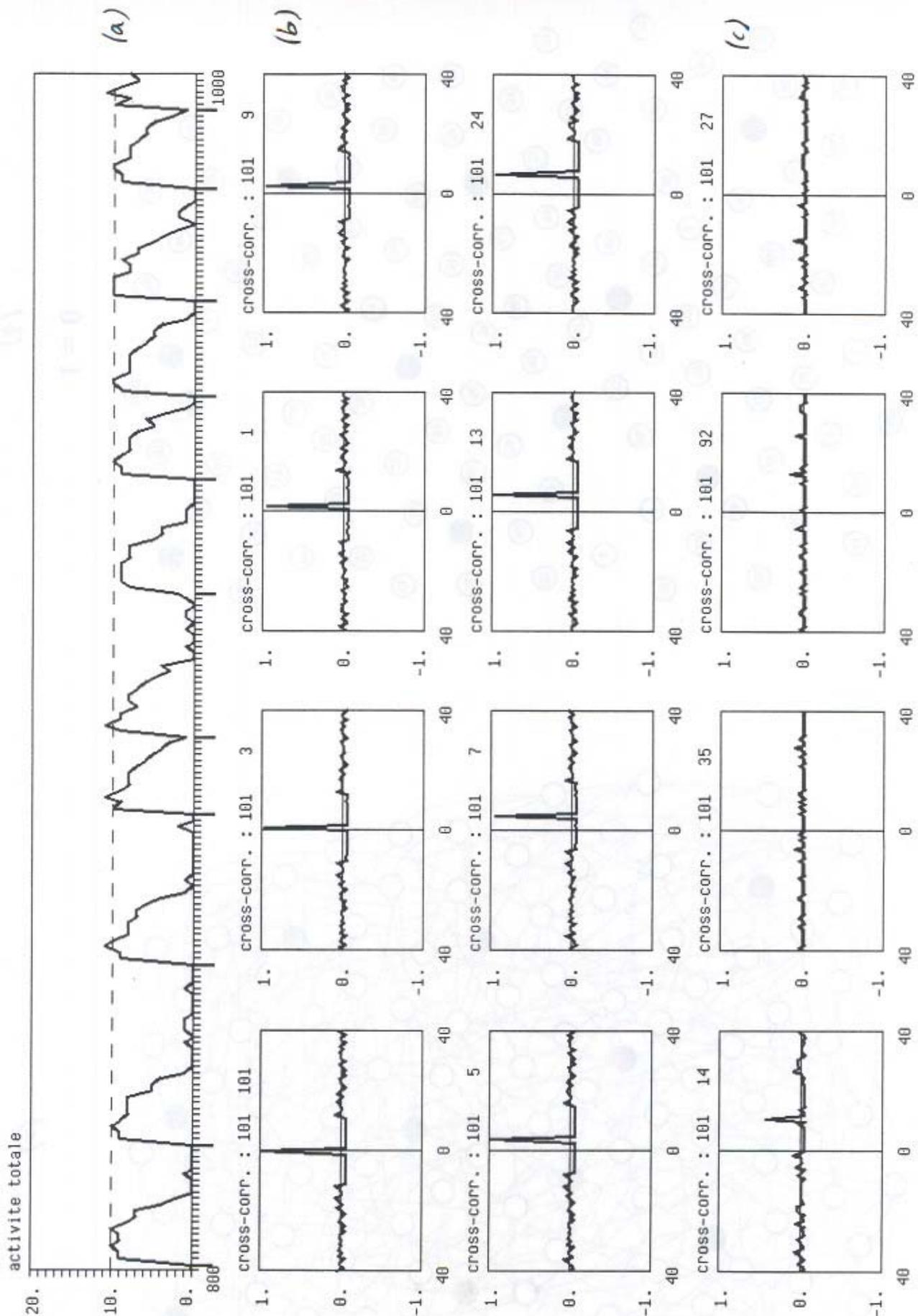
Figure 18 : Corrélations des activités neuronales dans le réseau organisé en chaîne. (a) La chaîne de longueur 12 que nous avons obtenue au bout de 4000 itérations est conservée telle quelle en stoppant l'évolution des connexions ($\alpha=0$), puis elle est parcourue par des ondes d'activation bruitées qui sont déclenchées par des pulsations de la racine Q . Ici, ces pulsations sont aléatoires, mais l'intervalle moyen est toujours de 20 itérations (la période affichée est longue de 200 itérations et contient 10 ondes complètes). (b) Les corrélogrammes ont été calculés dans cet environnement, sur une période de test de largeur $\Delta t=1000$, et pour des délais τ variant entre -40 et +40 (en abscisse : τ ; en ordonnée : la valeur de la corrélation entre -1 et +1). Les deux premières rangées montrent les corrélations entre $i=\#101$ d'une part et huit neurones j choisis dans les premiers paquets de la chaîne : il s'agit de #101 lui-même, #3, #1, #9, #5, #7, #13 et #24, qui appartiennent à respectivement aux paquets Q, P_1, P_2, \dots, P_7 (cf. figure 16) (puisque les neurones sont synchrones à l'intérieur des paquets, il est clair que ces corrélogrammes seraient les mêmes en remplaçant j par un autre neurone du même paquet, ou bien i par un autre neurone pulsatile). On constate alors que le pic de corrélation est obtenu logiquement pour un délai $\tau=\tau_{ij}^*$ qui correspond à la "distance synaptique" entre i et j , c'est à dire à l'éloignement de leurs groupes dans la chaîne : ainsi, dans ces huit corrélogrammes, τ_{ij}^* vaut respectivement 0, 1, 2, ..., 7. (c) Les quatre derniers corrélogrammes présentent au contraire des cas où les neurones j sont faiblement corrélés avec i , voire non-corrélés : pour $j=\#14$, on obtient un pic nettement moins élevé en $\tau_{ij}^*=11$ (voir le statut de #14 sur la figure 16a), et pour $j=\#35, \#92$ ou #27, aucun délai de corrélation maximale n'est détectable.

Figure 19 : Vue d'ensemble du réseau de neurones et de la focalisation des connexions, tels qu'ils pourraient apparaître dans la réalité. Sur cette illustration les neurones occupent des places déterminées dans l'espace, comme c'est le cas à l'intérieur du cortex : ils ne sont pas disposés en groupes ou en "paquets", et la structure de chaîne qui s'organise n'est donc pas directement visible (comparer avec la figure 20). Les 10 neurones pulsatiles, ou "neurones-horloges", sont représentés en gris, et les 100 autres en blanc. (a) Etat initial du réseau : l'innervation est large et diffuse, et les valeurs de connexions sont faibles et homogènes. Cet état est symbolisé par des arcs en pointillés reliant les neurones de façon aléatoire (pour ne pas surcharger ce dessin, on a omis les flèches précisant la direction des connexions, et la connectivité du graphe est environ le vingtième de la connectivité totale). (b) Etat initial représenté sans les connexions. Les neurones de R sont numérotés de #1 à #100. (c) Etat du réseau après 200 itérations (à comparer avec la figure 20a) : un début de chaîne s'est formé à partir du groupe des neurones pulsatiles, dont les efférences se sont focalisées vers huit neurones, lesquels ont eux-mêmes commencé à renforcer leurs efférences en direction de deux autres neurones (le #28 et le #59), qui constituent ainsi les premiers éléments du paquet P_2 . Les connexions fortes sont symbolisées par des arcs en trait plein (à nouveau, la connectivité est partielle sur cette illustration). Cette situation de début de chaîne correspond au premier profil de la figure 16c (à $t=200$, une pulsation est suivie de l'activation de 8 neurones, puis 2). Par ailleurs, les neurones de P_1 ont aussi créé des contacts plus faibles avec le #14 et le #1. (d) Etat du réseau après 4000 itérations (à comparer avec la figure 20b) : le principe de cette illustration est le même qu'au (c) (cependant, seuls les premiers et derniers "maillons" de la chaîne de longueur 12 y sont représentés).

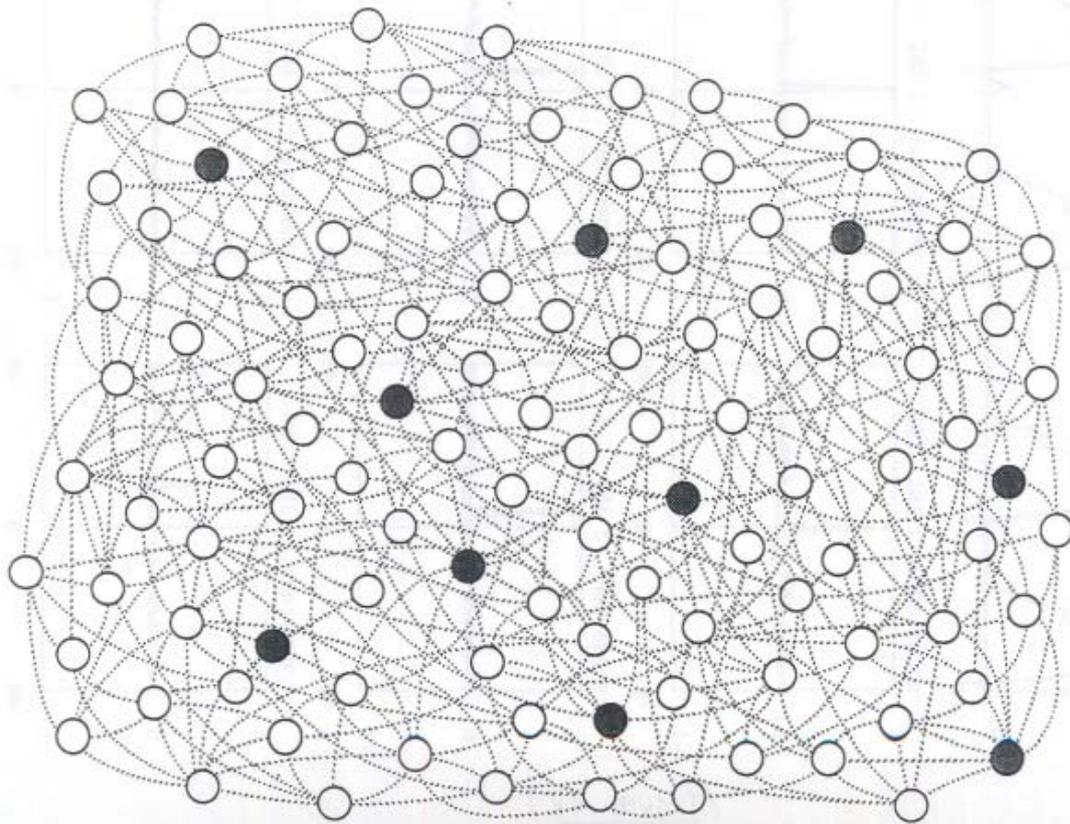
Figure 20 : Vue d'ensemble du réseau "réorganisé" spatialement. Ces deux illustrations correspondent aux instants $t=200$ et $t=4000$: elles sont donc respectivement équivalentes aux figures 19c et 19d. Comme nous avons eu l'habitude de le faire jusqu'à présent, les neurones ont été "déplacés" et regroupés en paquets pour mettre en évidence la structure de chaîne.



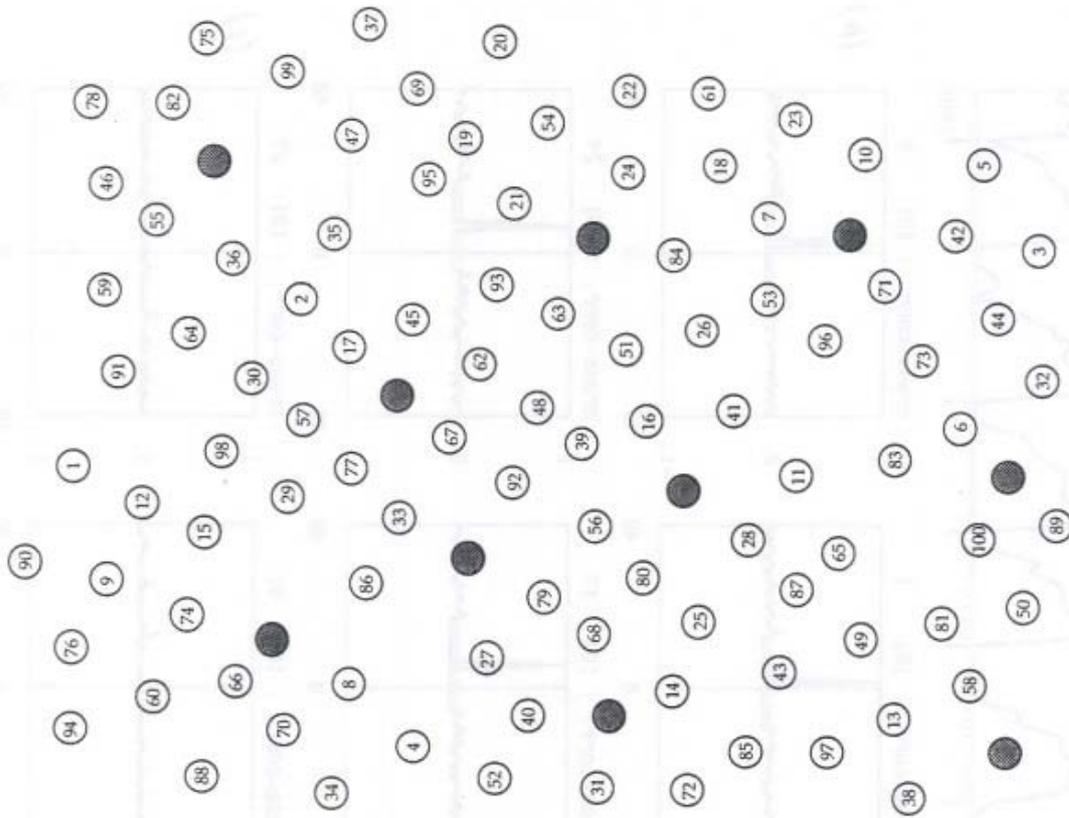
(figure 17)



(figure 18)



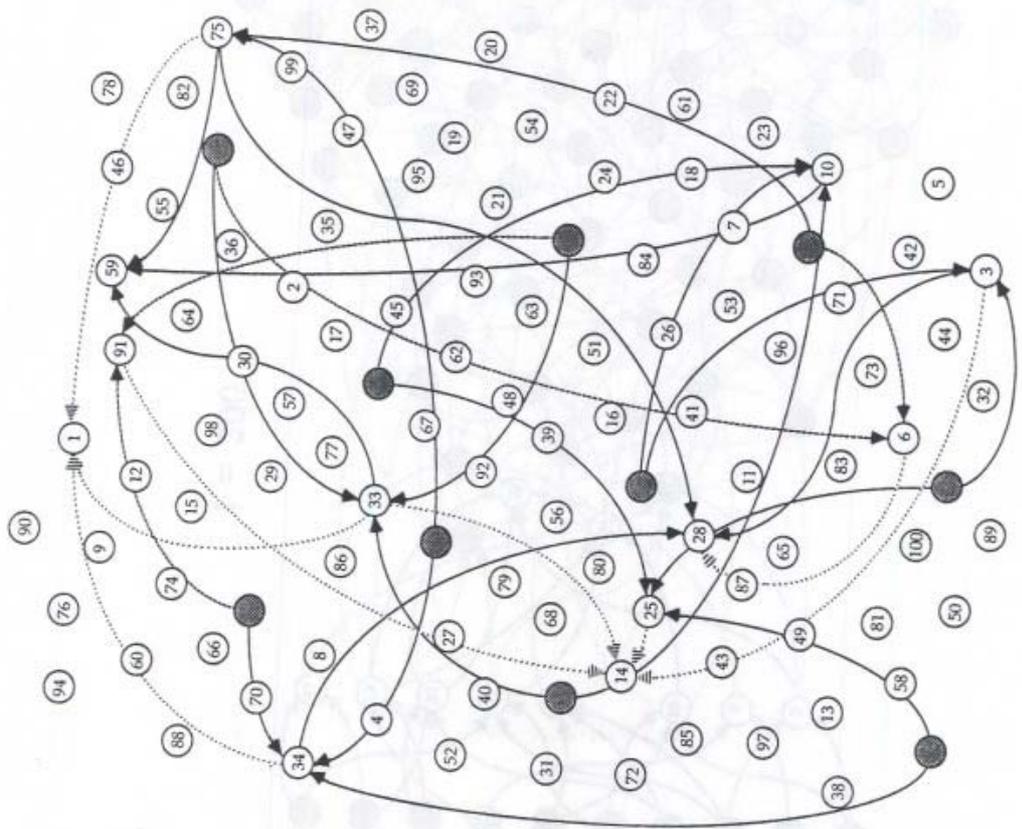
(a)



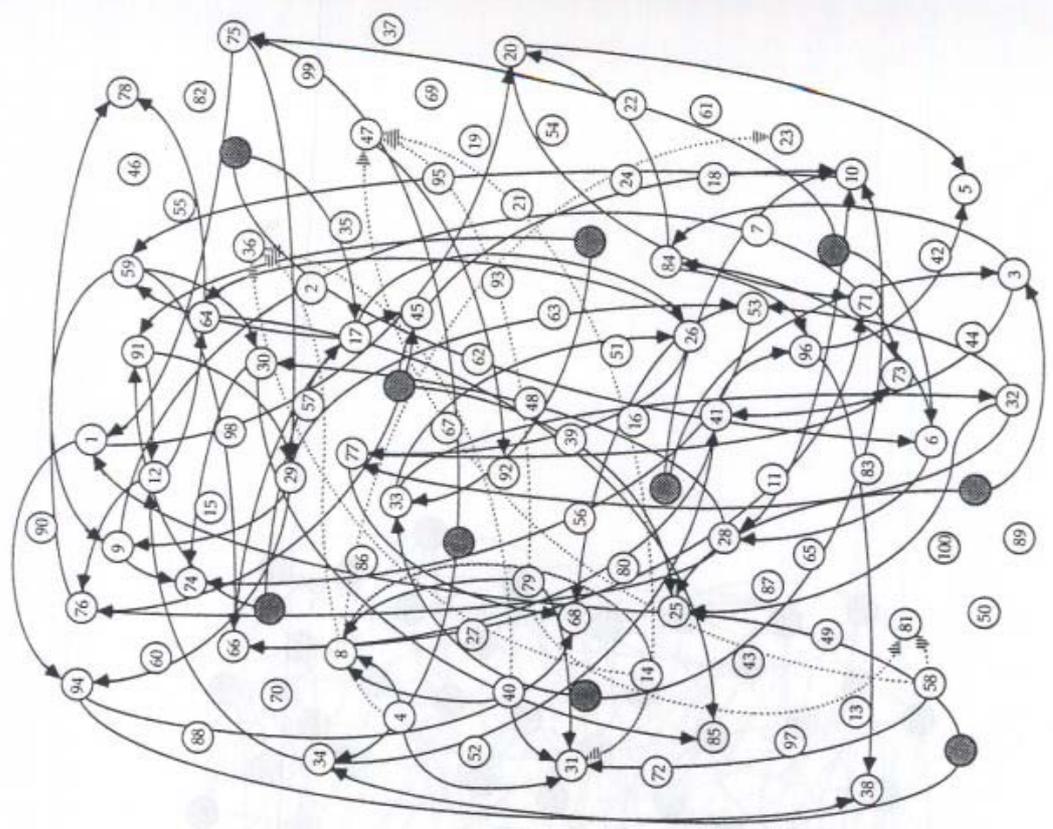
$t = 0$

(b)

(figure 19)



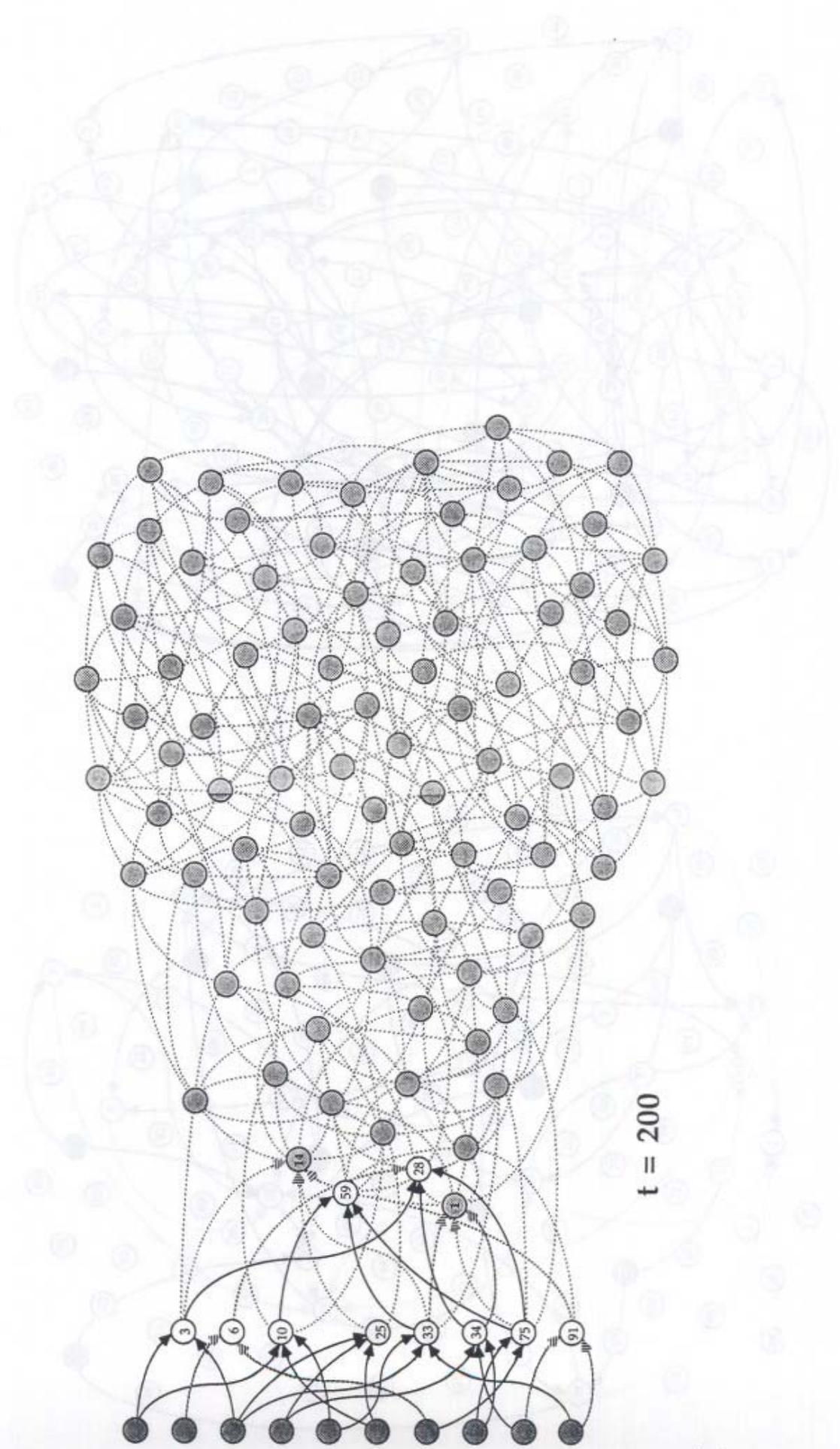
t = 200
(c)



t = 4000
(d)

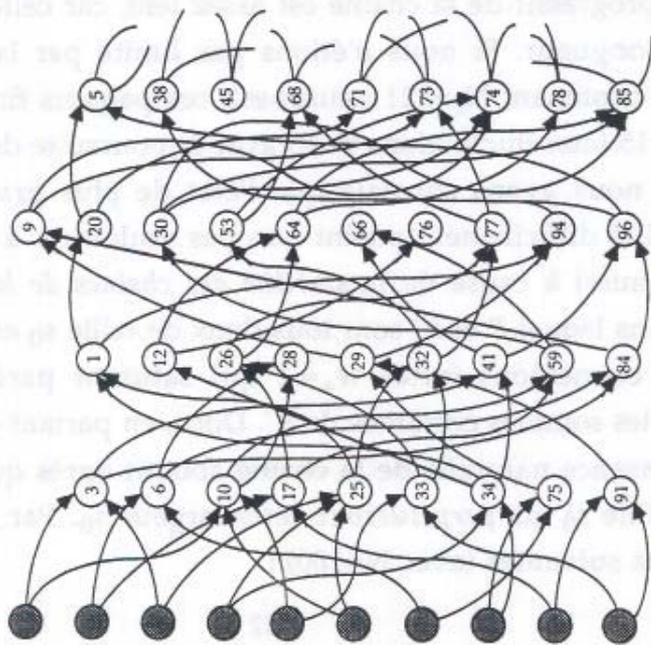
(v)
 $t = 1000$

(vi)
 $t = 500$



$t = 200$

(figure 20)



7 11 19 39 69 86 89 100

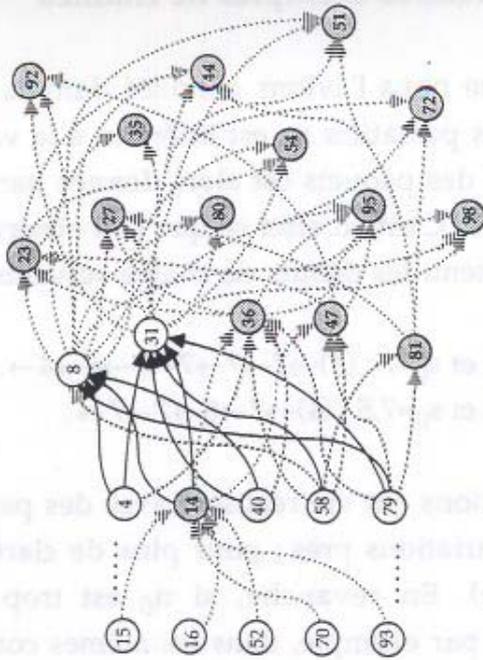
13 18 21 43 46 60 62 88

24 42 50 55 61 87 97 99

2 37 48 56 57 65 82

22 49 63 67 83 90

$t = 4000$



5.1.c Autres exemples de chaînes

Comme nous l'avions annoncé dans la discussion du §4.5, lorsque le nombre de neurones pulsatiles n_0 est inférieur à la valeur de conservation des sommes s_0 , la taille-limite des paquets est alors donnée par s_0 , et les connexions fortes valent dans ce cas $w_+ \approx 1$. C'est en effet ce que l'on observe dans les simulations : nous avons par exemple obtenu les débuts de chaîne suivants (pour $N=100$) :

- $n_0=5$ et $s_0=7$: $(5) \rightarrow 7 \rightarrow 7 \rightarrow 7 \rightarrow 7 \rightarrow 6 \rightarrow 4 \rightarrow \dots$
- $n_0=4$ et $s_0=7,5$: $(4) \rightarrow 7 \rightarrow 8 \rightarrow 7 \rightarrow 7 \rightarrow \dots$

(ces simulations ont été réalisées avec des paramètres semblables à ceux du §5.1.b, à quelques variations près : pour plus de clarté, nous ne les précisons pas dans ce paragraphe). En revanche, si n_0 est trop petit, les chaînes ne peuvent pas se construire : par exemple, dans les mêmes conditions paramétriques, mais avec $n_0=3$ et $s_0=7$, nous n'avons obtenu aucune structure de chaîne. Comme on l'a montré au §4.5, il existe en effet une "masse critique" pour les germes susceptibles de provoquer la cristallisation de la chaîne, et, plus précisément, le nombre de neurones pulsatiles n_0 doit être supérieur au seuil moyen $\Theta_0=3$ (la raison étant que les neurones pulsatiles doivent être capables de provoquer des PPS $n_0 w_+$ suffisamment forts dans les neurones de P_1 , c'est à dire supérieurs au seuil).

Au contraire, lorsque n_0 est plus grand que s_0 , la taille du premier paquet, et donc celle des paquets suivants, est en principe déterminée par n_0 (et dans ce cas, les connexions fortes sont inférieures à 1). Par exemple, les simulations ont donné :

- $n_0=15$ et $s_0=10$: $(15) \rightarrow 14 \rightarrow 13 \rightarrow 12 \rightarrow 11 \rightarrow 10 \rightarrow 9 \rightarrow 8 \rightarrow 6 \rightarrow 7 \rightarrow 7 \rightarrow 5 \rightarrow 4 \rightarrow \dots$

(avec $N=150$). Comme on le voit, ce profil de chaîne décroît assez régulièrement, et la taille des quatre premiers paquets a déjà largement dépassé $s_0=10$: l'épaississement progressif de la chaîne est assez lent, car celle-ci cherche avant tout à s'agrandir en longueur. Si nous n'étions pas limité par la taille du réseau (ici, $N=150$, la chaîne contenant $N_c=121$ neurones), ces paquets finiraient théoriquement par atteindre $n_0=15$ (aux fluctuations près) avec la poursuite des itérations.

Cependant, nous avons constaté que l'état de plus grande largeur prévu *en théorie* est en réalité difficilement atteint non pas seulement à cause de la taille finie du réseau, mais aussi à cause de la *stabilité des chaînes de largeur s_0* : en effet, un maillon $P \rightarrow P'$ dans lequel P et P' sont tous deux de taille s_0 est très stable, puisqu'il est composé de connexions fortes $w_+ \approx 1$ qui saturent parfaitement les sommes sortantes de P et les sommes entrantes de P' . Donc, en partant d'un paquet Q de taille $n_0 > s_0$, la décroissance naturelle de la chaîne aboutit après quelques maillons à des paquets P_k de taille s_0 qui *perpétueront* cette largeur s_0 . Par exemple, nous avons obtenu les chaînes suivantes (avec $N=100$) :

- $n_0=15$ et $s_0=7$: $(15) \rightarrow 12 \rightarrow 10 \rightarrow 8 \rightarrow 7 \rightarrow 7 \rightarrow 7 \rightarrow 7 \rightarrow 6 \rightarrow 5 \rightarrow 2 \rightarrow \dots$
- $n_0=12$ et $s_0=8$: $(12) \rightarrow 11 \rightarrow 10 \rightarrow 9 \rightarrow 8 \rightarrow 8 \rightarrow 8 \rightarrow 8 \rightarrow \dots$

Bien sûr, si on suppose que la conservation des sommes est respectée, ceci voudrait dire que les efférences des premiers paquets ne sont pas saturées (voir la figure 10), puisque ceux-ci sont plus grands que s_0 : donc ils devraient finir par recruter des neurones supplémentaires en aval, ce qui ferait augmenter la taille du paquet voisin, et ainsi de suite ("propagation de l'élargissement" discutée au §4.5.d). En réalité, le problème est que les sommes ne sont justement pas respectées sur ces paquets (pénalité quadratique non-nulle) à cause d'un γ relativement trop faible : ainsi, des valeurs w_+ plus fortes que prévuaturent prématurément les sommes sortantes du début de chaîne (et créent des sommes entrantes plus grandes que s_0). La "propagation de l'élargissement" n'aura donc pas lieu, et la stabilité du tronçon de largeur s_0 n'est pas remise en cause (l'une des plus longues chaînes que nous ayons obtenue dans ces conditions contenait 24 paquets (après 50000 itérations), dont la largeur fluctuait entre 8 et 9, dans un réseau de $N=400$ neurones, avec $n_0=10$ et $s_0=8$).

Enfin, les deux exemples suivants prouvent qu'il est possible de faire croître des chaînes sous une connectivité "diluée", c'est à dire avec un coefficient $p < 1$ ($N=300$) :

- $n_0=10, s_0=8$ et $p=0,5$: $(10) \rightarrow 14 \rightarrow 13 \rightarrow 13 \rightarrow 13 \rightarrow 11 \rightarrow 5 \rightarrow \dots$
- $n_0=10, s_0=8$ et $p=0,8$: $(10) \rightarrow 9 \rightarrow 8 \rightarrow 9 \rightarrow 9 \rightarrow 8 \rightarrow 8 \rightarrow 4 \rightarrow \dots$

Dans le premier cas, on remarque que la taille des paquets est nettement supérieure à n_0 ou s_0 : en effet, avec $p=0,5$, il faudrait théoriquement un paquet P_1 de taille 16 pour que les sommes sortantes du paquet pulsatile Q soient égales en moyenne à 8; puis le paquet P_2 devrait aussi contenir 16 neurones, et ainsi de suite (en pratique, on voit que les paquets n'ont pas encore atteint cette largeur limite). Le raisonnement est le même avec $p=0,8$: pour $s_0=8$, les paquets devraient être de largeur 10.

5.2 Neurones inhibiteurs et délais non-uniformes

Dans ce paragraphe, nous présentons une expérience réalisée dans les conditions complètes du modèle décrites au chapitre 3 (que nous avons jusqu'à présent simplifiées, au §4 et au §5.1, afin de mettre clairement en évidence les mécanismes de fonctionnement) : le réseau contient maintenant des neurones inhibiteurs ($N_I > 0$) et les délais τ_{ij} des connexions entre excitateurs ne sont pas uniformes.

Pour cette expérience, les paramètres seront les suivants (cf. notations du §3.1) : $N_E=200$, $N_I=40$ (nombres de neurones); $p_{EE}=1$, $p_{EI}=0,1$, $p_{IE}=0,1$ (degrés de connectivité); $1 \leq \tau_{ij}^{EE} \leq 6$ et $\tau_{ij}^{EI} = \tau_{ij}^{IE} = 1$ (délais); $\Theta_{0E}=3$, $\Theta_{0I}=2$ (seuils); $P_{0E}=5$, $P_{0I}=0$ (périodes réfractaires); $\alpha=0,1$, $\beta=0,1$, $\gamma=0,005$, $s_0=10$, $T=0,5$. Par ailleurs, les valeurs initiales des connexions $E \times I$ seront tirées aléatoirement dans l'intervalle $[0; 2]$, tandis que les connexions $I \times E$ vaudront uniformément -5 : on rappelle que ces connexions sont constantes, et que la potentiation ne concerne que les connexions $E \times E$. Les neurones pulsatiles, en nombre $n_0=20$, seront régulièrement activés toutes les 20 itérations (il s'agit de neurones excitateurs, mais ceux-ci ne contacteront pas les inhibiteurs). Avec ces paramètres, on voit qu'un neurone inhibiteur est innervé en moyenne par 20 neurones excitateurs, lesquels sont innervés en retour par 4 neurones inhibiteurs. Les délais des connexions $E \times I$ et $I \times E$ sont tous égaux à 1, ce qui signifie qu'un neurone inhibiteur réagira relativement rapidement à l'augmentation de l'activité moyenne des excitateurs (sans période réfractaire), et provoquera un instant après la désactivation de quatre d'entre eux à travers des connexions négatives très fortes.

La figure 21 montre l'évolution de l'activité totale de E et I : on constate que, parallèlement à la croissance de la chaîne des neurones excitateurs, apparaît une "contre-chaîne" inhibitrice. Il ne s'agit pas d'une focalisation entre neurones inhibiteurs, puisque leurs connexions sont constantes, mais de la répétition d'une même succession d'activations provoquée par la propagation des ondes d'activation de la chaîne des excitateurs. Chaque paquet P_k de la chaîne excitatrice innerve un certain nombre d'inhibiteurs qui eux-mêmes innervent d'autres excitateurs (donc l'activation de P_k aura pour conséquence indirecte l'inhibition de ces excitateurs, 2 instants plus tard). La figure 22 présente le profil de la chaîne obtenu après 10000 itérations : contrairement aux chaînes du §5.1, celui-ci est très irrégulier, le paquet le plus large étant situé en sixième position après le groupe pulsatile.

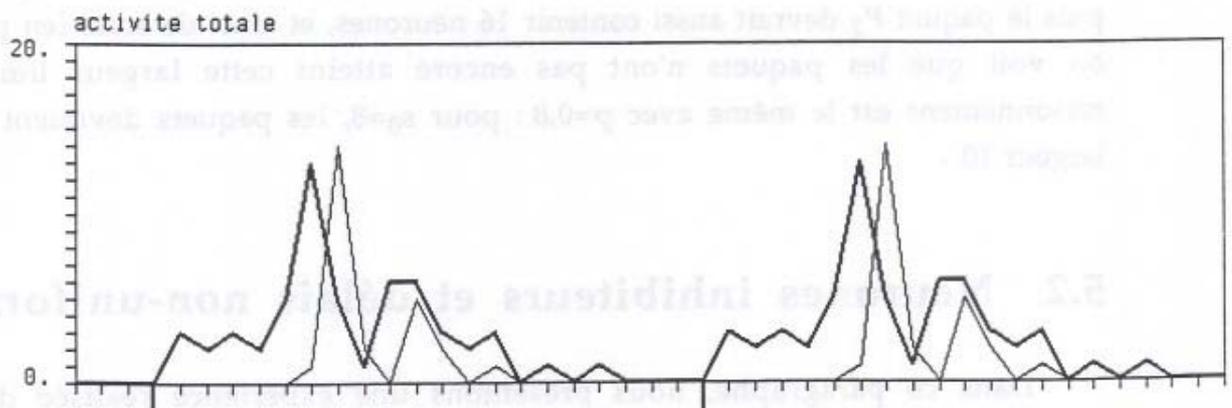
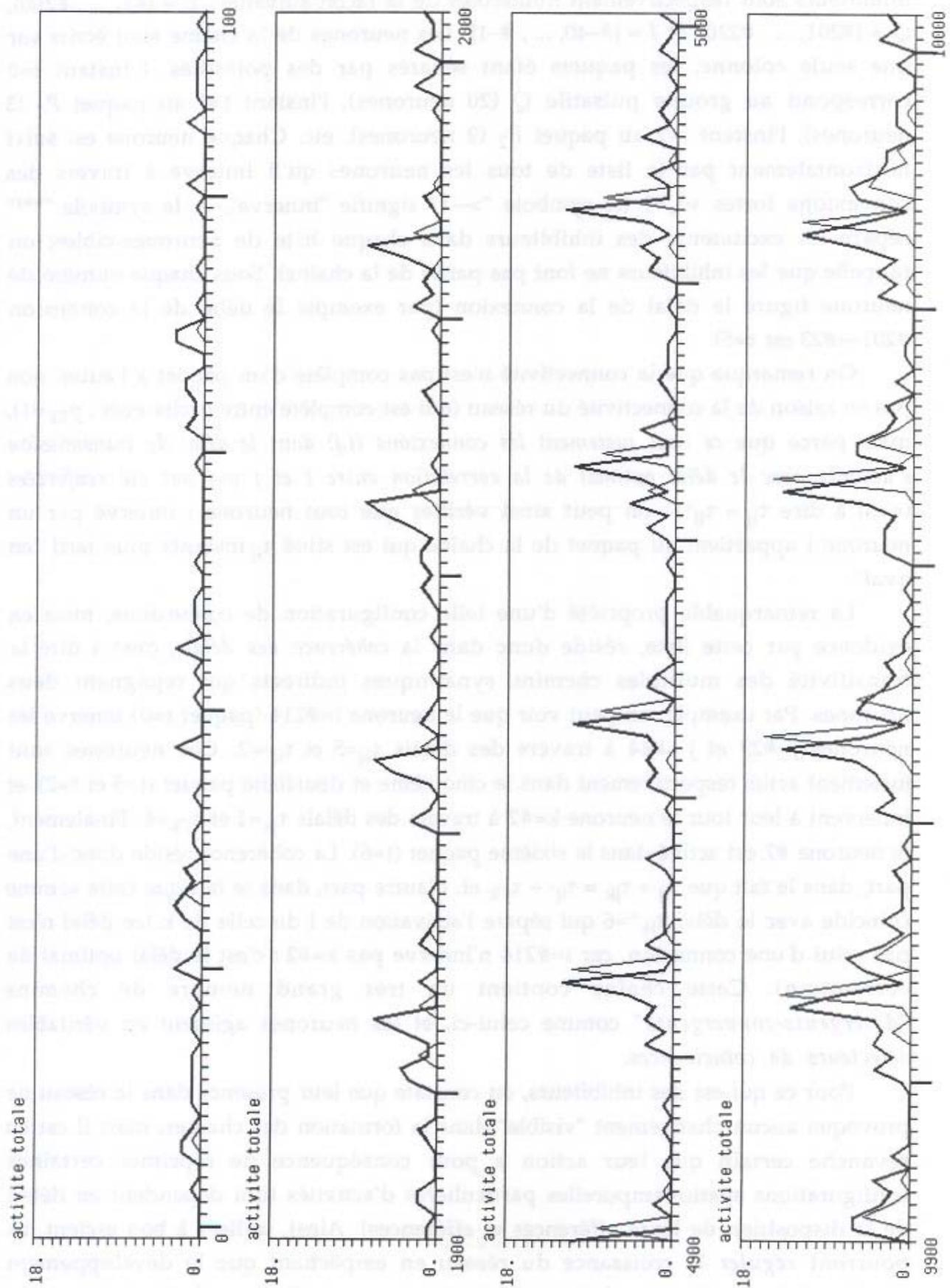


Figure 22 : Profil de la chaîne obtenu à température nulle. Les tailles des paquets (excitateurs) sont : (20), 3, 2, 3, 2, 5, 13, 5, 1, 6, 6, 3, 2, 3, 0, 1, 0, 1. L'irrégularité est due aux délais non-uniformes (voir texte).

Figure 21 : Evolution de l'activité totale : formation d'une chaîne. Sur le même graphique, on a représenté l'activité du réseau des excitateurs E (en trait fort) et l'activité du réseau des inhibiteurs I (en trait fin). On voit clairement la réaction des inhibiteurs aux fortes activités excitatrices (les pics d'activité de E sont suivis un instant plus tard par des pics d'activité sur I).



(figure 21)

Dans la table 1 est inscrite la liste de tous les neurones participant à la chaîne ainsi que toutes leurs connexions sortantes (connexions fortes internes à la chaîne), dont on montre également les délais. Les neurones excitateurs, pulsatiles et inhibiteurs sont respectivement numérotés de la façon suivante : $E = \{\#1, \dots, \#200\}$, $Q = \{\#201, \dots, \#220\}$ et $I = \{\#-40, \dots, \#-1\}$. Les neurones de la chaîne sont écrits sur une seule colonne, les paquets étant séparés par des pointillés. L'instant $t=0$ correspond au groupe pulsatile Q (20 neurones), l'instant $t=1$ au paquet P_1 (3 neurones), l'instant $t=2$ au paquet P_2 (2 neurones), etc. Chaque neurone est suivi horizontalement par la liste de tous les neurones qu'il innerve à travers des connexions fortes $w_+ \approx 1$ (le symbole ">-|" signifie "innerve", et le symbole "****" sépare les excitateurs des inhibiteurs dans chaque liste de neurones-cibles; on rappelle que les inhibiteurs ne font pas partie de la chaîne). Sous chaque numéro de neurone figure le délai de la connexion (par exemple le délai de la connexion $\#201 \rightarrow \#23$ est $\tau=5$).

On remarque que la connectivité n'est pas complète d'un paquet à l'autre, non pas en raison de la connectivité du réseau (qui est complète entre excitateurs : $p_{EE}=1$), mais parce que *ce sont justement les connexions (i,j) dont le délai de transmission s'accorde avec le délai optimal de la corrélation entre i et j qui ont été renforcées* (c'est à dire $\tau_{ij} = \tau_{ij}^*$) : on peut ainsi vérifier que tout neurone j innervé par un neurone i appartient au paquet de la chaîne qui est situé τ_{ij} instants plus tard "en aval".

La remarquable propriété d'une telle configuration de connexions, mise en évidence par cette liste, réside donc dans la *cohérence des délais*, c'est à dire la transitivité des multiples chemins synaptiques indirects qui rejoignent deux neurones. Par exemple, on peut voir que le neurone $i=\#214$ (paquet $t=0$) innerve les neurones $j=\#23$ et $j'=\#44$ à travers des délais $\tau_{ij}=5$ et $\tau_{ij'}=2$. Ces neurones sont justement actifs respectivement dans le cinquième et deuxième paquet ($t=5$ et $t=2$), et innervent à leur tour le neurone $k=\#2$ à travers des délais $\tau_{jk}=1$ et $\tau_{j'k}=4$. Finalement, le neurone $\#2$ est activé dans le sixième paquet ($t=6$). La cohérence réside donc d'une part, dans le fait que $\tau_{ij} + \tau_{jk} = \tau_{ij'} + \tau_{j'k}$ et, d'autre part, dans le fait que cette somme coïncide avec le délai $\tau_{ik}^*=6$ qui sépare l'activation de i de celle de k (ce délai n'est pas celui d'une connexion, car $i=\#214$ n'innerve pas $k=\#2$: c'est le délai optimal de corrélation). Cette chaîne contient un très grand nombre de chemins "divergents-convergeants" comme celui-ci, et les neurones agissent en véritables *détecteurs de coïncidences*.

Pour ce qui est des inhibiteurs, on constate que leur présence dans le réseau ne provoque aucun changement "visible" dans la formation des chaînes, mais il est en revanche certain que leur action a pour conséquence de réprimer certaines configurations spatio-temporelles particulières d'activités (qui dépendent en détail de la disposition de leurs afférences et efférences). Ainsi, utilisés à bon escient, ils pourront *réguler* la croissance du réseau en empêchant que le développement simultané de nombreuses chaînes n'aboutisse à une "explosion" de l'activité.

Table 1 : Liste exhaustive de la chaine et des connexions, avec delais de transmission.

t=0																	
201 >--	23	37	61	116	128	158	160	164	166	179							
	5	6	4	6	5	4	6	1	6	1							
202 >--	20	23	73	122	158	161	199										
	1	5	6	2	4	6	3										
203 >--	2	20	23	114	128	145	176										
	6	1	5	6	5	5	3										
204 >--	15	60	73	158	161	179	194	199									
	6	6	6	4	6	1	3	3									
205 >--	73	79	110	132	160	166	176	179									
	6	5	6	5	6	6	3	1									
206 >--	15	31	110	122	132	176	194	199									
	6	6	6	2	5	3	3	3									
207 >--	2	23	31	37	122	123	132	160	164	176							
	6	5	6	6	2	6	5	6	1	3							
208 >--	15	20	23	44	73	128	160	161	164	166	199						
	6	1	5	2	6	5	6	6	1	6	3						
209 >--	23	37	60	110	199												
	5	6	6	6	3												
210 >--	2	20	31	122	132	158	161	179									
	6	1	6	2	5	4	6	1									
211 >--	2	15	20	23	44	145	160	161	164	166	194						
	6	6	1	5	2	5	6	6	1	6	3						
212 >--	31	37	61	79	116	128	160	166	199								
	6	6	4	5	6	5	6	6	3								
213 >--	15	20	44	61	123	145	158	161									
	6	1	2	4	6	5	4	6									
214 >--	23	31	37	44	114	116	122										
	5	6	6	2	6	6	2										
215 >--	2	61	79	123	132	161	166	176									
	6	4	5	6	5	6	6	3									
216 >--	37	73	110	160	161	176											
	6	6	6	6	6	3											
217 >--	15	31	61	114	158	164	176	194	199								
	6	6	4	6	4	1	3	3	3								
218 >--	20	44	60	116	122	128	145	166	179								
	1	2	6	6	2	5	5	6	1								
219 >--	31	44	79	122	123	158	176										
	6	2	5	2	6	4	3										
220 >--	15	20	37	44	73	116	123	145	161	194							
	6	1	6	2	6	6	6	5	6	3							
t=1																	
20 >--	44	114	122	123	128	158	***	-38	-36	-27	-4	-3					
	1	5	1	5	4	3	***	1	1	1	1	1					
164 >--	2	114	116	176	***	-36	-35	-26	-23	-5							
	5	5	5	2	***	1	1	1	1	1							
179 >--	27	37	44	61	79	110	114	125	145	166	194	***	-33	-32	-14	-6	-1
	6	5	1	3	4	5	5	6	4	5	2	***	1	1	1	1	1
t=2																	
44 >--	2	7	23	31	37	79	110	161	189	***	-37	-25	-20	-11	-4		
	4	5	3	4	4	3	4	4	5	***	1	1	1	1	1		
122 >--	2	15	27	110	123	189	***	-26	-16	-14							
	4	4	5	4	4	5	***	1	1	1							
t=3																	
176 >--	24	27	31	35	60	103	145	167	189	***	-25	-3					
	6	4	3	6	3	5	2	6	4	***	1	1					
194 >--	15	21	23	60	116	145	165	***	-31	-30	-26	-24	-21				
	3	6	2	3	3	2	6	***	1	1	1	1	1				
199 >--	7	21	116	160	166	189	***	-23	-20								
	4	6	3	3	3	4	***	1	1								
t=4																	
61 >--	15	35	69	79	116	123	189	***	-28	-11	-8						
	2	5	6	1	2	2	3	***	1	1	1						
158 >--	7	24	55	69	103	114	123	125	132	167	***	-34	-1				
	3	5	6	6	4	2	2	3	1	5	***	1	1				
t=5																	
23 >--	2	7	13	43	55	69	107	116	123	166							
	1	2	5	5	5	5	6	1	1	1							
79 >--	31	43	60	77	110	154	160	165	***	-34	-21	-5					
	1	5	1	4	1	5	1	4	***	1	1	1					
128 >--	7	18	21	69	73	103	116	161	167	***	-26	-10	-2				
	2	2	4	5	1	3	1	1	4	***	1	1	1				
132 >--	7	15	21	27	31	35	125	165	167	***	-37	-25	-19	-8	-5		
	2	1	4	2	1	4	2	4	4	***	1	1	1	1	1		

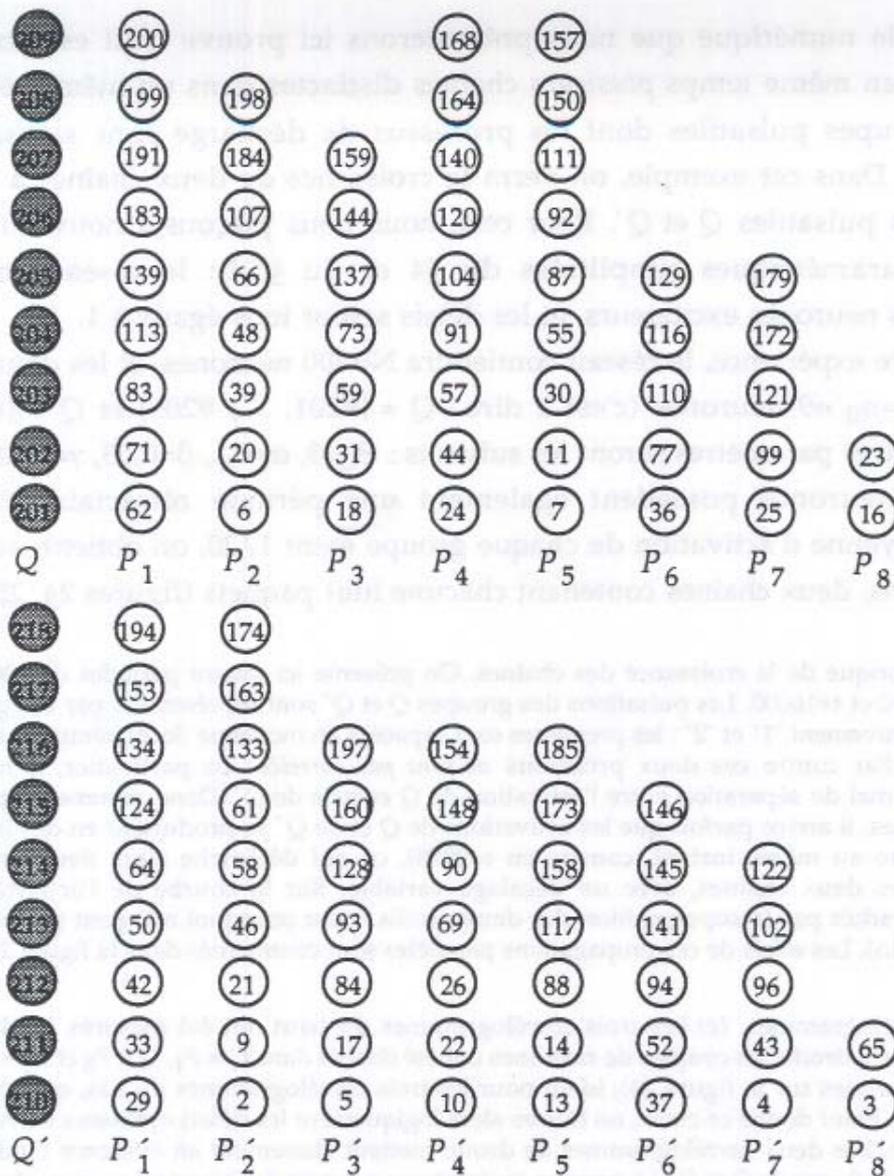
6 COMPOSITION ET CONFIGURATIONS COMPLEXES

Nous avons montré tout au long des chapitres 4 et 5 que les "mécanismes de développement" très généraux exposés au chapitre 3 peuvent, dans des conditions appropriées, assurer la croissance de configurations spatio-temporelles (CST) possédant une topologie linéaire "en chaîne". Ces chaînes sont définies par une *structure temporelle* précise, caractérisée par de nombreuses synchronisations et corrélations entre activités neuronales, cette structure temporelle étant soutenue par une *structure spatiale* particulière de connexions (focalisation). Dans ce chapitre, nous aborderons le mécanisme de *composition dynamique* des chaînes en commençant par présenter un exemple numérique (§6.1), puis en généralisant ces résultats à un cadre moins schématique que celui des simulations (CST à délais "réels", §6.2). Enfin, nous mettrons en perspective la croissance et la composition des CST dans un vaste processus de structuration du cortex, où elles interagiront de différentes manières pour produire des configurations *complexes* (§6.3).

6.1 Composition dynamique de deux chaînes

Dans ce paragraphe, on exposera le mécanisme de *composition des chaînes* à travers un exemple numérique. Comme nous l'avons signalé au chapitre III.1, nous ferons l'hypothèse que la composition dynamique des symboles est assurée par des "synapses de von der Malsburg", c'est à dire une forme de plasticité synaptique hebbienne *rapide et réversible*. En résumé, dans la théorie qu'il propose, von der Malsburg (1981) prévoit deux variables pour décrire un poids synaptique W_{ij} : l'une, s_{ij} , sert de base à une "structure permanente" du réseau, et ses variations s'effectuent sur une échelle de temps lente (mémoire à long terme); l'autre, a_{ij} , est une variable dynamique, qui change sur une échelle de temps rapide (elle est appelée "état d'activité" de la synapse (i,j)). Lorsque les signaux entre i et j sont fortement corrélés, la synapse est "activée" par une augmentation de a_{ij} , et lorsqu'ils sont à nouveau décorrélés, a_{ij} diminue et la synapse est "inactivée" (lorsqu'aucun signal n'est transmis de i à j, a_{ij} se stabilise à une "valeur de repos" a_0). Il s'agit donc d'un phénomène de modulation rapide, dont une formulation simple pourrait être, par exemple : $W_{ij} = a_{ij} s_{ij}$ avec $0 \leq a_{ij} \leq 1$.

Nous décrirons tout d'abord une simulation numérique ayant abouti à la croissance simultanée de deux chaînes indépendantes, puis nous verrons comment ces chaînes, considérées comme des "morceaux élémentaires de symboles", peuvent être *composées*, c'est à dire liées entre elles par des synapses rapides, grâce à la synchronisation de leurs ondes d'activation.



(a)

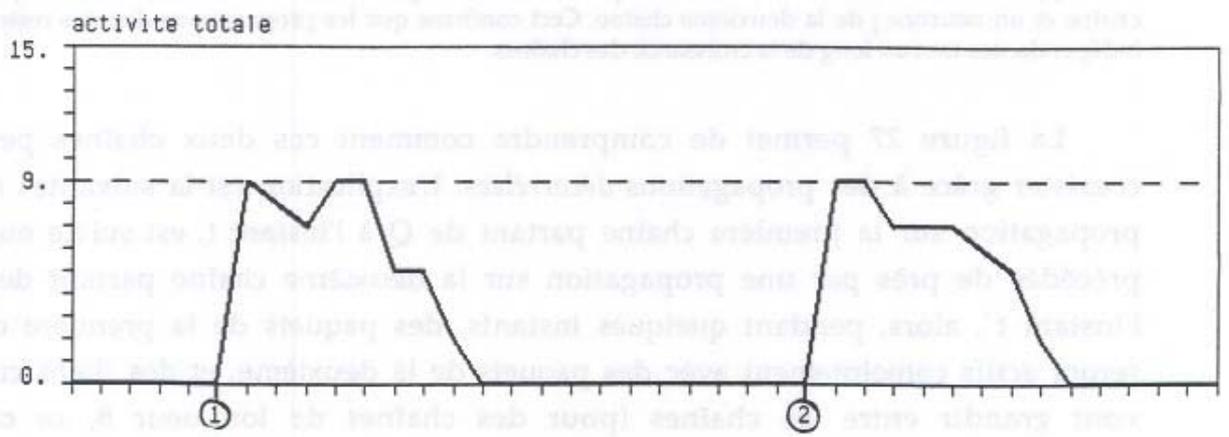


Figure 24 : Croissance simultanée de deux chaînes indépendantes. (a) A l'instant $t=16000$, les chaînes sont toutes deux de longueur 8 : la première contient en tout 54 neurones, et la deuxième 52 neurones (donc au total, environ la moitié du réseau s'est structurée). (b) Leurs profils sont également visibles par propagation d'une onde non-bruitée. La période de 40 itérations montrée ici a été extraite au-delà de $t=16000$, et contient une pulsation de chaque groupe : on voit que l'activation de Q (graduation étiquetée "1") fait apparaître une chaîne de profil (9), 9, 8, 7, 9, 9, 5, 5, 2, tandis que l'activation de Q' (graduation "2"), fait apparaître une chaîne de profil (9), 9, 9, 7, 7, 7, 6, 5, 2.

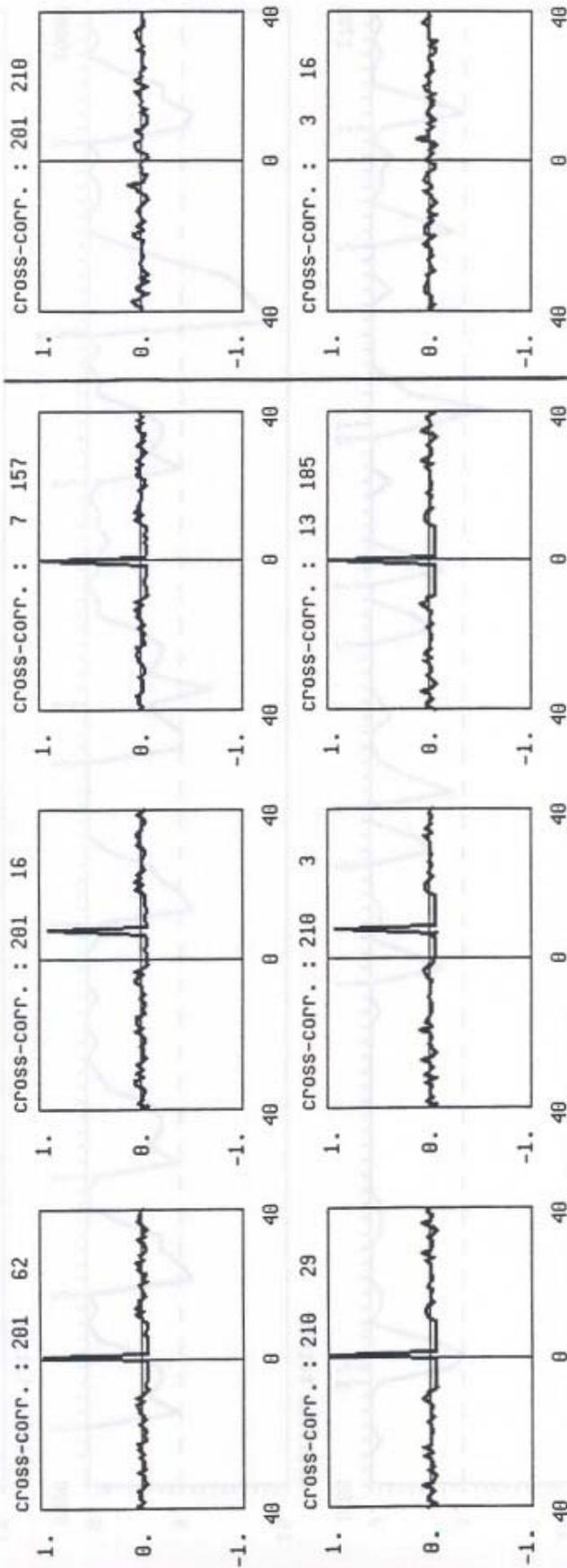
L'exemple numérique que nous présenterons ici prouve qu'il est possible de faire grandir en même temps *plusieurs* chaînes distinctes dans un même réseau R , à partir de groupes pulsatiles dont les processus de décharge sont *statistiquement indépendants*. Dans cet exemple, on verra la croissance de deux chaînes à partir de deux groupes pulsatiles Q et Q' . Pour cela, nous nous plaçons à nouveau dans les conditions paramétriques simplifiées du §4 et du §5.1 : le réseau contiendra seulement des neurones excitateurs, et les délais seront tous égaux à 1.

Pour cette expérience, le réseau contiendra $N=200$ neurones, et les deux groupes pulsatiles $n_0=n_0'=9$ neurones (c'est à dire : $Q = \{\#201, \dots, \#209\}$ et $Q' = \{\#210, \dots, \#218\}$). Les autres paramètres seront les suivants : $\Theta_0=3$, $\alpha=0,1$, $\beta=0,05$, $\gamma=0,005$, $T=0,45$ et $s_0=9$ (les neurones possèdent également une période réfractaire $P_0=2$). La fréquence moyenne d'activation de chaque groupe étant $1/20$, on obtient, au bout de 16000 itérations, deux chaînes contenant chacune huit paquets (figures 24, 25 et 26).

Figure 25 : Historique de la croissance des chaînes. On présente ici quatre périodes de 100 itérations extraites entre $t=0$ et $t=16000$. Les pulsations des groupes Q et Q' sont représentées par des graduations étiquetées respectivement "1" et "2" : les premières sont espacées en moyenne de 20 itérations tout comme les deuxièmes. Par contre ces deux processus *ne sont pas corrélés* : en particulier, il n'existe pas d'intervalle minimal de séparation entre l'activation de Q et celle de Q' . Donc, comme on peut le voir sur ces graphiques, il arrive parfois que les activations de Q et de Q' se produisent en des instants très rapprochés (voire au même instant, comme en $t=9978$), ce qui déclenche alors deux propagations parallèles sur les deux chaînes, avec un décalage variable. Sur la courbe de l'activité totale, ce phénomène se traduit par la superposition des deux profils : c'est pourquoi $n(t)$ peut parfois atteindre $9+9=18$ (vers la fin). Les effets de ces propagations parallèles sont commentés dans la figure 27.

Figure 26 : Corrélogrammes. (a) Les trois corrélogrammes du haut ont été mesurés sur la première chaîne : de gauche à droite, les couples de neurones ont été choisis dans $Q \times P_1$, $Q \times P_8$ et $P_5 \times P_5$ (voir les numéros des neurones sur la figure 24); idem pour les trois corrélogrammes du bas, qui concernent la deuxième chaîne. Etant donné ce choix, on trouve alors logiquement les délais optimaux suivants : $\tau_{ij}^*=1$, $\tau_{ij}^*=8$ et $\tau_{ij}^*=0$. (b) Les deux corrélogrammes de droite mettent clairement en évidence l'indépendance des processus de décharge de Q et Q' . Le premier confronte un neurone de Q avec un neurone de Q' (#201 et #210), et le deuxième un neurone de P_8 avec un neurone de P_8' (#3 et #16) : aucun délai optimal τ_{ij}^* n'est alors apparent. Bien sûr, il en est de même pour tout corrélogramme entre un neurone i de la première chaîne et un neurone j de la deuxième chaîne. Ceci confirme que les propagations d'ondes restent bien indépendantes tout au long de la croissance des chaînes.

La figure 27 permet de comprendre comment ces deux chaînes peuvent coexister grâce à des propagations *décorrélées*. L'explication est la suivante : si une propagation sur la première chaîne partant de Q à l'instant t , est suivie ou a été précédée de près par une propagation sur la deuxième chaîne partant de Q' à l'instant t' , alors, pendant quelques instants, des paquets de la première chaîne seront actifs conjointement avec des paquets de la deuxième, et des "liens croisés" vont grandir entre les chaînes (pour des chaînes de longueur 8, ce cas de "propagation parallèle" correspond donc à $|t - t'| < 8$). Plus précisément : si les paquets P_k et P_{1+k}' sont simultanément actifs (avec $1 - k = t' - t$), alors P_{k+1} et P_{1+k+1}' seront simultanément actifs à l'instant suivant, et par conséquent, toutes les connexions reliant les neurones de P_k à ceux de P_{1+k}' augmenteront de $+\alpha$, ainsi que toutes les connexions reliant P_{k+1} à P_{1+k+1}' (il s'agit ici de la plasticité lente habituelle : nous envisagerons le cas des synapses rapides un peu plus bas).



(figure 26)

(b)

(a)

Or, si le prochain événement de propagation parallèle est caractérisé par un délai relatif $t-t'$ différent du précédent, ce seront d'autres liens croisés qui augmenteront, et qui feront par conséquent rediminuer ceux qui avaient déjà commencé à s'installer, à cause de la compétition. Ainsi, les chaînes pourront grandir chacune de leur côté, sans être importunées par les propagations d'activation qui ont lieu sur l'autre chaîne, grâce au fait que le retard relatif n'est jamais le même, donc ne permet pas la stabilisation des connexions croisées.

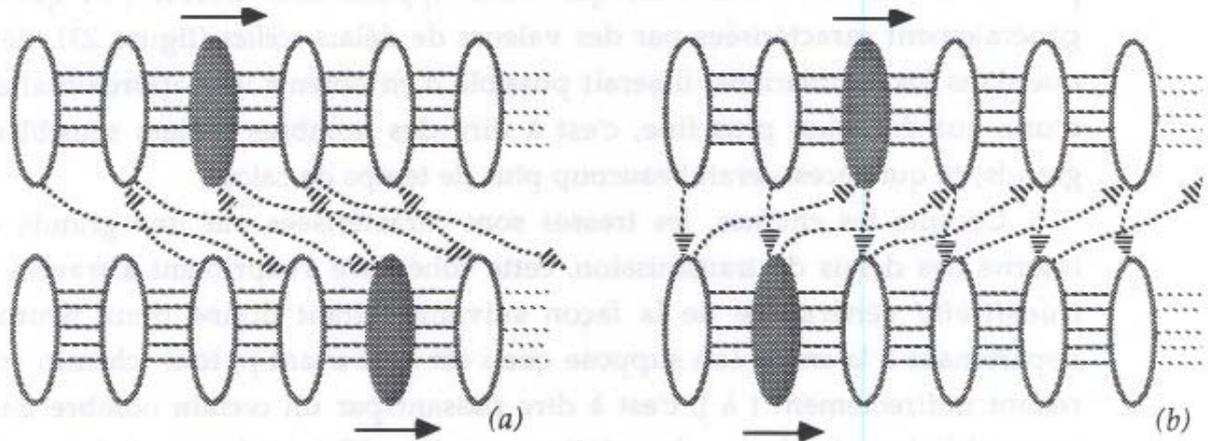


Figure 27 : Stabilité de la coexistence des deux chaînes. Comme on a pu le constater sur les graphiques de la figure 25, il arrive parfois que deux ondes d'activations se propagent "en parallèle" sur les deux chaînes, à cause de pulsations de Q et Q' rapprochées (on notera t et t' les instants d'activation de ces groupes : donc nous envisageons ici la situation où $|t-t'| < 8$). Dans ce cas, en vertu de la règle de potentiation immédiate, des "connexions croisées" vont alors se former entre les chaînes. (a) Par exemple, ici, $t'=t+2$ (le dessin montre un état instantané du réseau, dans lequel P_2 et P_4' sont actifs simultanément) : dans ce cas, toutes les connexions du type $P_k \rightarrow P_{k+3}'$ et $P_k' \rightarrow P_{k-1}$ augmenteront de $+\alpha$. (b) Dans cet autre exemple, on a $t'=t-1$, ce qui provoquera l'augmentation des connexions $P_k \rightarrow P_k'$ et $P_k' \rightarrow P_{k-2}$, au détriment des connexions du (a), qui rediminueront. En conclusion, on voit que si l'écart relatif $t-t'$ change incessamment de valeur, ces liens croisés ne seront jamais renforcés.

Par quel mécanismes ces deux chaînes peuvent-elles être alors *composées*, c'est à dire reliées par des connexions qui les rendent constituants d'un graphe plus grand ? La coexistence même de ces chaînes décorréliées nous en fournit, a contrario, la réponse : c'est précisément par la *synchronisation* de leurs propagations que ces chaînes seront assemblées. Si la différence de phase entre les ondes d'activations reste stable, les connexions croisées correspondant à ce décalage seront renforcées plus durablement : elles contribueront alors à leur tour à maintenir la stabilité du déphasage, car lorsqu'une onde se propagera sur l'une des chaînes, elle induira la propagation d'une onde sur l'autre chaîne à travers ces connexions. Il s'agit donc d'un collage "perpendiculaire" provoqué par des propagations "parallèles" (ce qui serait comparable au mécanisme d'une "fermeture éclair"). Comme on le voit, ce modèle pourrait permettre d'expliquer le mécanisme élémentaire du "liage dynamique" des symboles ("dynamic binding"), par l'utilisation de synapses rapides : en résumé, la compositionnalité des CST est réalisée à travers leur synchronisation.

6.2 Propriétés générales des CST

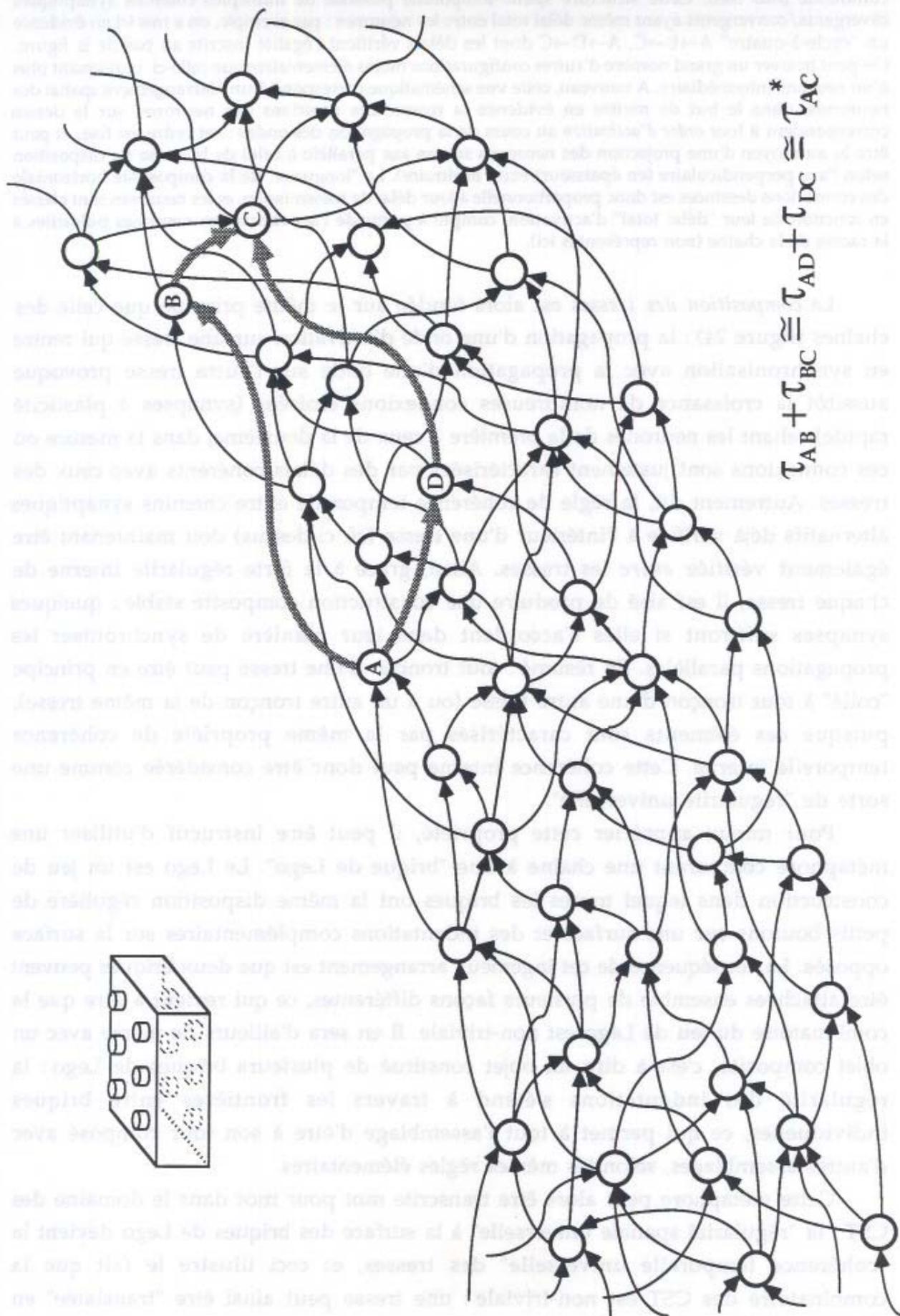
Pour les besoins des simulations numériques, nous avons jusqu'à présent discrétisé l'axe des temps et donc supposé que les délais étaient des nombres entiers (uniformes ou distribués entre deux bornes). C'est pourquoi les chaînes que nous avons décrites étaient constituées de "groupes" de neurones synchrones. Or, comme nous l'avions annoncé dans l'introduction du §4, ces chaînes ne sont que des cas particuliers de CST linéaires, que nous appellerons "tresses", et qui sont plus généralement caractérisées par des valeurs de délais *réelles* (figure 23). Remarquons que dans les simulations, il serait possible d'en obtenir une approximation à l'aide d'une numérisation plus fine, c'est à dire des nombres entiers sensiblement plus grands, ce qui nécessiterait beaucoup plus de temps de calcul.

Comme les chaînes, les tresses sont caractérisées par une grande cohérence interne des délais de transmission, cette cohérence s'exprimant à travers une *loi de transitivité* généralisée de la façon suivante : étant donné deux neurones i et j appartenant à la tresse (on suppose que i est actif avant j), tout "chemin synaptique" reliant indirectement i à j , c'est à dire passant par un certain nombre de neurones intermédiaires k_1, k_2, \dots, k_n , réalise un *même délai total* entre i et j , ce délai total correspondant au délai optimal τ_{ij}^* de la corrélation des décharges entre i et j (sans qu'il existe nécessairement une connexion (i,j)) :

$$\tau_{ik_1} + \tau_{k_1k_2} + \dots + \tau_{k_nj} = \tau_{ij}^*$$

Cependant, on ne peut pas délimiter ici des sous-ensembles P_k de neurones parfaitement synchrones comme dans le cas des délais entiers, et il est même rare que deux neurones d'une tresse soient actifs simultanément : la propagation d'une onde consiste donc en une succession temporelle d'activations individuelles séparées par des intervalles de temps irréguliers fixés. Cette succession temporelle est *stable et reproductible* : elle constitue le "motif rythmique" caractéristique de la tresse (code d'"ordre élevé" avec délais non-nuls; cf. notations du chapitre 1). Sur l'axe temporel, les neurones ne sont donc pas repérés selon leur appartenance à des groupes synchrones mais plus généralement par un "délai absolu" qui sépare leur activation de celle d'un neurone pris comme référence.

Les tresses ont également besoin d'une "inhomogénéité" pour croître, c'est à dire d'un neurone-source i_0 dont les afférences innervent fortement un petit groupe de neurones Q : cependant, dans la mesure où les délais de ces afférences ne sont désormais pas nécessairement identiques, la pulsation générique de Q ne sera pas constituée de décharges synchrones, mais de décharges corrélées séparées par des intervalles de temps irréguliers. Un neurone j recruté par Q pour faire partie du premier paquet P_1 devra donc posséder des afférences cohérentes avec les corrélations internes aux neurones $i \in Q$: par exemple, pour des seuils $\Theta_0=3$, il devra exister au moins trois chemins synaptiques $i_0 \rightarrow i \rightarrow j$ ayant même délai total.



$$\tau_{AB} + \tau_{BC} \cong \tau_{AD} + \tau_{DC} \cong \tau_{AC}^*$$

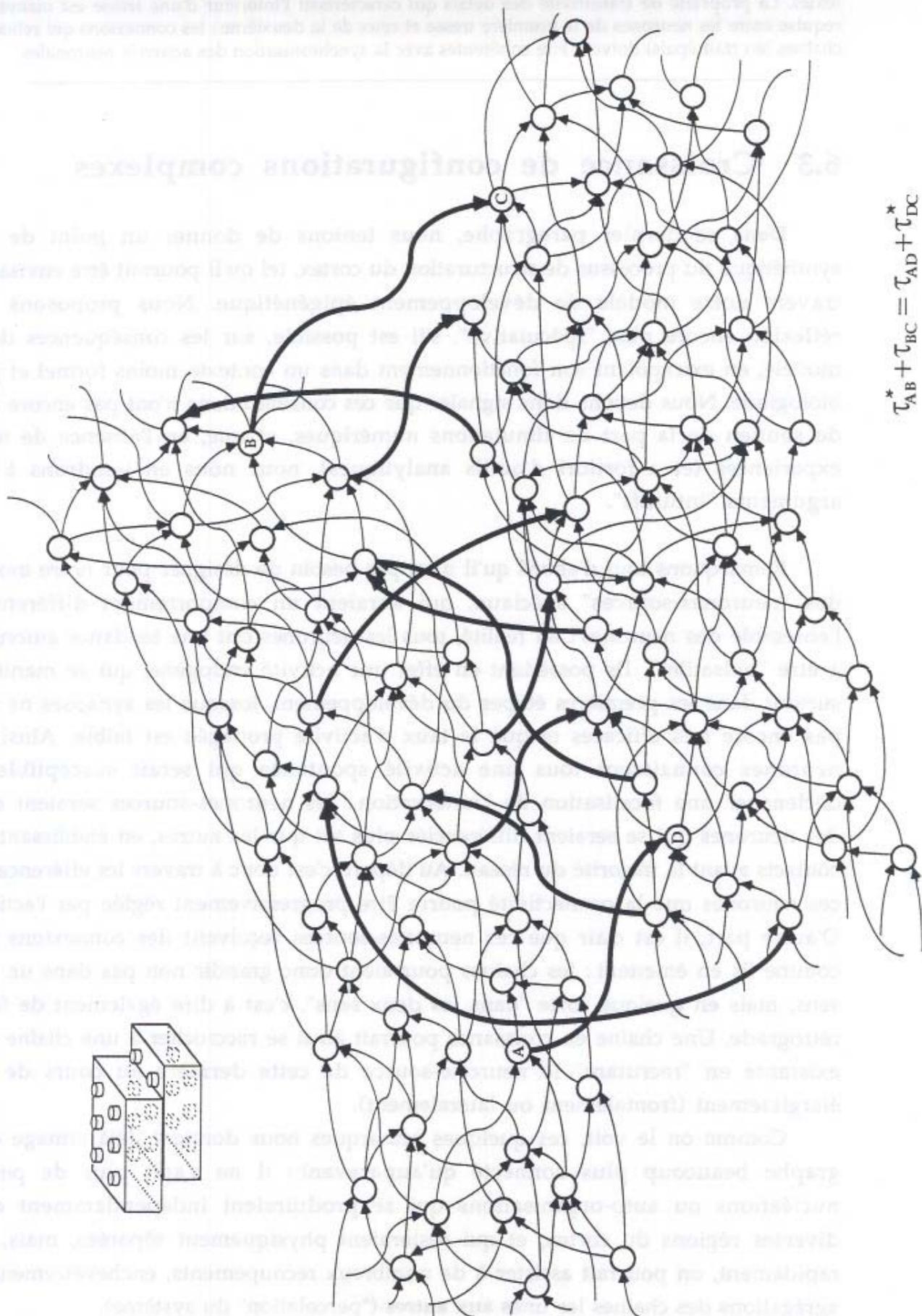
(figure 23)

Figure 23 : Illustration d'une CST linéaire ou "tresse" (le pictogramme de la brique de "Lego" sera commenté plus bas). Cette structure spatio-temporelle possède de multiples chemins synaptiques divergents/convergents ayant même délai total entre les neurones : par exemple, on a mis ici en évidence un "cycle-à-quatre" $A \rightarrow B \rightarrow C$, $A \rightarrow D \rightarrow C$ dont les délais vérifient l'égalité inscrite au bas de la figure. On peut trouver un grand nombre d'autres configurations moins élémentaires que celle-ci, impliquant plus d'un neurone intermédiaire. A nouveau, cette vue schématique correspond à un réarrangement spatial des neurones, dans le but de mettre en évidence la tresse. Les positions des neurones sur le dessin correspondent à leur *ordre d'activation* au cours de la propagation des ondes : cet ordre est fixe, et peut être lu au moyen d'une projection des neurones sur un axe parallèle à celui de la tresse (la disposition selon l'axe perpendiculaire (en épaisseur) étant arbitraire). La "longueur" de la composante horizontale des connexions dessinées est donc proportionnelle à leur délai de transmission, et les neurones sont classés en fonction de leur "délai total" d'activation, compté à partir de l'activation des neurones pulsatiles à la racine de la chaîne (non représentés ici).

La *composition des tresses* est alors fondée sur le même principe que celle des chaînes (figure 24) : la propagation d'une onde d'activation sur une tresse qui rentre en synchronisation avec la propagation d'une onde sur l'autre tresse provoque aussitôt la croissance de nombreuses connexions croisées (synapses à plasticité rapide) reliant les neurones de la première à ceux de la deuxième, dans la mesure où ces connexions sont justement caractérisées par des délais cohérents avec ceux des tresses. Autrement dit, la règle de cohérence temporelle entre chemins synaptiques alternatifs déjà vérifiée à l'intérieur d'une tresse (cf. ci-dessus) doit maintenant être également vérifiée *entre* les tresses. Ainsi, grâce à la forte régularité interne de chaque tresse, il est aisé de produire une construction composite stable : quelques synapses suffiront si elles s'accordent dans leur manière de synchroniser les propagations parallèles. En résumé : tout tronçon d'une tresse peut être en principe "collé" à tout tronçon d'une autre tresse (ou à un autre tronçon de la même tresse), puisque ces éléments sont caractérisés par la même propriété de cohérence temporelle interne. Cette cohérence interne peut donc être considérée comme une sorte de "régularité universelle".

Pour mieux apprécier cette propriété, il peut être instructif d'utiliser une métaphore comparant une chaîne à une "brique de Lego". Le Lego est un jeu de construction dans lequel toutes les briques ont la même disposition régulière de petits boutons sur une surface et des indentations complémentaires sur la surface opposée. La conséquence de cet ingénieux arrangement est que deux briques peuvent être attachées ensemble de plusieurs façons différentes, ce qui revient à dire que la combinatoire du jeu de Lego est non-triviale. Il en sera d'ailleurs de même avec un objet composite, c'est à dire un objet constitué de plusieurs briques de Lego : la régularité des indentations s'étend à travers les frontières entre briques individuelles, ce qui permet à tout l'assemblage d'être à son tour composé avec d'autres assemblages, selon les mêmes règles élémentaires.

Cette métaphore peut alors être transcrite mot pour mot dans le domaine des CST : la "régularité spatiale universelle" à la surface des briques de Lego devient la "cohérence temporelle universelle" des tresses, et ceci illustre le fait que la combinatoire des CST est non-triviale : une tresse peut ainsi être "translatée" en plusieurs endroits le long d'une autre tresse.



$$\tau_{AB}^* + \tau_{BC} = \tau_{AD} + \tau_{DC}^*$$

(figure 24)

Figure 24 : Illustration de la composition de deux tresses. (pictogramme des briques de "Lego" : voir texte). La propriété de transitivité des délais qui caractérisait l'intérieur d'une tresse est maintenant requise entre les neurones de la première tresse et ceux de la deuxième : les connexions qui relient les chaînes (en trait épais) doivent être cohérentes avec la synchronisation des activités neuronales.

6.3 Croissance de configurations complexes

Dans ce dernier paragraphe, nous tentons de donner un point de vue synthétique du processus de structuration du cortex, tel qu'il pourrait être envisagé à travers notre modèle de développement épigénétique. Nous proposons une réflexion encore plus "spéculative", s'il est possible, sur les conséquences de ce modèle, en extrapolant son fonctionnement dans un contexte moins formel et plus biologique. Nous devons donc signaler que ces considérations n'ont pas encore reçu de soutien de la part de simulations numériques, et que, en l'absence de telles expériences (et a fortiori d'outils analytiques), nous nous en tiendrons à des arguments "intuitifs".

Remarquons tout d'abord qu'il n'est pas besoin de désigner pour notre modèle des "neurones-sources" spéciaux, qui auraient un comportement différent de l'ensemble des neurones : en réalité, tous les neurones ont une tendance autonome à être "pulsatiles". Ils possèdent en effet une activité endogène, qui se manifeste surtout dans les premières étapes du développement, lorsque les synapses ne sont pas encore très efficaces et que le taux d'activité propagée est faible. Ainsi, les neurones connaissent tous une activité spontanée qui serait susceptible de déclencher une focalisation de l'innervation : les neurones-sources seraient donc des neurones qui se seraient différenciés plus tôt que les autres, en établissant des contacts avant la majorité du réseau. Au départ, c'est donc à travers les efférences de ces neurones que la connectivité pourra être progressivement réglée par l'activité. D'autre part, il est clair que ces neurones-sources reçoivent des connexions tout comme ils en émettent : les chaînes pourraient donc grandir non pas dans un seul sens, mais en quelque sorte "dans les deux sens", c'est à dire également de façon rétrograde. Une chaîne en croissance pourrait ainsi se raccrocher à une chaîne déjà existante en "recrutant" le neurone-source de cette dernière au cours de son élargissement (frontalement ou latéralement).

Comme on le voit, ces quelques remarques nous donnent déjà l'image d'un graphe beaucoup plus connecté qu'auparavant : il ne s'agit plus de petites nucléations ou auto-organisations qui se produiraient indépendamment dans diverses régions du cortex, et qui resteraient physiquement séparées, mais, très rapidement, on pourrait assister à de nombreux recoupements, enchevêtrements et agrégations des chaînes les unes aux autres ("percolation" du système).

Quelle description plus précise pourrait-on alors donner de ces interactions ?

Dans un premier temps, nous pouvons envisager la formation d'une vaste structure arborescente : comme nous venons de l'évoquer, il se produira des bifurcations, et ces bifurcations pourront être stables si elles se sont formées indépendamment les unes des autres. Il semble en effet difficile qu'une bifurcation apparaisse pendant le processus de croissance d'une chaîne (chaîne "bifide"), car il y aurait aussitôt "fermeture" des deux parties à cause de la nécessaire synchronisation des propagations. En revanche, si les différentes voies se sont organisées au départ indépendamment les unes des autres (puis se sont rejointes par des connexions latérales), elles auront ensuite la possibilité de rester séparées, surtout si l'on fait par ailleurs l'hypothèse que le niveau global d'activité est contrôlé par des systèmes inhibiteurs et ne peut donc pas dépasser un certain seuil.

Ceci conduit donc à penser que ces bifurcations, résultat de la fusion d'éléments indépendants, sont le siège d'"instabilités" qui font en sorte que l'activité ne puisse pas se propager systématiquement à la fois sur les deux branches, à cause des mécanismes inhibiteurs (car si c'était le cas, on aboutirait à un embrasement du reste du réseau). Dans ce cadre, on voit que les neurones inhibiteurs auraient un rôle très important à jouer, rôle qui était peu visible dans la construction d'une chaîne unique (comme dans l'exemple du §5.2) : ces neurones ne bouleversent pas le processus de croissance du réseau, mais ont pour effet de réprimer certaines configurations, qui, autrement, auraient pu être sélectionnées et stabilisées.

En résumé, il existe donc des contraintes fortes contre lesquelles l'activité vient se heurter (un "mur" infranchissable, qui correspond à un niveau moyen d'activité), implémentées par les cellules inhibitrices. Bien sûr, en plus des voies divergentes, dont l'"aguillage" est contrôlé par l'inhibition, il pourra aussi se produire des fusions par convergence : deux extrémités de chaînes se rejoignent et forment une chaîne unique.

Cette image relativement complexe pourrait alors se complexifier davantage en supposant que les chaînes qui se sont déjà organisées et rattachées les unes aux autres établissent ensuite des "points de contacts" entre elles : il s'agirait d'interactions relativement plus faibles réalisant un collage parallèle entre tronçons de chaînes élémentaires, en suivant le mécanisme de composition que nous avons décrit plus haut.

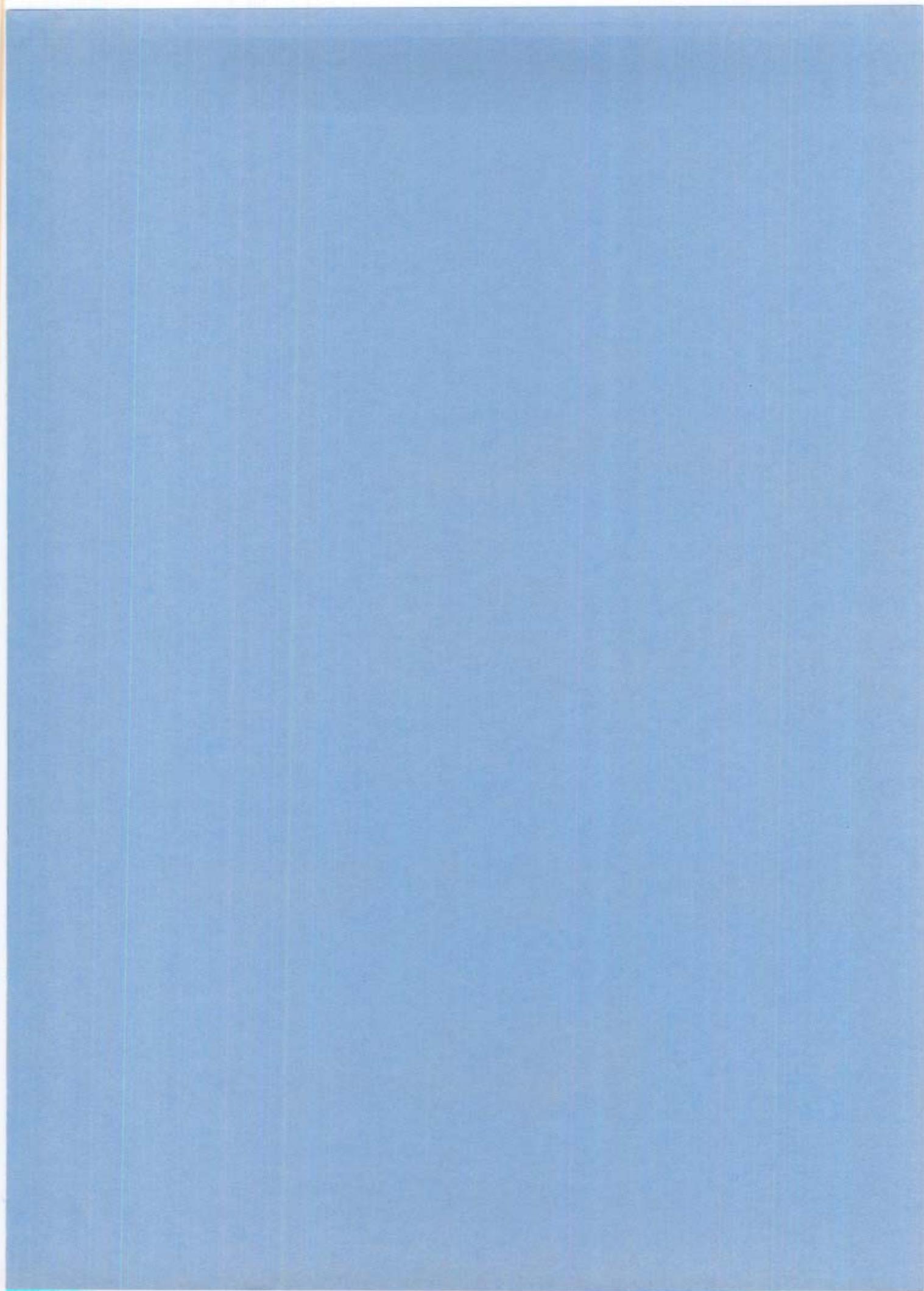
Pour mieux apprécier cette possibilité, il est sans doute utile ici de faire intervenir une métaphore extraite de la biologie moléculaire : comme on le sait, la synthèse des protéines commence par la formation de chaînes linéaires d'acides aminés reliés les uns aux autres par des liaisons covalentes fortes, cet enchaînement étant suivi par un processus de repli, qui est déterminé par des interactions d'intensité plus faible entre molécules distantes sur la chaîne. Il s'établit donc des "ponts" entre différentes régions du polymère, qui confèrent à celui-ci sa structure tridimensionnelle très spécifique, responsable de ses propriétés chimiques et biologiques. Il peut alors s'avérer particulièrement intéressant d'établir une analogie

entre chaînes polypeptidiques et chaînes de neurones, c'est à dire entre protéines et CST complexes : ainsi, les chaînes de notre modèle, constituées de neurones reliés par des connexions synaptiques fortes, seraient également susceptibles de se "replier", tout comme les protéines, ce repli s'effectuant par l'apparition de connexions supplémentaires (plus faibles que les premières), accompagnées de nouvelles corrélations temporelles stables entre activités neuronales "éloignées" le long de la chaîne (remarquons que, dans ce cas, les liens faibles ne respectent pas la cohérence des délais propre à la structure des liens forts, car ils créent des "courts-circuits" entre différents éléments; cependant, la cohérence est toujours conservée localement). C'est donc ici que l'on peut entrevoir toute la richesse des CST complexes : en résumé, celles-ci posséderaient des conformations très particulières qui leur permettraient de se composer entre elles de façon hautement spécifique.

A travers cette éventualité, nous avons décrit ce que pourrait être l'implémentation d'une "machine à composer les symboles" dans le cortex. En résumé : (a) les symboles sont des CST complexes, caractérisés localement par une forte régularité temporelle interne, et globalement par une structure spécifique, qui serait due, par exemple, au "repli" d'une chaîne linéaire; (b) ces symboles peuvent être composés entre eux à travers un mécanisme élémentaire de collage par synchronisation (voir explications du §6.1 et du §6.2); (c) les "règles syntaxiques", c'est à dire les contraintes guidant cette composition et lui conférant un caractère non-trivial, sont dictées directement par la structure spécifique de ces CST (tout comme les réactions allostériques en biologie moléculaire, c'est à dire les "emboîtements" de molécules ayant des formes complémentaires)

Pour conclure : nous avons d'abord décrit le développement épigénétique à un niveau très local, en comparant la focalisation de l'innervation et la formation des chaînes de connexions à un processus de croissance agrégative ("croissance cristalline") à partir d'une inhomogénéité locale ("germe" de la croissance). Cette première partie a reçu une confirmation expérimentale numérique solide, et a constitué l'objet principal de l'étude exposée dans la présente thèse (chapitres 3, 4 et 5 de la partie III). Cependant, comme nous avons tenté de le montrer ici, cette description est très limitée et incomplète : il faut ensuite passer à un niveau d'intégration supérieur, qui envisage des interactions complexes entre chaînes. Nous disposons alors de plusieurs métaphores pour extrapoler le comportement du modèle (qui devront être confirmées ultérieurement par des simulations numériques), et nous aboutissons finalement à une image très particulière d'une construction auto-organisatrice qui serait peut-être caractéristique du système nerveux.

BIBLIOGRAPHIE



- ABELES, M. (1982). Local cortical circuits: an electrophysiological study. *Studies of brain function* (vol. 7). V. Braitenberg ed., Springer-Verlag, Berlin.
- ABELES, M. and GERSTEIN, G. L. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.* 60: 909-924.
- ACKLEY, D. H., HINTON, G.E. and SEJNOWSKI, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Sci.* 9: 147-169.
- BARLOW, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* 1: 371-394.
- BELLMANN, R. E. (1961). *Adaptive Control Processes*. Princeton University Press. Princeton, New Jersey.
- BIENENSTOCK, E. (1988) Neural-like Graph-matching Techniques for Image Processing. In: *Organization of Neural networks*. W. von Seelen, G. Shaw, and U.M. Leinhos, Eds. VCH Verlags- gesellschaft, Weinheim, Germany.
- BIENENSTOCK, E. and von der MALSBERG, C. (1987). A neural network for invariant pattern recognition. *Europhys. Lett.* 4 (1): 121-126.
- BIENENSTOCK, E., and DOURSAT, R. (1989). Elastic matching and pattern recognition in neural networks. In: *Neural networks: From Models to Applications*. L. Personnaz and G. Dreyfus, Eds. IDSET, Paris.
- BIENENSTOCK, E., and DOURSAT, R. (1991). Issues of representation in neural networks. In: *Representation of Vision: Trends and Tacit assumptions in Vision research*. A. Gorea, ed. Cambridge University Press.
- BLAKEMORE, C. and COOPER, G. F. (1970). Development of the brain depends on the visual environment. *Nature* 228: 477-478.
- BUHMANN, J., LANGE, J., von der MALSBERG, C., VORBRÜGGEN, J. C. and WÜRZT, R. P. (1989). Object recognition in the dynamic link architecture - Parallel implementation on a transputer network. In: *Neural Networks: A dynamic systems approach to machine intelligence*. B. Kosko, ed. Prentice Hall, New York.
- BURR, D. J. (1980). Elastic Matching of line drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-3, 6: 708-713.
- CHANGEUX, J.-P. (1983). *L'homme neuronal*. Fayard, Paris.
- CHANGEUX, J.-P. and DANCHIN A. (1976). Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. *Nature* 264: 705-712.
- CHOMSKY, N. (1957). *Syntactic structures*. Mouton and Co., The Hague.
- CONSTANTINE-PATON, M. and LAW, M. I. (1978). Eye-specific termination bands in tecta of three-eyed frogs. *Science* 202: 639-641.
- CREPEL, F., MARIANI, J. and DELHAYE-BOUCHAUD, N. (1976). Evidence for a multiple innervation of Purkinje cells by climbing fibers in the immature rat cerebellum. *J. Neurobiol.* 7: 567-578.
- CRICK, F. (1982). Do dendritic spines twitch ? *Trends in Neurosci.* 5: 44-46.
- DAMASIO, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation* 1: 123-132.
- DE BEER, G. R. (1958). *Embryos and Ancestors*. Clarendon Press, Oxford.

- DUDA, R. O. and HART, P. E. (1973). *Pattern Classification and Scene analysis*. John Wiley & Sons, New York.
- ECKHORN, R., BAUER, R., JORDAN, W., BROSCHE, M., KRUSE, W., MUNK, M. and REITBÖCK, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybern.* 60: 121-130.
- EDELMAN, G. M. (1983). Cell adhesion molecules. *Science* 219: 450-457.
- FODOR, J. A. (1975). *The language of thought*. Harvard University Press. Cambridge, Massachusetts.
- FODOR, J. A. and PYLYSHYN, Z. (1988). Connectionism and Cognitive architecture: a critical analysis. *Cognition* 28, 1/2: 3-71
- GEMAN, S., BIENENSTOCK, E. and DOURSAT, R. (1991, submitted). Neural networks and the bias/variance dilemma.
- GOULD, S. J. (1977). *Ontogeny and Phylogeny*. Harvard University Press. Cambridge, Massachusetts.
- GRAY, C. M., KÖNIG, P., ENGEL, A. K. and SINGER, W. (1989). Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature* 338: 334-337.
- GUYON, I. (1988). Réseaux de neurones pour la reconnaissance des formes : architectures et apprentissage. Travail de doctorat non-publié (décembre 1988). Université Paris VI.
- HAMBURGER, V. (1934). The effects of wing bud extirpation on the development of the central nervous system in chick embryos. *J. Exp. Zool.* 68: 449-494.
- HEBB, D. O. (1949). *The organization of behavior*. John Wiley & Sons, New York.
- HINTON, G.E. and SEJNOWSKI, T. J. (1986). Learning and relearning in Boltzmann machines. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, Rumelhart, McClelland and the PDP group, eds. MIT Press, Cambridge, Massachusetts.
- HOPFIELD, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79: 2554-2558.
- HUBEL, D. H. and WIESEL, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. Lond.* 148: 574-591.
- HUBEL, D. H. and WIESEL, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. Lond.* 160: 106-154.
- HUBEL, D. H. and WIESEL, T. N. (1965). Binocular interaction in striate cortex of kittens reared with artificial squint. *J. Neurophysiol.* 28: 1041-1059.
- HUBEL, D. H. and WIESEL, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. Lond.* 195: 215-243.
- HUBEL, D. H. and WIESEL, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J. Physiol. Lond.* 206: 419-436.
- HUBEL, D. H. and WIESEL, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp. Neurol.* 158: 267-294.
- HUBEL, D. H., WIESEL, T. N. and LEVAY, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Phil. Trans. R. Soc. Lond. B* 278: 131-163.

LE CUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W. and JACKEL, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comp.* 1: 541-551.

LEVI-MONTALCINI, R. (1953). Effects of mouse tumor transplantation on the nervous system. *Ann. NY Acad. Sci.* 55: 330-343.

McCULLOCH, W. S. and PITTS, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophysics* 5: 115-133.

von der MALSBERG, C. (1973). Self-organization of orientation-sensitive cell in the striate cortex. *Kybernetik* 14: 85-100.

von der MALSBERG, C. (1981). The correlation theory of brain function. Internal Report 81-2. Max Planck Institute for Biophysical Chemistry, Dept. of Neurobiology. Göttingen, Germany.

von der MALSBERG, C. (1986). Am I thinking assemblies? In: *Brain Theory*. G. Palm and A. Aertsen eds. Springer-Verlag, Heidelberg.

von der MALSBERG, C. (1987). Synaptic Plasticity as Basis of Brain Organization. In: *The neural and Molecular Bases of Learning*. J.-P. Changeux and M. Komishi, pp. 411-432. John Wiley & Sons, New York.

von der MALSBERG, C. and BIENENSTOCK, E. (1986). Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain. In: *Disordered systems and biological organization*: 247-272. E. Bienenstock, F. Fogelman and G. Weisbuch eds. Springer Verlag, Berlin.

von der MALSBERG, C. and BIENENSTOCK, E. (1987). A neural network for the retrieval of superimposed connection patterns. *Europhys. Lett.* 3 (11): 1243-1249.

von der MALSBERG, C. and WILLSHAW, D. J. (1977). How to label nerve cells so that they can interconnect in an ordered fashion. *Proc. Natl. Acad. Sci. USA* 74: 11, 5176-5178.

MARIANI, J. and CHANGEUX, J.-P. (1981). Ontogenesis of olivocerebellar relationships. Studies by intracellular recordings of the multiple innervation of Purkinje cells by climbing fibers in the developing rat cerebellum. *J. Neurosci.* 1: 696-702.

MASTRONARDE, D. N. (1983). Correlated firing of cat retinal ganglion cells. *J. Neurophysiol.* 49: 303-324.

PERRETT, D., ROLLS, E. and CAAN, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Exp Brain Res.* 47: 329-342.

PIATTELLI-PALMARINI, M. (1978). Théories du langage, théories de l'apprentissage: le débat entre Jean Piaget et Noam Chomsky, au "Centre Royaumont pour une science de l'homme" (débat organisé et recueilli par M. P.-P.). Points, Seuil, Paris.

PROCHIANTZ, A. (1989). La construction du cerveau. Hachette, Paris.

PURVES, D. and LICHTMAN, J. W. (1980). Elimination of synapses in the developing nervous system. *Science* 210: 153-157.

PURVES, D. and LICHTMAN, J. W. (1985). Principles of Neural Development. Sinauer. Sunderland, Massachusetts.

RAKIC, P. (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* 145: 61-84.

- REDFERN, P. A. (1970). Neuromuscular transmission in new-born rats. *J. Physiol. London* 209: 701-709.
- ROSENBLATT, F. (1962). Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. *Spartan Books*, Washington.
- RUMELHART, D. E., HINTON, G. E. and WILLIAMS, R. J. (1986a). Learning internal representations by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, Rumelhart, McClelland and the PDP group, eds., pp318-362. MIT Press, Cambridge, Massachusetts.
- RUMELHART, D. E., HINTON, G. E. and WILLIAMS, R. J. (1986b). Learning internal representations by backpropagating error. *Nature (London)* 323: 553-556.
- SCHMIDT, J. (1985). Formation of Retinotopic Connections: Selective Stabilization by an Activity-Dependent Mechanism. *Cellular and Molecular Neurobiology* 5, 1/2: 65-84
- SEJNOWSKI, T. (1981). Skeleton Filters in the Brain. In: *Parallel Models of Associative Memory*. G. E. Hinton and J. A. Anderson eds., Erlbaum Associates. Hillsdale, New Jersey.
- SMOLENSKY, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11: 1-74.
- SOLLA, S. A. (1989). Learning and generalization in layered neural networks: The contiguity problem. In: *Neural networks: From Models to Applications*. L. Personnaz and G. Dreyfus, Eds. IDSET, Paris.
- SPEMANN, H. (1938). *Embryonic Development and Induction*. Yale University Press, New Haven.
- SPERRY, R. W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl. Acad. Sci. USA* 50: 703-710.
- STENT, G. S. (1973). A physiological mechanism for Hebb's postulate of learning. *Proc Natl. Acad. Sci. USA* 70: 997-1001.
- TAPPERT, C. C. (1982). Cursive Script Recognition by Elastic Matching. *IBM Journal of Research and Development* 26: 765-771.
- WHITE, H. (1990). Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings. *Neural Networks* 3: 535-549.
- WILLSHAW, D. J. and von der MALSBERG, C. (1976). How patterned neural connections can be set up by self-organization. *Proc. R. Soc. Lond. B* 194: 431-445.
- WILLSHAW, D. J. and von der MALSBERG, C. (1979). A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. *Phil. Trans. R. Soc. Lond. B* 287: 203-243.