

Université Lumière Lyon2

École Doctorale “Cerveau et Cognition”

Laboratoire d’InfoRmatique en Images et Systèmes d’information

LIRIS - UMR CNRS 5205 - Équipe TURING

# UNE MODÉLISATION ÉVOLUTIONNISTE DU LIAGE TEMPOREL

**Par David MEUNIER**

Thèse de doctorat de **Sciences Cognitives**  
**Mention Informatique**

*Dirigée par* Professeur Hélène PAUGAM-MOISY

*Présenté et soutenue publiquement le* **19 Octobre 2007**

*Devant un jury composé de :*

Docteur Hugues BERRY (Examineur)

Docteur Olivier BERTRAND (Examineur)

Docteur Agnès GUILLOT (Examineur)

Professeur Hélène PAUGAM-MOISY (Directeur de thèse)

Professeur Marc SCHOENAUER (Rapporteur)

Docteur Catherine TALLON-BAUDRY (Rapporteur)

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problématique du liage des traits . . . . .	1
1.2	Modèle de liage temporel . . . . .	2
1.3	Emergence par l'évolution . . . . .	2
1.4	Théorie des réseaux complexes . . . . .	3
1.5	Plan de la thèse . . . . .	3
<b>2</b>	<b>Neurophysiologie</b>	<b>4</b>
2.1	Fonctionnement des neurones biologiques . . . . .	4
2.1.1	Propriétés des neurones biologiques . . . . .	4
2.1.2	Codage de l'information par les neurones . . . . .	5
2.2	Plasticité synaptique . . . . .	8
2.2.1	Assemblées cellulaires . . . . .	8
2.2.2	Plasticité synaptique par taux de décharge . . . . .	9
2.2.3	Plasticité synaptique temporelle . . . . .	9
2.3	Neuro-imagerie et électrophysiologie . . . . .	10
2.3.1	Fonction cognitive localisée . . . . .	11
2.3.2	Fonctionnement "en réseau" . . . . .	11
<b>3</b>	<b>Hypothèse du liage temporel</b>	<b>13</b>
3.1	Problématique du liage des traits . . . . .	13
3.2	Liage par convergence . . . . .	14
3.3	Hypothèse du liage temporel . . . . .	15
3.3.1	Assemblées temporelles . . . . .	15
3.3.2	Oscillations . . . . .	18
3.4	Évidence expérimentale du liage temporel . . . . .	21
3.4.1	Perception . . . . .	21
3.4.2	Processus de groupage . . . . .	22
3.4.3	Attention et influence <i>top-down</i> . . . . .	24
3.4.4	Mémorisation . . . . .	25
3.4.5	Intégration multi-modale . . . . .	26
3.4.6	Pathologies . . . . .	27
<b>4</b>	<b>Neurosciences computationnelles</b>	<b>29</b>
4.1	Réseaux de neurones artificiels . . . . .	29
4.1.1	Réseaux de neurones artificiels "classiques" . . . . .	29

4.1.2	Réseaux de neurones temporels . . . . .	32
4.2	Modèles de synchronisation neuronale . . . . .	36
4.2.1	Synchronisation d'oscillateurs . . . . .	37
4.2.2	Modèle des <i>synfire chains</i> . . . . .	37
4.2.3	Polychronisation . . . . .	38
4.2.4	Emergence d'oscillations dans une structure particulière . . . . .	39
4.2.5	Modèles de liage des traits . . . . .	42
4.2.6	Modèle <i>a priori</i> versus <i>a posteriori</i> . . . . .	43
<b>5</b>	<b>Evolution</b>	<b>45</b>
5.1	Evolution biologique . . . . .	45
5.1.1	Théorie synthétique de l'évolution . . . . .	45
5.1.2	Effet Baldwin . . . . .	47
5.2	Algorithme évolutionniste . . . . .	48
5.2.1	Principes de fonctionnement . . . . .	48
5.2.2	Explications du fonctionnement . . . . .	50
5.2.3	Applications . . . . .	50
5.3	Evolution et cerveau . . . . .	51
5.3.1	Evolution et apprentissage . . . . .	51
5.3.2	Simulation de l'évolution du cerveau . . . . .	53
5.3.3	Applications en robotique . . . . .	57
5.3.4	Retour à la problématique . . . . .	60
<b>6</b>	<b>Théorie des réseaux complexes</b>	<b>61</b>
6.1	Introduction . . . . .	61
6.1.1	Systèmes complexes . . . . .	61
6.1.2	Réseaux complexes . . . . .	61
6.1.3	Applications de la théorie des réseaux complexes . . . . .	62
6.2	Modèles et mesures pour les réseaux complexes . . . . .	63
6.2.1	Composantes fortement connexes . . . . .	63
6.2.2	Modèle de réseaux petit-monde et mesures d'efficacité . . . . .	64
6.2.3	Réseaux invariants d'échelle . . . . .	68
6.2.4	Modularité . . . . .	68
6.3	Dynamique dans les modèles de réseaux complexes . . . . .	73
6.3.1	Dynamique dans les réseaux petit-monde . . . . .	74
6.3.2	Dynamique dans les réseaux invariants d'échelle . . . . .	74
<b>7</b>	<b>Réseaux complexes et cerveau</b>	<b>76</b>
7.1	Modularité du cerveau . . . . .	77
7.1.1	Module Fodorien . . . . .	77
7.1.2	Module Darwinien . . . . .	77
7.2	Réseaux complexes et neuro-imagerie . . . . .	78
7.2.1	Niveau anatomique . . . . .	78
7.2.2	Niveau dynamique . . . . .	79
7.3	Réseaux complexes et Neurosciences computationnelles . . . . .	81
7.4	Réseaux complexes et évolution . . . . .	83

<b>8</b>	<b>Le modèle EvoSNN</b>	<b>85</b>
8.1	Modèle de neurone . . . . .	85
8.2	Modèle de synapse . . . . .	87
8.2.1	STDP . . . . .	87
8.2.2	Règles additive ou multiplicative . . . . .	87
8.2.3	Application de la STDP . . . . .	88
8.3	Modèle d'évolution . . . . .	88
8.3.1	Construction de la topologie du réseau . . . . .	88
8.3.2	Environnement virtuel . . . . .	93
8.3.3	Algorithme évolutionniste . . . . .	95
<b>9</b>	<b>Résultats sur l'évolution et l'apprentissage</b>	<b>97</b>
9.1	Résultats sur le comportement . . . . .	97
9.1.1	Résultats sur l'ensemble des simulations . . . . .	97
9.1.2	Exemple d'une simulation d'évolution . . . . .	99
9.1.3	Exemple du passage dans l'environnement . . . . .	100
9.2	Résultats sur l'apprentissage . . . . .	101
9.2.1	Protocole de test . . . . .	101
9.2.2	Mise en œuvre du protocole de test . . . . .	102
<b>10</b>	<b>Résultats sur la topologie</b>	<b>105</b>
10.1	Résultats sur la connectivité . . . . .	105
10.1.1	Projections "interfaces" . . . . .	105
10.1.2	Projections internes au réseau . . . . .	106
10.2	Résultats sur la composante fortement connexe géante . . . . .	107
10.3	Mesures "petit-monde" et efficacité . . . . .	108
10.4	Résultats sur la modularité . . . . .	109
10.4.1	Algorithme NG étendu aux graphes orientés . . . . .	109
10.4.2	Exemple d'application de l'algorithme NG étendu . . . . .	110
10.4.3	Résultats avec l'algorithme NG étendu . . . . .	111
10.5	Interprétation des résultats topologiques . . . . .	112
<b>11</b>	<b>Résultats sur la dynamique</b>	<b>114</b>
11.1	Protocole de stimulation . . . . .	114
11.2	Résultats sur les bandes de fréquences . . . . .	115
11.2.1	Calcul des fréquences à partir des PA . . . . .	116
11.2.2	Calcul des fréquences à partir des signaux continus . . . . .	117
11.3	Résultats sur les cross-corrélogrammes . . . . .	120
11.3.1	Ajustements des cross-corrélogrammes . . . . .	121
11.3.2	K-moyennes sur les paramètres d'ajustement . . . . .	123
11.3.3	Résultats des ajustements . . . . .	124
11.4	Différences entre les stimuli . . . . .	128
11.5	Interprétation des résultats dynamiques . . . . .	130

<b>12 Discussion</b>	<b>132</b>
12.1 Evolution . . . . .	133
12.2 Apprentissage . . . . .	134
12.2.1 Apprentissage dynamique . . . . .	134
12.2.2 Connaissances innées . . . . .	135
12.2.3 Effet Baldwin . . . . .	135
12.3 Topologie . . . . .	136
12.3.1 Augmentation du nombre de projections “interfaces” . . . . .	136
12.3.2 Absence d’émergence de structures modulaires . . . . .	137
12.3.3 Optimisation du coût de câblage . . . . .	138
12.4 Dynamique . . . . .	138
12.4.1 Considérations méthodologiques . . . . .	139
12.4.2 Activité de fond . . . . .	140
12.4.3 Injection d’un stimulus . . . . .	142
12.4.4 Formation d’une assemblée temporelle et liage temporel . . . . .	145
<b>13 Conclusion</b>	<b>147</b>
<b>14 Perspectives</b>	<b>148</b>
14.1 Améliorations du modèle . . . . .	148
14.1.1 Inclusion d’un “coût” métabolique . . . . .	148
14.1.2 Réseau petit-monde . . . . .	149
14.1.3 Réseau exponentiel . . . . .	149
14.2 Liens entre la topologie et la dynamique . . . . .	149
14.3 Application aux études de neuro-imagerie . . . . .	150
<b>A Théorie des graphes</b>	<b>152</b>
A.1 Plusieurs types de graphes . . . . .	152
A.2 Représentation matricielle d’un graphe . . . . .	153
A.3 Mesures sur un graphe . . . . .	153
A.3.1 Degré . . . . .	153
A.3.2 Plus court chemin . . . . .	153
A.4 Réseaux aléatoires et réseaux de voisinage . . . . .	154
A.5 Parité des circuits . . . . .	154
<b>B Méthodes d’analyse des signaux continus</b>	<b>156</b>
B.1 Potentiels évoqués . . . . .	156
B.2 Oscillations induites . . . . .	157
B.2.1 Diagrammes Temps-Fréquence . . . . .	157
B.2.2 Oscillations évoquées et induites . . . . .	157
B.2.3 Synchronisation de phase . . . . .	158
<b>C Cross-corrélogrammes</b>	<b>159</b>
C.1 Calcul d’un cross-corrélogramme . . . . .	159
C.2 Cross-corrélogramme corrigé . . . . .	160
C.3 Normalisation . . . . .	160

<b>D</b>	<b>Statistiques</b>	<b>161</b>
D.1	ANOVA . . . . .	161
D.2	$\chi^2$ réduit . . . . .	161
<b>E</b>	<b>K-moyennes</b>	<b>162</b>

# Chapitre 4

## Neurosciences computationnelles

Fondée sur l’hypothèse d’Alan Turing, l’approche computationnelle des neurosciences suppose que tout processus cognitif est le résultat d’un calcul. Ainsi, les modèles développés en neurosciences computationnelles cherchent à reproduire par des simulations informatiques les types de calculs effectués par le système nerveux.

Nous aborderons dans un premier temps les modèles de réseaux de neurones artificiels (section 4.1.1), puis les modèles de réseaux de neurones temporels (section 4.1.2), pour présenter ensuite les modèles qui cherchent à reproduire le phénomène de synchronisation neuronale (section 4.2).

### 4.1 Réseaux de neurones artificiels

Les réseaux de neurones artificiels (RNA) correspondent à un ensemble d’objets mathématiques, les neurones artificiels, connectés et qui interagissent entre eux. Chaque objet a accès à une partie de l’information, mais l’ensemble des éléments peut effectuer un traitement complet au niveau global. On parle alors de fonctionnement distribué.

Le connexionisme est le domaine qui étudie les propriétés des RNA. À partir de modèles d’inspiration biologique, un certain nombre de travaux s’éloignent ensuite de la biologie pour étudier les propriétés mathématiques des RNA, et permettre la mise en place d’outils utilisés dans le domaine des sciences de l’ingénieur, tels que l’apprentissage automatique, l’optimisation, etc.

Nous introduisons ici un certain nombre de modèles de RNA classiques, ainsi que leurs principales propriétés. Les réseaux de neurones temporels sont ensuite introduits.

#### 4.1.1 Réseaux de neurones artificiels “classiques”

Nous ferons ici un bref rappel des modèles de réseaux de neurones artificiels classiques, correspondant à des modèles mis en place jusqu’au début des années 1980. Cette introduction nous permettra d’une part d’aborder certains aspects propres aux réseaux de neurones, tels que leur capacité d’apprentissage et l’étude de leurs dynamiques, et d’autre part de mettre en évidence les différences qu’il existe avec les modèles de réseaux

de neurones temporels de la section 4.1.2.

## Modèle de neurone à seuil

A partir des données expérimentales de la section 2.1.1, McCulloch et Pitts (1943) ont défini le neurone à seuil, un modèle mathématique schématisant le fonctionnement du neurone biologique. L'état du neurone est décrit par une variable binaire, correspondant à la loi du tout-ou-rien : soit le neurone est actif (sa sortie vaut 1), soit le neurone est au repos (sa sortie vaut 0). L'état du neurone est calculé à partir des activités qu'il reçoit en entrée : si la somme des états des activités entrantes dépasse un certain seuil, alors sa sortie vaut 1, sinon elle vaut 0.

Dans un réseau, les activités entrantes d'un neurone sont les sorties des autres neurones, pondérées par les poids des synapses que les signaux traversent. Selon la nature excitatrice ou inhibitrice d'une synapse, son poids est positif ou négatif. La contribution d'un signal traversant une synapse excitatrice permet à la somme des activités de se rapprocher de la valeur du seuil. Au contraire, la contribution d'une synapse inhibitrice empêche la somme des activités de se rapprocher de la valeur de seuil.

## Modèle de neurone sigmoïdal

Un deuxième modèle de neurone classique est le modèle de neurone sigmoïdal (Hopfield, 1984; Rumelhart et Mc Clelland, 1986). Dans ce modèle, la sortie du neurone n'est plus binaire, mais prend des valeurs réelles entre 0 et 1. Cette valeur correspond au taux de décharge du neurone (voir section 2.1.2). Les valeurs transmises entre les neurones sont donc cette fois des valeurs continues.

## Apprentissage

Le paradigme de l'encodage des souvenirs par l'intermédiaire de la plasticité synaptique (section 2.2.1) est l'idée de base sur laquelle reposent la plupart des méthodes d'apprentissage automatique dans les RNA. Dans certains modèles connexionnistes, un **apprentissage supervisé** est réalisé en modifiant les valeurs des poids synaptiques, de manière à minimiser l'erreur réalisée par le réseau, vis-à-vis d'une sortie désirée. Pour le perceptron (Rosenblatt, 1958), utilisant des neurones à seuil, la modification des poids synaptiques est basée sur un algorithme de correction d'erreur. Dans le perceptron multicouches (Rumelhart et Mc Clelland, 1986), basé sur le modèle de neurone sigmoïdal, on utilise la retropropagation du gradient d'erreur pour modifier les poids synaptiques. Ces modèles, qui ont abouti à la théorie statistique de l'apprentissage (Vapnik, 1998), sont largement utilisés dans le domaine de l'apprentissage automatique. Cependant, la plausibilité biologique de leurs mécanismes, bien qu'initialement d'inspiration biologique, a été largement remise en cause du fait de la nature supervisée de l'apprentissage (voir par exemple Thorpe et Imbert (1989)). De même, le lien entre plasticité synaptique et apprentissage est encore discuté au niveau biologique (Edeline, 1996, 1999).



D'autres modèles reposent sur des principes assez éloignés de la plasticité synaptique. Par exemple, Les cartes auto-organisatrices de Kohonen (1982)) sont inspirées par l'organisation topologique des cartes corticales, et les réponses des neurones se font par auto-organisation. Les neurones utilisés ici calculent les distances au motif présenté, et le neurone le plus proche du motif est le neurone gagnant, et va se spécialiser dans la réponse à ce motif. Dans ce cas, on parle d'**apprentissage non-supervisé**, puisque il n'y a pas de sortie désirée, seulement une organisation topologique .

Un autre type d'apprentissage est l'**apprentissage par renforcement** (Sutton et Barto, 1998). Par exemple, avec l'algorithme de *Q-Learning* (Watkins et Dayan, 1992), on cherche à ce qu'un agent apprenne un comportement global. L'agent doit trouver ce comportement par essais/erreurs, et on lui communique à chaque étape un signal de punition ou de récompense, selon que le comportement local du modèle se rapproche, ou non, de la réponse globale. Ces signaux vont modifier les paramètres du modèle, de manière à ce qu'il adapte sa réponse et parvienne à effectuer le comportement global.

## Dynamique et chaos dans les RNA

Certains modèles de réseaux de neurones s'intéressent à la dynamique d'un ensemble de neurones reliés entre eux. C'est par exemple le cas du modèle de Hopfield (1982, 1984). Dans ce modèle, qui utilise le modèle du neurone à seuil (Hopfield, 1982), puis le modèle du neurone sigmoïdal (Hopfield, 1984), tous les neurones sont connectés entre eux. Les poids synaptiques, calculés en utilisant une règle hebbienne, permettent de stocker dans le réseau un certain nombre de motifs. On présente ensuite un nouveau motif dans le réseau, correspondant aux activités des neurones, puis on laisse la dynamique du réseau se développer. Après un certain nombre d'itérations, le réseau atteint un état stable. Lorsque le motif initial correspond à une version légèrement bruitée d'un des motifs stockés dans le réseau, la dynamique du réseau permet de compléter le motif. On parle de **mémoire associative**, car le réseau peut associer un motif bruité à la version du motif complet contenu en mémoire.

Ce type de réseau de neurones, appartenant à la famille des réseaux récurrents, présente un intérêt biologique, du fait que la reconnaissance du motif se fait sur un certain nombre d'itérations. On parle alors de dynamique d'attracteur : la reconnaissance d'un motif présenté se fait lorsque la dynamique converge vers un des états stables du réseau, correspondant aux motifs stockés dans le réseau. Cependant, l'utilisation d'une matrice de connexion symétrique fait que la dynamique dans le réseau de Hopfield est relativement simple. Dans le cas d'une mise à jour séquentielle des sorties des neurones, les motifs stockés dans le réseau correspondent à des attracteurs dit points fixes de la dynamique. Une fois que la dynamique a atteint un état stable, il est impossible de sortir de cet attracteur. D'autres auteurs (Amit, 1989, 1995; Daucé et al., 1998) ont étudiés les dynamiques en utilisant le modèle de Hopfield, avec une matrice est assymetrique. Les dynamiques qui en résultent sont plus complexes, avec l'apparition d'attracteurs cycle-limite, ou d'attracteurs chaotiques.

Tsuda (2001) et par la suite Kay (2003), proposent quand à eux que la dynamique du cerveau fonctionnent par **itinérance chaotique** : une dynamique chaotique permet que la reconnaissance d'un motif se fasse par un passage transitoire dans un attracteur, puis la possibilité d'en sortir pour visiter un autre attracteur, etc. Ce fonctionnement expli-

querait l'émergence d'une activité oscillante transitoire (voir section 3.3.2), permettant au système de ne pas rester bloqué dans un bassin d'attraction.

## Apprentissage dynamique dans les RNA

L'approche dynamique de la cognition (van Gelder, 1998) suppose un fonctionnement du système nerveux basé sur la théorie des systèmes dynamiques. Dans le cadre de cette théorie, un système de  $N$  variables est représenté sous la forme d'un vecteur d'état  $\mathbf{X}(t)$ , correspondant à l'état des  $N$  variables au temps  $t$ . Dans le cas du système nerveux, le vecteur d'état correspond à l'ensemble des états des neurones, par exemple la valeur de leurs potentiels de membrane, au temps  $t$ . De part ces interactions, l'état présent du système  $\mathbf{X}(t)$  dépend de l'ensemble des états précédents, de  $\mathbf{X}(0)$  à  $\mathbf{X}(t-1)$ , car il possède une dynamique intrinsèque générée par les activités de ses éléments, qui peuvent fluctuer même en l'absence de perturbations extérieures (Guillot et Daucé, 2002). L'existence d'une telle dynamique permet d'introduire une mémoire des événements rencontrés par l'individu, du fait que les stimuli perturbent la dynamique intrinsèque du système, et le font bifurquer vers un état différent de celui dans lequel il se trouve en l'absence de stimuli (Beer, 2000). Dans le cadre d'un fonctionnement dynamique ou chaotique, la plasticité synaptique n'est plus vu comme le support de l'apprentissage, mais le régulateur de l'activité neuronale (Freeman, 1994).

Enfin, Thomas et Kaufman (2001) proposent que l'existence de circuits positifs (voir annexe A.5) dans un système permette d'obtenir la propriété de multistationnarité, c.à.d l'existence de plusieurs états stables dans le système. Dans le cadre de l'hypothèse du liage temporel, cette propriété permet d'expliquer la coexistence de plusieurs états stables dans le système, ce qui conduit au stockage de souvenirs.

Les réseaux de neurones artificiels développés jusqu'à la fin des années 1980 se basaient sur le postulat que les temps d'émissions des PA ne portaient pas d'information suffisamment précises pour qu'il soit intéressant de les prendre en compte. L'apparition de données expérimentales sur le codage de l'information du neurone biologique (voir section 2.1.2) a poussé les modélisateurs à s'intéresser à des modèles où l'on tient compte de certaines propriétés temporelles intrinsèques du neurone.

### 4.1.2 Réseaux de neurones temporels

#### Modèles de neurone temporel

Les données expérimentales évoquées dans la section 2.1.2 ont eu un grand retentissement dans le domaine de la modélisation des réseaux de neurones. Certains modèles reproduisant au plus près le fonctionnement du neurone biologique ont été utilisés dans le cadre de simulations. Les modèles de neurones présentés ci-dessous prennent en compte le temps de manière explicite, et ont tous la propriété d'émettre en sortie un PA, caractérisé par son temps d'émission. Ces différents modèles sont regroupés sous le nom de neurones impulsifs (*spiking neurons*).

*Le modèle de Hodgkin et Huxley*

Les modèles les plus anciens sont les modèles à conductance, tel que le modèle **électrique** bien connu de Hodgkin et Huxley (1952a,b) (HH). Ce modèle repose sur quatre équations différentielles non-linéaires couplées. L'idée de base est de modéliser le comportement temporel du potentiel de membrane  $u(t)$  du neurone par des circuits électriques, composés de résistances et de condensateurs :  $C$  est la capacité de la membrane, les  $g_i$  sont les conductances des différents canaux ioniques (Potassium K+, sodium Na+, etc.) et les  $E_i$  sont les potentiels d'équilibre correspondants. Les variables  $m$ ,  $h$  and  $n$  décrivent l'ouverture et la fermeture des canaux.

$$C \frac{du}{dt} = -g_{Na} m^3 h (u - E_{Na}) - g_K n^4 (u - E_K) - g_L (u - E_L) + I(t) \quad (4.1)$$

$$\tau_n \frac{dn}{dt} = -[n - n_0(u)] \quad \tau_m \frac{dm}{dt} = -[m - m_0(u)] \quad \tau_h \frac{dh}{dt} = -[h - h_0(u)]$$

Le modèle reproduit différents modes de fonctionnement du neurone biologique. Une activation peut se traduire par l'émission d'un potentiel d'action unique, d'une bouffée (*burst*) de quelques PA à très haute fréquence, ou par une activité périodique soutenue (Gerstner et Kistler, 2002). Son implémentation requiert la mise en oeuvre de techniques relativement complexes d'analyse numérique. C'est cependant le modèle qui permet de reproduire au mieux le fonctionnement du neurone biologique.

#### *Le modèle LIF*

Le modèle "Leaky Integrate & Fire" (LIF) (Lapicque (1907), cité par Abbott (1999)) se place à un niveau de description moins précis que le modèle HH, permettant une implémentation plus aisée. Le neurone est modélisé par un circuit électrique disposant d'un condensateur  $C$  en parallèle avec une résistance  $R$ , et la dynamique du neurone est décrite en fonction de son potentiel de membrane  $u(t)$  et de l'intensité  $I(t)$  qui circule dans le neurone. Ce modèle repose sur une équation différentielle du premier ordre. En définissant la constante de temps  $\tau_m = RC$ , comme le temps caractéristique de la fuite du potentiel de membrane pour retourner à son potentiel de repos  $u_{rest}$ , on a alors l'équation :

$$\tau_m \frac{du}{dt} = u_{rest} - u(t) + RI(t) \quad (4.2)$$

De plus, le temps d'émission d'un PA est défini comme le temps où le potentiel de membrane croise la valeur de seuil  $u(t^{(f)}) = \vartheta$ , avec la condition  $u'(t^{(f)}) > 0$ . Juste après  $t^{(f)}$ , le potentiel est réinitialisé à une valeur  $u_{abs}$  durant une durée  $d_{abs}$ , afin de modéliser une période réfractaire absolue.

#### *Le modèle SRM*

Le modèle "Spike Response Model" (SRM) (Gerstner et van Hemmen, 1992; Gerstner et Kistler, 2002) est une modélisation **phénoménologique** du neurone, où le calcul effectué est une sommation de PPS ayant un decours temporel précis. On ne s'intéresse pas ici aux mécanismes sous-jacents à la génération des PPS, ni à l'émission d'un PA, seul le comportement du neurones est modélisé. Le modèle exprime le potentiel de membrane  $u$

au temps  $t$  comme la somme de deux termes : le premier terme correspond à la contribution des PPS arrivés au soma du neurone avant le temps  $t$ . Le second terme exprime la contribution des périodes réfractaires rencontrées par le neurone avant le temps  $t$ . On définit  $\mathcal{F}_j = \{t_j^{(f)}; 1 \leq f \leq n\} = \{t \mid u_j(t) = \vartheta \wedge u'_j(t) > 0\}$  comme l'ensemble des instants d'émission du neurone  $N_j$ , et  $\Gamma_j = \{i \mid N_i \text{ est en amont de } N_j\}$  l'ensemble des neurones présynaptiques. L'état du neurone  $N_j$  à  $t$  est alors défini par :

$$u_j(t) = \sum_{t_j^{(f)} \in \mathcal{F}_j} \omega_{ij} \eta \left( t - t_j^{(f)} - d_{ij}^{ax} \right) + \sum_{i \in \Gamma_j} \sum_{t_i^{(f)} \in \mathcal{F}_i} \omega_{ij} \epsilon \left( t - t_i^{(f)} - d_{ij}^{ax} \right) \quad (4.3)$$

où  $d_{ij}^{ax}$  est le délai de transmission axonal et  $\omega_{ij}$  est le poids de la synapse, entre les neurones  $N_i$  et  $N_j$ , avec les fonctions noyau suivantes :  $\eta$  est négative pour  $s > 0$  et modélise la remise à zero du potentiel après l'émission d'un PA et  $\epsilon$  décrit la forme des PPS. On considère que tous les PPS ont la même forme, la variation provenant du fait que  $\omega_{ij}$ , selon la nature excitatrice ou inhibitrice de la synapse, peut être positif ou négatif, et fait varier l'amplitude du PPS.

On a alors émission d'un PA lorsque :

$$t_j^{(n+1)} = t \iff u_j(t) = \vartheta, \text{ avec } \vartheta \text{ la valeur du seuil} \quad (4.4)$$

On prend habituellement, pour le neurone  $N_j$ , les expressions suivantes pour les fonctions noyau :

$$\eta(s) = -\vartheta \exp \left( -\frac{s}{\tau} \right) \mathcal{H}(s), \text{ ou } \eta_j(s) = -\eta_0 \exp \left( -\frac{s - \delta^{abs}}{\tau} \right) \mathcal{H}(s - \delta^{abs}) - K \mathcal{H}(s) \mathcal{H}(\delta^{abs} - s) \quad (4.5)$$

où  $\mathcal{H}$  est la fonction de Heaviside,  $\vartheta$  est la valeur de seuil et  $\tau$  une constante de temps, ou  $K \rightarrow \infty$  assure une période réfractaire absolue  $\delta^{abs}$  et  $\eta_0$  définit l'amplitude de la période réfractaire relative.

$$\epsilon(s) = \frac{s}{\tau_s} \exp \left( -\frac{s}{\tau_s} \right) \text{ ou } \epsilon(s) = \left[ \exp \left( -\frac{s}{\tau_m} \right) - \exp \left( -\frac{s}{\tau_s} \right) \right] \mathcal{H}(s) \quad (4.6)$$

où  $\tau_m$  et  $\tau_s$  sont des constantes de temps. Ces deux dernières fonctions sont appelées des fonctions  $\alpha$ .

Une variante plus simple du modèle, appelée le modèle  $SRM_0$ , ne prend en compte que le dernier temps d'émission  $\hat{t}_j$  du neurone  $N_j$  pour la période réfractaire. On remplace alors la somme dans la formule (4.3) par une contribution unique :

$$u_j(t) = \sum_{i \in \Gamma_j} w_{ij} \epsilon(t - \hat{t}_i - d_{ij}^{ax}) + \eta_j(t - \hat{t}_j) \quad \text{avec émission d'un PA si } u_j(t) = \vartheta \implies \hat{t}_j = t \quad (4.7)$$

L'intérêt du modèle par rapport aux modèles temporels évoqués dans les sections précédentes est qu'il ne fait pas appel à la résolution d'équations différentielles couplées. Malgré sa simplicité, le modèle SRM avec délais possède la propriété de détection de corrélation temporelle des PA afférents, surpassant en cela les modèles de neurones à

seuil et sigmoïdaux (Maass, 1997b). D'autre part, (Gerstner et Kistler, 2002) démontrent qu'il permet de reproduire le comportement du neurone HH.

### *Autres modèles*

D'autres modèles, intermédiaires entre le neurone HH et le neurone LIF, apparaissent dans la littérature. On citera le modèle de Hindmarsch et Rose (1984), qui permet de reproduire des propriétés de *bursts* observés au niveau du thalamus chez le rat avec trois équations linéaires couplées. Plus intéressant est le modèle relativement récent de Izhikevich (2004), basé sur deux équations différentielles couplées où interviennent quatre paramètres. Selon les valeurs prises par ces quatre paramètres, le modèle peut reproduire pratiquement tous les comportements observés dans les neurones biologiques. Ce modèle est notamment le seul à reproduire le phénomène de PA post-inhibition, un comportement observé chez les cellules mitrales du rat, qui émettent un potentiel d'action en réaction à l'arrivée d'une inhibition, et sans excitation ultérieure.

## **Modèles de synapse temporelle**

Le modèle Bienenstock et al. (1982) (BCM) explique comment la PLT et la DLT (voir section 2.2.2) peuvent être activées en fonction des taux de décharge des neurones pré- et postsynaptiques. Si le produit des deux taux de décharge est en-dessous d'une certaine valeur seuil, alors la synapse est dépréciée. Si le produit est au-dessus de ce seuil, alors la synapse est potentiée. La grande nouveauté du modèle BCM a été d'ajouter la possibilité de faire varier ce seuil, avec une valeur qui dépend de manière non linéaire de l'activité postsynaptique.

En ce qui concerne la STDP (section 2.2.2), Gerstner et al. (1996) propose un modèle basé sur la différence entre les émissions des neurones pré- et postsynaptiques, par l'intermédiaire d'une fenêtre temporelle.

La modélisation de cette forme de plasticité implique un stockage des derniers temps d'émissions pré- et postsynaptiques, et une mise à jour du poids de la synapse à chaque émission, du neurone présynaptique ou du neurone postsynaptique (Nowotny et al., 2003; Izhikevich et al., 2004). D'autres modèles prennent en compte des influences plus complexes, tels que des triplets de PA (Kepecs et al., 2002; Pfister et Gerstner, 2006), ou les influences de l'ensemble de la séquence de PA (Morrison et al., 2007). Enfin, la STDP permet de maximiser la transmission d'information entre les neurones (Chechik, 2003; Toyozumi et al., 2007).

Les équations de la STDP ont été mises en relation avec le modèle BCM (Izhikevich et Desai, 2003). Kempter et al. (1999) et Gerstner et Kistler (2002) montrent que la STDP est une extension de la loi de Hebb, où les temps d'émissions des PA sont pris en compte (voir également Sejnowski (1999)). D'autre part, les équations de la STDP permettent d'implémenter un mécanisme utilisé dans l'apprentissage par renforcement (Rao et Sejnowski, 2001) (voir section 4.1.1). Enfin, la fenêtre temporelle de STDP permet de reproduire, en utilisant une variable unique, le mécanisme existant du fait des interactions entre la quantité de neurotransmetteurs et la quantité de récepteurs disponibles (Senn

et al., 2001).

## Réseaux de neurones temporels

Tous ces modèles de neurones temporels existent depuis de nombreuses années. Cependant, à la fin des années 1990, les ordinateurs sont devenus suffisamment puissants pour réaliser des simulations impliquant plusieurs centaines ou milliers de ces neurones, permettant l'étude de réseaux de neurones impulsionnels. Dans un premier temps, plusieurs auteurs ont cherché à reproduire les RNA classiques avec des réseaux de neurones temporels. Ainsi, Gerstner et van Hemmen (1992) définissent une mémoire associative basée sur les neurones impulsionnels pour l'apprentissage de motifs, modèle semblable au réseau de Hopfield. Maass (1997a) montre qu'il est possible d'émuler en codage temporel des RNA classiques avec des réseaux de neurones impulsionnels. D'autre part, la capacité de discrimination des réseaux de neurones temporels surpasse celle des réseaux basés sur les anciens modèles de neurones, notamment du fait de l'utilisation des délais de transmission entre les neurones. Selon (Maass, 1997b), les neurones impulsionnels constituent la troisième génération de neurones artificiels.

Un modèle spécifique de réseaux de neurones impulsionnels est apparu au début des années 2000. D'une part, la machine à états liquides ("liquid state machine", Maass et al. (2002)), d'autre part le réseau à états résonnants ("echo-state network", Jaeger et Haas (2004)), ont été définis indépendamment, mais sont basés sur un principe similaire. On regroupe sous le nom de *reservoir computing* le principe de fonctionnement commun à ces deux modèles fondateurs (Schrauwen et al., 2007). Dans ce type de réseau, la connectivité est clairsemée (voir annexe A.4), et les connexions entre les neurones sont définies aléatoirement. Le modèle se compose d'une entrée, par laquelle les motifs sont injectés, d'une partie centrale, correspondant à un sac de neurones (également appelé **réservoir**), et d'une sortie, permettant de lire l'état du réseau. Seules les connexions entre le sac et les sorties sont apprises, en fonction de l'erreur réalisée par le réseau vis-à-vis de la réponse attendue. Un des intérêts de cette approche est de permettre l'apprentissage de séquences de motifs, puisque la sortie du réseau peut être réinjectée dans la partie centrale du modèle, lors de la présentation du motif suivant.

Les modèles de réseaux de neurones temporels évoqués ici correspondent à un fonctionnement plus proche des neurosciences, du fait de la prise en compte explicite du temps dans le modèle de neurone. Ils ne se défont cependant pas de la critique majeure vis-à-vis de la modélisation des phénomènes observés en neurosciences, qui concerne le problème de la supervision du modèle par la définition d'une erreur attendue (voir section 4.1.1). Nous verrons par la suite (section 4.2.6) qu'une supervision par l'intermédiaire d'un algorithme évolutionniste, dans lequel on indique jamais explicitement quelles sont les réponses attendues, permet d'avoir une forme d'apprentissage plus biologiquement plausible.

Des modèles de réseaux de neurones temporels ont été explicitement définis pour simuler le phénomène de synchronisation neuronale observée en biologie (voir section 3.3).

## 4.2 Modèles de synchronisation neuronale

Le phénomène de synchronisation a été étudié dans le cadre de la physique théorique. Le premier exemple date de Huygens, qui étudiait la synchronisation des pendules, en 1673. D'autres applications ont vu le jour, par exemple dans l'étude des oscillateurs en électronique, des phénomènes ondulatoires en transmission du signal (par exemple pour les antennes) ou des ressorts en mécanique des solides.

### 4.2.1 Synchronisation d'oscillateurs

En biologie, la synchronisation d'oscillateurs est étudiée pour expliquer la coordination des cellules cardiaques, la synchronisation des lucioles, la production et la lyse d'enzymes en biochimie, etc.

Une des premières hypothèses pour la génération d'oscillations électriques au niveau du système nerveux est que les neurones sont responsables, du fait de leurs propriétés physiologiques, de la génération d'un rythme intrinsèque : certains neurones génèrent des oscillations dites "sous le seuil" au niveau de leurs potentiels de membrane (Llinas et al., 1991; Wang, 1993), et peuvent être assimilés à des oscillateurs.

Différents modèles ont été utilisés, dont les modèles à phase variable (Kuramoto, 1984; Mirollo et Strogatz, 1990) et les modèles à phase variable avec délais (Yeung et Strogatz, 1999; Earl et Strogatz, 2003). Un oscillateur  $i$  est décrit par sa phase  $\phi_i$ , variable dans le temps. Ces modèles se basent sur une constante de couplage ( $\epsilon$ ) dans l'équation 4.8, permettant de faire varier le comportement du modèle entre un oscillateur isolé, avec une phase propre  $\Omega_i$  lorsque  $\epsilon = 0$ , et un oscillateur influencé par les différences de phase avec les autres oscillateurs du système lorsque  $\epsilon = 1$ .

$$\frac{d\phi_i(t)}{dt} = \Omega_i + \frac{\epsilon}{N} \sum_{j=1}^N \sin(\phi_j(t) - \phi_i(t)) \quad (4.8)$$

Ces modèles permettent l'étude des propriétés de synchronisation d'un grand nombre d'éléments. On montre que l'émergence d'une population complètement synchronisée peut se faire avec une constante de couplage faible (*weak coupling*), dans le cas où  $\epsilon$  est petit devant 1, mais non nul).

Même si les propriétés physiologiques des neurones jouent un rôle dans la génération d'oscillations, la génération d'un rythme interne n'explique pas la formation d'oscillations fonctionnelles, c.a.d liées à une fonction cognitive particulière. D'autre part, les neurones ayant une période propre, tels les neurones *pace-maker*, responsables des mécanismes de respiration ou des battements cardiaques, correspondent à des mécanismes automatiques et donc ne sont pas directement influencés par la cognition.

### 4.2.2 Modèle des *synfire chains*

Abeles (1982, 1991) a introduit le modèle des *synfire chains*. Une *synfire chain* est constituée de neurones impulsionnels répartis dans des couches successives, liées entre elles par des connexions unidirectionnelles. Du fait des émissions quasi-simultanées des

neurones d’une couche, les neurones de la couche suivante vont être activés de manière synchrone, et vont à leur tour propager leurs activités sur les neurones de la couche suivante, et ainsi de suite.

Diesmann et al. (1999) montrent ainsi que l’activité neuronale peut se propager dans un tel réseau sous la forme de paquets de PA (*pulse packet*), correspondant à une activité intermédiaire entre une synchronisation de tous les neurones dans la couche et une activité complètement asynchrone. Ils montrent qu’en partant d’un état avec une dispersion assez importante des temps d’émission de PA, l’activité propagée dans les couches suivantes atteint un attracteur, avec une dispersion beaucoup plus faible dans le paquet de PA. Ainsi, à partir d’un état relativement désynchronisé sous l’effet d’un bruit dans les temps d’émissions de PA, l’activité du réseau peut atteindre un état stable, où les temps d’émissions des neurones se synchronisent.

L’une des limites de ce modèle est son caractère *feed-forward* : la propagation de l’activité se fait d’une couche à l’autre, toujours dans la même direction. L’activité des dernières couches ne peut pas influencer l’activité des premières couches, ce qui, du point de vue de la modélisation du fonctionnement cérébral, n’est pas très plausible.

D’autre part, le caractère *feed-forward* de ce modèle ne permet pas d’expliquer l’émergence d’oscillations liées au phénomène de synchronisation. Afin de compléter cette lacune, Bienenstock (1995) propose un modèle de *synfire chains* superposées, où plusieurs *synfire chains* peuvent se combiner, permettant ainsi aux informations convoyées par différentes chaînes d’interagir, de s’associer les unes aux autres. D’autre part, le modèle suppose l’existence de *synfire chains* “retour”, permettant ainsi la prise en compte par les premières couches de la *synfire chain* de traitements effectués par les dernières. De même, Hertz et Prügel-Bennett (1996) proposent un modèle basé sur une idée similaire, et étudient la capacité d’apprentissage de ces *synfire chains* récurrentes.

L’introduction de boucles dans le modèle des *synfire chains* induit des difficultés dans la régulation de la synchronisation. D’autres modèles ont permis d’étudier les conditions d’émergence d’oscillations liées à la synchronisation neuronale.

### 4.2.3 Polychronisation

En réalité, du fait l’existence de délais de transmission différents entre les neurones, il est difficile d’imaginer comment des PA émis simultanément par différents neurones peuvent impacter simultanément un neurone en aval. Izhikevich et al. (2004) puis Izhikevich (2006) proposent qu’un stimulus soit représenté dans le système nerveux par l’activation d’un **groupe polychrone**, c.à.d par le déclenchement d’une séquence spatio-temporelle de PA, spécifique d’un stimulus. Dans ce modèle, les temps d’activation des neurones ne sont pas synchrones, au sens d’une coactivation simultanée, mais sont corrélés, puisqu’il existe un intervalle fixe entre les temps d’émission des neurones, cet intervalle dépendant des délais de transmission entre les neurones. Paugam-Moisy et al. (2007) utilisent la notion de polychronisation pour expliquer comment fonctionne l’apprentissage dans un sac de neurones impliqué dans un réseau sur lequel est définie une règle d’apprentissage s’apparentant au *reservoir computing* (voir section 4.1.2).



De plus, en imaginant que le système soit bouclé, la reproduction d'un motif spatio-temporel lors de plusieurs itérations successives peut conduire à l'émergence d'oscillations (Izhikevich, 2006). Cet auteur propose une explication à la variabilité de la latence d'un sujet, d'un essai sur l'autre : si l'on considère une oscillation interne, générée même au repos, et si l'on suppose qu'un groupe polychrone est déclenché à partir du pic de l'oscillation, alors un premier facteur de variabilité serait la phase. Un second facteur est le degré d'attention porté au stimulus, qui va déterminer à quel pic les neurones déclencheurs vont mettre en action le groupe polychrone. Si, par exemple, le sujet est concentré sur une tâche, le déclenchement du groupe polychrone se fera de manière plus précoce que si la concentration du sujet s'est relâchée. Cette interprétation est particulièrement intéressante pour expliquer les résultats sur les oscillations en EEG et en MEG (voir section 3.4). Le déclenchement du groupe polychrone à des instants différents en fonction de l'état du sujet au moment où l'essai commence permet d'expliquer l'origine d'oscillations d'induites, de latences variables d'un essai sur l'autre (voir annexe B.2).

#### 4.2.4 Emergence d'oscillations dans une structure particulière

Différentes approches sont possibles pour lier l'émergence d'oscillations au phénomène de synchronisation. Nous nous intéresserons tout d'abord à des modèles permettant de reproduire les phénomènes biologiques observés dans certaines structures particulières, puis des modèles étudiant les conditions d'émergence d'oscillations, par une approche empruntée à la physique statistique, seront abordés.

##### Modèles biologiques

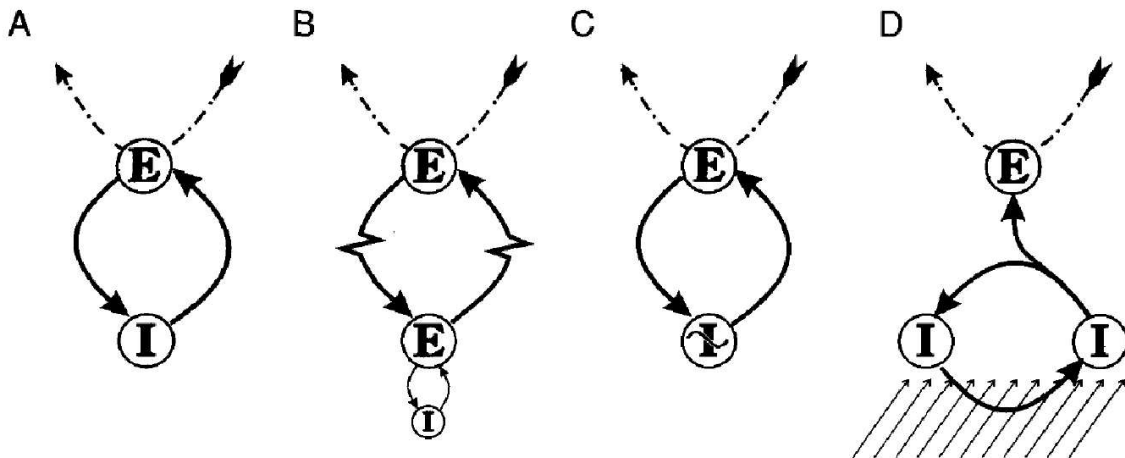


FIG. 4.1 – Différents modèles permettant la génération d'oscillations  $\gamma$ . (A) Inhibition récursive. (B) Excitation mutuelle. (C) Oscillateurs intrinsèques. (D) Inhibition mutuelle. Repris de Jeffery et al. (1996).

Un certain nombre de modèles s'intéressent à la génération d'oscillations - pour la plupart des modèles, oscillations de type  $\gamma$  - dans des structures cérébrales précises, en

particulier l'hippocampe du rat (Wang et Buzsáki, 1996; Whittington et al., 1997; Kopell et al., 2000; Bibbig et al., 2002; Brunel et Wang, 2003), le lobe antennaire de l'insecte (Bazhenov et al., 2001; Laurent et al., 2001) ou le bulbe olfactif du rat (Eeckman et Freeman, 1990; Hendin et al., 1998; Galán et al., 2006).

Ces modèles décrivent les interactions au sein d'une population de neurones (Wang et Rinzel, 1992; Jeffery et al., 1996; Wang et Buzsáki, 1996), ou entre plusieurs populations de neurones ayant des propriétés différentes, classiquement une population de neurones excitateurs et une population de neurones inhibiteurs (Eeckman et Freeman, 1990; Lytton et Sejnowski, 1991; Wang et Buzsáki, 1996; Whittington et al., 1997; Karbowski et Kopell, 2000; Kopell et al., 2000; Bibbig et al., 2002). La figure 4.1 montre différents mécanismes permettant la génération d'oscillations au sein d'une structure particulière.

Un des premiers modèles (Wang et Rinzel, 1992; Wang et Buzsáki, 1996) se base sur les interactions au sein d'une population de neurones inhibiteurs, permettant de reproduire les oscillations générées au niveau de l'hippocampe chez le rat.

A l'aide d'un modèle basé sur deux populations de neurones inhibiteurs et excitateurs, Whittington et al. (1997) proposent un modèle permettant d'expliquer l'émergence d'oscillations  $\gamma$  et  $\beta$  dans CA1, une structure de l'hippocampe. Ils montrent comment l'introduction d'une plasticité synaptique dans le modèle permet de maintenir une oscillation  $\beta$  dans le système lorsque le celui-ci est stimulé, cette propriété étant perdue lors d'épisodes non-oscillants.

Kopell et al. (2000), à partir d'un modèle identique, étudient les propriétés topologiques liées aux oscillations  $\gamma$  et  $\beta$ . Ils montrent que le rythme  $\gamma$  implique une interaction plutôt locale, tandis que le rythme  $\beta$ , plus global, permet de synchroniser plusieurs groupes de neurones, qui seraient indépendants en son absence. Cette simulation va dans le sens du lien entre valeur de la fréquence et taille du réseau impliqué (voir section 3.3.2).

Enfin, Bibbig et al. (2002) étudient de manière systématique quels sont les paramètres des neurones qui permettent la génération d'oscillations  $\gamma$  et  $\beta$ , respectivement. Ils montrent que le rapport entre les délais de transmission et le temps de décroissance des PPS inhibiteurs détermine l'existence d'un rythme  $\gamma$ , ou d'un rythme  $\beta$ .

## Modèles issus de la physique statistique

L'approche champ moyen (*mean field*), inspirée de la physique statistique, permet de connaître le comportement "en moyenne" d'un neurone temporel, plongé dans un réseau, en considérant des influences stochastiques en provenance des autres neurones du réseau. Cette approche permet d'établir une dérivation analytique du comportement d'un modèle de neurone temporel, et de décrire les conditions dans lesquelles il peut y avoir émergence d'oscillations, à partir du comportement d'une population de neurones. Les modèles étudiés dans ce domaine ne se basent pas nécessairement sur la modélisation d'une structure neurobiologique précise; ils s'intéressent plutôt aux conditions d'émergence d'oscillations en fonction des paramètres du modèle de neurone considéré.

Les premières études en champ moyen à partir du modèle LIF (Tsodyks et al., 1993; Sompolinsky et Tsodyks, 1994) ont montré, dans un système où le couplage entre les

neurones est faible (voir la section 4.2.1), que deux comportements pouvaient coexister pour les neurones : un premier comportement correspondant à une activité complètement synchronisée entre les neurones, et un second comportement correspondant à des neurones dont les émissions sont apériodiques. Hansel et al. (1995) étudient les conditions d'émergence d'une synchronisation complète dans un ensemble de neurones excitateurs. Il ressort de cette étude que les influences excitatrices extérieures ont pour effet de désynchroniser le système plutôt que de le synchroniser.

van Vreeswijk et al. (1994) étudient l'influence de connexions excitatrices ou inhibitrices dans des modèles à deux neurones de différents types (oscillateur, modèle LIF ou modèle HH). Ils montrent que, sous certaines conditions, une influence inhibitrice permet une synchronisation entre les émissions des neurones, ce que ne permet pas une influence excitatrice. Ce résultat est intéressant, car on pourrait penser intuitivement que l'excitation joue un rôle plus important que l'inhibition dans la coordination entre neurones.

A la suite de ce résultat, van Vreeswijk (1996) étudie une population de neurones de type LIF, avec une fonction  $\alpha$  pour la forme des PPS. Une influence inhibitrice avec un couplage fort permet une synchronisation partielle avec l'émergence de plusieurs assemblées. Les neurones au sein d'une même assemblée sont synchronisés, les neurones appartenant à des assemblées différentes ne sont pas synchronisés. Ce résultat est intéressant pour l'hypothèse du liage temporel, car il montre à quelles conditions plusieurs assemblées synchronisées peuvent cohabiter dans le système.

Amit et Brunel (1997) étudient l'émergence d'une activité spécifique à un stimulus dans un réseau consistant en une population de neurones excitateurs et une population de neurones inhibiteurs, en utilisant le modèle LIF. D'une part, la formation d'une assemblée capable de se maintenir pendant un certain temps après la disparition du stimulus, est basée sur la sélection de délais de transmission entre les neurones. Si la plasticité synaptique, responsable de cette sélection, n'est pas assez forte, l'activité liée au stimulus ne persiste pas. D'autre part, La réponse sélective d'une assemblée de neurones est plus rapide si le réseau a une activité spontanée que si le réseau n'a aucune activité au moment de la présentation du stimulus. Le fait de maintenir une activité spontanée permet au système d'être plus réactif.

Brunel et Wang (2003) reprennent les travaux précédents, en utilisant un modèle de PPS avec une forme biologiquement plausible, pour étudier l'interaction entre une population d'inhibiteurs et d'excitateurs. Ils montrent qu'il est possible d'avoir un rythme rapide ( $\sim 40$  Hz) au niveau de l'ensemble des neurones, tout en conservant pour chaque neurone un rythme relativement lent ( $\sim 2$  Hz), ce qui correspond aux données physiologique au niveau des enregistrements unitaires et PCL. Cette étude permet d'expliquer l'émergence d'un rythme  $\gamma$  au niveau d'une population de neurones, alors que chaque neurone émet des PA de manière irrégulière, avec un rythme beaucoup plus lent.

La dérivation analytique, basée sur les études du champ moyen, permet d'étudier sous quelles conditions émerge une synchronisation totale de tous les éléments. Cependant, la synchronisation de tous les éléments du système ne porte pas plus d'information qu'un état complètement désynchronisé (voir par exemple la critique d'Izhikevich (2006)). D'autre part, même l'émergence de plusieurs assemblées, dont les neurones sont synchronisés entre eux, mais dont les neurones appartenant à deux assemblées différentes sont

désynchronisés, ne correspond pas à la situation de liage temporel. En effet, ces assemblées émergent de toute façon si certaines conditions au niveau de leur paramètres sont réunies. Ils ne sont pas le fait de percepts à séparer, et il est donc difficile de déterminer ce que représente chacune des assemblées.

Enfin, ce type de dérivation analytique permet de prévoir l'ensemble des comportements possibles à partir d'un modèle de neurone donné, mais ne tient pas compte des détails. Par exemple, la définition du couplage faible se base sur une constante de couplage utilisée pour l'ensemble des influences synaptiques, chaque poids synaptique individuel étant défini par une déviation aléatoire vis-à-vis de cette constante. L'approche champ moyen permet d'appréhender de manière globale les mécanismes mis en l'œuvre dans un réseau, mais pas de connaître dans le détail l'historique d'une connexion synaptique donnée. A mon sens, une meilleure approche pour comprendre le fonctionnement cérébral est de faire des simulations, de manière à appréhender également le détail des comportements du réseau.

## 4.2.5 Modèles de liage des traits

### Modèles de liage par convergence

Différents modèles (Hopfield et Brody, 2000, 2001) ont été proposés pour étudier la convergence hiérarchique de l'information vers un neurone unique, responsable de la détection d'un percept particulier (voir section 3.2). Ces modèles se basent sur la spécialisation hiérarchique des aires cérébrales, et permettent ainsi de rendre compte de réponses sélectives à un stimulus dans un laps de temps relativement court (Hopfield et Brody, 2000).

Une autre manière de modéliser l'activation spécifique d'un neurone pour un stimulus donné est d'utiliser un mécanisme de *winner-take-all* (Maass, 2000). Dans ce mécanisme, des inhibitions entre neurones permettent au premier neurone qui émet un PA d'empêcher les autres neurones auquel il est connecté d'émettre des PA à leur tour (Thorpe, 1990). Ce mécanisme a notamment été utilisé pour la détection de visages, où chaque visage provoque le déclenchement d'un PA (Thorpe et al., 2004).

### Modèles de liage temporel

A partir de l'énonciation de l'hypothèse du liage temporel (voir section 3.3), des modèles se basant sur cette hypothèse ont été développés. L'un des premiers modèles, le *neural cocktail party processor*, a été proposé par von der Malsburg et Schneider (1986). Dans ce modèle, les neurones se comportent comme des oscillateurs (voir section 4.2.1), et la présentation d'un stimulus correspond à l'activation d'un sous-ensemble d'oscillateurs, dont les activités sont corrélées entre elles. L'un des aspects novateurs du modèle est l'introduction d'une forme de plasticité synaptique rapide, permettant de sélectionner la représentation d'un stimulus attendu parmi les autres. Même si le neurone artificiel utilisé est relativement simple, ce modèle est le premier à avoir tenté de reproduire un fonctionnement de liage temporel par la corrélation d'activité d'un sous-ensemble d'unités.

Sporns et al. (1991) montrent comment l'application de l'hypothèse du liage temporel permet de discriminer, dans une scène visuelle, une forme en mouvement d'un fond statique. Les activités des neurones correspondant à la forme en mouvement sont corrélées entre elles, tandis qu'il n'existe pas de corrélations entre les activités des neurones correspondant à la forme et les activités des neurones correspondant au fond. L'extension de ce modèle a conduit à la prise en compte des influences thalamo-corticales dans la génération de corrélations entre les neurones (Lumer et al., 1997a,b). Dans ce modèle de synchronisation large-échelle, impliquant des synapses excitatrices et inhibitrices, les rythmes rapides produits par les neurones dépendent en premier lieu des boucles poly-synaptiques (i.e. impliquant plusieurs relais synaptiques) mais aussi des délais de transmission entre les neurones et des constantes de temps synaptiques, particulièrement celle des synapses inhibitrices.

Shastri et Ajjanagadde (1993) proposent un modèle connexionniste basé sur la synchronisation, permettant de définir des règles d'inférence. Dans ce modèle, les unités correspondant à la conclusion d'une règle se synchronisent avec les unités correspondant à la prémisse de cette règle. La nature de ces règles peut être très générale : elles peuvent aussi correspondre à des inférences de bas-niveau (par exemple de la completion de motifs) que des règles permettant de réaliser un raisonnement basé sur le langage. Les auteurs ne font aucune hypothèse sur la nature des unités (neurones ou ensemble de neurones), mais leur système permet de générer un moteur inférentiel, et de réaliser rapidement des inférences en série. Ce modèle montre comment le liage temporel peut permettre à un système connexionniste de manipuler des règles, de la même manière qu'un système symbolique. Le but des auteurs est de résoudre la question soulevée par Fodor et Pylyshyn (1988), à savoir comment un système distribué peut produire des règles symboliques. La réponse apportée au dilemme soulevé par Fodor et Pylyshyn (1988) passe ainsi par l'introduction du concept de liage temporel pour réaliser des associations dynamiques entre les différentes parties de la règle.

Dans Watanabe et al. (2001), les auteurs proposent un modèle pour simuler un phénomène de liage temporel avec des connexions bidirectionnelles entre des cartes représentant des aires visuelles de plus en plus spécialisées. L'utilisation de liens bidirectionnels permet d'avoir des interactions neuronales corrélées, sans avoir recours à des oscillateurs intrinsèques. Les assemblées de neurones sont déterminées par des mesures de connectivité fonctionnelle : la présentation d'un stimulus induit la création d'une assemblée cellulaire dynamique, dont la composition peut varier dans le temps sans être complètement déterminée par la connectivité.

Raffone et van Leeuwen (2003) introduisent un modèle chaotique de neurones temporels pour résoudre le problème de la transitivité de la synchronisation neuronale (voir section 3.3.1). Nous rappelons que ce problème apparaît lorsque deux objets partageant une caractéristique commune doivent être traités simultanément par le système. Pour ce faire, les auteurs introduisent un modèle chaotique de neurone temporel, et montrent sous cette condition qu'un modèle où les neurones correspondant aux caractéristiques non partagées des deux objets (voir figure 3.3) se synchronisent de manière aléatoire avec les neurones correspondant à la caractéristique commune permet de résoudre le problème de

la transitivité de la synchronisation.

Les modèles évoqués dans la section 4.2.4 font la plupart du temps intervenir une population de neurones excitateurs en interaction avec une population de neurones inhibiteurs. En restant au niveau des liens entre les populations de neurones, ces modèles correspondent à un circuit impair, et provoquent l'émergence d'oscillations par régulation de l'activité (voir annexe A.5).

Cependant, la plupart de ces modèles ne peuvent rendre compte de la synchronisation à large-échelle. En effet, la connectivité de ces réseaux est trop simple pour rendre compte de l'interaction entre des sous-populations de neurones, et ainsi rendre compte de la synchronisation entre différentes structures. Le parti-pris de cette thèse a été d'introduire la complexité au niveau du modèle de réseau de neurones par l'intermédiaire de la topologie sous-jacente au réseau de neurones. Cependant, le problème est la régulation d'un tel réseau.

#### 4.2.6 Modèle *a priori* versus *a posteriori*

La plupart des modèles décrits précédemment se basent sur une approche dite *a priori* : en partant des descriptions issues de la biologie, pour les propriétés des neurones et pour l'anatomie, ils tentent de reproduire les propriétés dynamiques observées au niveau macroscopique. Cependant, ces réseaux de neurones modélisent souvent une structure particulière (par exemple le bulbe olfactif, l'hippocampe, le cervelet), et les interactions avec les autres structures, ou en provenance des entrées sensorielles, ne sont modélisées que par des signaux aléatoires (par exemple des trains de spikes obéissant à une distribution de Poisson). Or, il y a une certaine incohérence à modéliser les influences internes de la structure de manière déterministe et les influences extérieures à cette structure de manière probabiliste. Qu'est ce qui justifie cette différence de nature entre les signaux ?

Une des réponses est la difficulté à faire fonctionner ensemble plusieurs structures. Si dans un premier temps chaque sous-partie est réalisée indépendamment, le fait de les mettre ensemble dans un même système induit des difficultés. En effet, les modules n'ont aucune raison de fonctionner ensemble, puisque leurs paramètres ont été réglés indépendamment, de manière ad-hoc. La mise en commun de ces modules suppose de réaliser un réglage fin, empirique, des différents paramètres, en réalisant de nombreux essais.

Une autre approche, l'approche *a posteriori*, suppose que la communication entre structures est une propriété émergente. En effet, c'est le système, dans son ensemble, qui doit fonctionner. Il existe de nombreuses manières de réaliser cette optimisation non-supervisée (simplexe, descente en gradient). Une des manières de réaliser cette émergence est de faire appel à une méthode évolutionniste, qui permet de reproduire de manière biologiquement plausible la construction d'un fonctionnement global du système nerveux.

En effet, dans la nature, cette communication entre sous-systèmes relativement complexes est possible car les modules ne sont pas construits indépendamment, mais en relation les uns avec les autres, au fur et à mesure de l'évolution. C'est l'individu dans sa totalité qui est sanctionné par son environnement, les sous-modules étant forcés de coopérer pour produire un comportement cohérent au niveau global (c.à.d. adapté au

monde, du point de vue de la sélection naturelle). Ainsi, une modification dans un sous-système ne sera propagée vers les générations ultérieures (c.à.d acceptée par la sélection naturelle) que si elle constitue un avantage pour l'individu qui l'apporte, sans induire d'effets négatifs dans les autres sous-systèmes.

La problématique de la modélisation de la synchronisation neuronale a été peu abordée jusqu'à présent. Or la synchronisation neuronale et l'adaptation d'un individu à son environnement se placent à la même échelle : au niveau de l'individu. Le parti pris de cette thèse est de faire **émerger** la synchronisation neuronale, en simulant l'évolution.

Nous introduisons dans le chapitre suivant la problématique de l'évolution. Nous introduirons les concepts issus de la biologie, puis différents modèles permettant de simuler les processus d'évolution, et l'implication de l'évolution dans l'étude des processus cognitifs.