Chapitre 6

MODELISATION ET CONTROLE ADAPTATIF DANS UN RESEAU ATM AVEC LE RNA

6.1 Introduction

Jusqu'à présent, nous nous sommes bornés à prédire et à modéliser des suites temporelles sans jamais faire usage de leurs propriétés caractéristiques, au demeurant souvent inconnues *a priori*. Nous quittons désormais le domaine de la prédiction pour nous attacher à la **modélisation** d'un réseau de communication avec un réseau de neurones récurrent. Au-delà de la modélisation, l'objet de cette étude est d'optimiser la gestion des ressources du réseau de communication soumis à des trafics *multimédia* à flot sporadique. Il s'agit là d'un problème particulièrement vaste auquel une thèse ne suffirait sans doute pas. Aussi, notre intention n'est pas d'apporter une réponse définitive mais d'inaugurer une approche ambitieuse visant à reproduire le comportement global du réseau. Le système que nous étudions est un réseau à intégration de services large-bande à commutation asynchrone dite ATM (Asynchrone Transfert Mode).

Dans un sens, la modélisation d'un réseau de communication se situe aux antipodes de la prédiction pure : elle exige la connaissance détaillée du système pour définir les caractéristiques du réseau connexioniste (i.e. sa fonction d'activation, ses connexions) et pour élaborer sa configuration (i.e. la connectivité). En revanche, nous retombons dans les ornières de la prédiction dans la mesure où nous visons, au moyen de la modélisation, à prévoir la congestion avant qu'elle ne survienne. En ces termes, le contrôle adaptatif est un domaine connexe à la prédiction. Il est donc naturel d'avoir recours à des techniques adaptatives capables d'anticiper l'évolution future des systèmes dynamiques et donc de commander ces derniers de façon à réduire une fonction coût. Certains résultats phares de la théorie des files d'attente suggèrent l'emploi d'une fonction rationnelle du premier ordre pour modéliser des tampons à l'équilibre stationnaire. C'est la raison pour laquelle nous délaissons le neurone formel de McCulloch et Pitts pour le neurone aléatoire de Gelenbe [Gelenbe 89b] décrit au § 2.8.

En vue de préserver le réseau des risques de *congestion* et de maintenir la *qualité de service* requise par les utilisateurs, un contrôle d'accès au réseau ATM des communications en attente est nécessaire, dont le terme consacré en anglais est Call Admission Control (CAC). Ce rôle d'arbitrage incombe au neuro-contrôleur qui, sur la base des paramètres déclarés par les souscripteurs, relaxe le RNA pour obtenir une évaluation de l'état futur du réseau ATM. Si cet état dévie anormalement des conditions normales d'utilisation ou que la qualité de service estimée n'est pas respectée pour les nouvelles communications, les requêtes d'appel jugées responsables sont rejetées.

Le positionnement du problème - La théorie des files d'attente s'applique avec succès à la modélisation du réseau téléphonique parce que les flots de trafic acheminés vers un commutateur peuvent être représentés fidèlement par un processus de Poisson en raison de l'indépendance des trafics. Néanmoins, les files d'attente qui s'établissent dans les tampons au sein des commutateurs ATM ont une nature statistique bien différente (au niveau appel) de celles du trafic téléphonique. Parce que les données sont transmises en *rafale*, les arrivées successives sont fortement corrélées et ne peuvent plus par conséquent être modélisées par une processus de Poisson. En raison de la nature irrégulière des flots de données, la commutation ATM donne naissance à de nouveaux problèmes en théorie des files d'attente qui ne peuvent être résolus avec les outils mathématiques conventionnels. Aussi, l'absence d'outils théoriques *ad hoc* motive l'emploi de méthodes adaptatives ne nécessitant aucune connaissance au préalable du comportement statistique des files d'attente. Dans cet esprit, il semble raisonnable d'avoir recours aux réseaux de neurones dans l'espoir d'appréhender l'influence complexe des trafic sporadiques sur les ressources du réseau.

Le contrôle d'admission des appels (CAC) est de loin la fonction la plus importante pour protéger le réseau de la congestion, dans la mesure où les trafics responsables de l'excédent de charge seront simplement rejetés. Un contrôleur adéquat doit être capable de s'adapter avec souplesse à la nature transitoire de l'état du réseau afin de prévoir le résultat *a priori* de la superposition des trafics sur la charge du réseau et donc des files d'attente qui jalonnent le réseau. Lorsque une demande de connexion est formulée par un ensemble de paramètres décrivant la nature statistique du trafic généré par la source, le réseau accède à cette requête ou la rejette au vu de l'estimation de la quantité de ressource que nécessite cette connexion.

Les réseaux connexionistes ont été récemment employés [Chien 91, Tran-Gia 91, Hiramatsu 90] avec l'intention d'approximer la relation qui existe entre, le nombre de requêtes, la nature des sources, et la probabilité de perte dans un multiplexeur. Bien que les résultats obtenus avec cette approche soient satisfaisants, il est clair que le contrôle à l'échelle du réseau entier est une tâche autrement plus ardue. Il est nécessaire, pour ce faire, de connaître la topologie du réseau, la capacité des liens, la technique de commutation, l'état courant des tampons, etc.

A la différence de ces méthodes où le CAC est formulé comme une approximation fonctionnelle statique, nous décrivons dans ce chapitre une méthode d'apprentissage fondée sur plusieurs réseaux de neurones distribués sur les commutateurs et interconnectés via les lignes de communication. L'objectif poursuivi est d'effectuer un contrôle d'admission des appels ainsi qu'un routage optimal des connexions. Avec ce dessein à l'esprit, nous examinons le moyen de modéliser les files d'attente localisées en entrée et en sortie des commutateurs avec trois neurones aléatoires [Gelenbe 89b], dont les activations sont respectivement

1. le nombre moyen de cellule en attente de service.
2. le taux d'utilisation de la file d'attente.
3. une estimation du degré de fluctuation du trafic entrant représenté par le carré du coefficient de variation de la distribution des inter-arrivées.

L'examen attentifs des résultats issus de la théorie des files d'attente à l'état stationnaire fournit des indications utiles sur la manière de connecter ces neurones. Le théorème de Pollazek-Khinchine suggère que la relation fonctionnelle entre les neurones 1 et 2 soit représentée par une fonction rationnelle du premier ordre. On est donc loin du cadre de Pitts et McCulloch et des fonctions sigmoïdales. Il est donc naturel de se tourner vers le réseau neuronal aléatoire de Gelenbe dont nous avons déjà parlé au § 2.8. Il est curieux de noter l'analogie que l'on peut établir entre le modèle RNA associé à un réseau Markovien d'unités stochastiques parmi lesquelles circulent des signaux, et le réseau non-Markovien de files d'attente que l'on cherche à modéliser.

Les réseaux de neurones situés sur les commutateurs sont entraînés à apprendre le comportement stochastique des tampons au niveau des appels. A cette échelle de temps (de l'ordre de la seconde), on peut spéculer sur une variation "relativement" lente et continue des propriétés statistiques des tampons. Aussi, le réseau de neurones s'attachera à anticiper la *variation* plutôt que la valeur exacte de ces grandeurs statistiques. Dans cette optique, chaque neurone reçoit un signal d'*inertie* qui compense l'erreur commise à l'itération précédente. Cela confère à chaque neurone une mémoire à court terme de son état précédent. Cette mémoire bouleverse quelque peu l'algorithme d'apprentissage standard discuté en § 2.8.1. On retrouve par des voies détournées les problèmes que nous nous sommes posés avec le DRNN. Une procédure d'apprentissage est développée dans l'esprit de la rétro-propagation récurrente [Aussem 94, Aussem 95e, Hertz 91, Pineda 87].

6.2 Le réseau ATM

Face à l'explosion des services multimédia, de nouveaux réseaux de communication flexibles doivent être imaginés capables de véhiculer des trafics à débit sporadique. Parmi ces services nouveaux, on compte la télévision haute définition, la vidéo-conférence, la vidéophonie, le transfert rapide de données informatiques, etc. sans parler de services encore inconnus qui ne manqueront pas d'apparaître. De cette floraison de services est née l'exigence d'un réseau "universel" qui permettra de satisfaire aux besoins croissants de ces nouvelles applications multimédia. L'avènement de la fibre optique a suscité d'emblée des améliorations notables de la qualité de la transmission tout en simplifiant les protocoles.

Les organismes internationaux de normalisation ont recommandé une technique de transfert asynchrone dite ATM (Asynchronous Transfer Mode) fondée sur la commutation

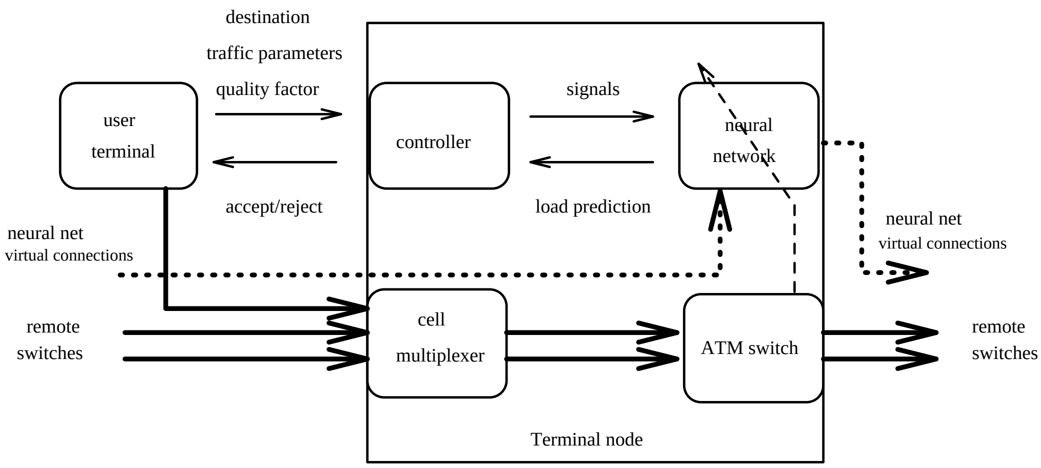

FIG. 6.1 - : Représentation du neuro-contrôleur d'admission d'appels implémenté sur chaque commutateur.

de paquets de petite taille appelés *cellules*. Ce mode de transfert se situe à la confluence de la commutation de circuit et de paquet. Il s'inspire de la commutation de circuit par une absence de recouvrement en cas d'erreur et du contrôle de flux (i.e. à charge des couches supérieures). Enfin, il hérite de la souplesse de la commutation de paquets en raison de la taille fixe des cellules. Cette technique permettra de transporter sur un support unique, la voix, les données et les images à des débits vertigineux. L'ATM repose sur un multiplexage des trafics pour bénéficier des fluctuations statistiques des sources de trafic et utiliser au mieux la bande passante de ces réseaux à haut-débit. La mission de l'opérateur de réseau consiste à assurer le transfert d'informations en respectant un niveau de qualité prédéfini, base d'un contrat préalable avec l'usager. Toutefois, le contrôle du trafic et la gestion des ressources font toujours l'objet de nombreuses études.

6.3 Un modèle de réseau ATM

6.3.1 Les descripteurs de Trafic

Il est essentiel, dans un réseau véhiculant des trafics sporadiques, d'être en mesure de caractériser le processus de génération des cellules pendant un appel, de façon à dimensionner le réseau. Néanmoins, il est devenu excessivement difficile de modéliser, et *a fortiori* de prédire, le comportement statistique d'une kyrielle de services. Il y eu un travail considérable effectué ces dernières années sur les modèles de source et sur le choix adéquat de descripteurs de trafic permettant d'identifier les propriétés essentielles du trafic et de développer de modèles d'évaluation de performance.

Il existe à ce jour plusieurs choix possibles pour décrire un trafic sporadique. L'ensemble des descripteurs doit être très restreint pour permettre un contrôle simple et efficace, tout en représentant fidèlement les propriété des flots de données. De surcroît, il doit être possible de déduire les caractéristiques de la superposition de trafics élémentaires à l'aide des paramètres individuels. Avec ces contraintes à l'esprit, nous nous limitons délibérément à deux grandeurs statistiques fondamentales

- le taux moyen d'arrivée λ.
- le carré du coefficient de variation de la distribution des inter-arrivées, C_a^2.

Le carré du coefficient de variation de la distribution des inter-arrivées est commode dans notre contexte car il peut être relié à l'*index de dispersion* $I(t)$ représentant un descripteur synthétique approprié lorsque le flot de cellules est vu comme un processus ponctuel stationnaire [Cost 91]. Soit $t = 0$ un instant arbitraire, et $N(0,t)$ le nombre des arrivées durant la période $[0,t]$, alors $I(t)$ est définie par

$$I(t) = \frac{\mathrm{var}(N(0,t))}{\mathrm{E}(N(0,t))} \tag{6.1}$$

Pour un processus de Poisson, $I(t) = 1, \forall t$. De plus, pour une processus de renouvellement $I(t)$ n'est autre que le carré du coefficient de variation de la distribution des inter-arrivées, C_a^2. Plus généralement, il est démontré [Cox 66] que $I(t)$ vérifie

$$\lim_{t\to\infty} I(t) = \frac{\mathrm{var}(\tau)}{\mathrm{E}^2(\tau)} = C_a^2 \tag{6.2}$$

où τ est le temps d'inter-arrivée des cellules. En conséquence, on peut facilement calculer et le taux moyen d'arrivée et la variance du nombre des arrivées du flot agrégé en sommant les moyennes et les variances individuelles et déduire le coefficient de variation des inter-arrivées correspondant. Considérons la superposition de n sources indépendantes et $N_i(0,t)$ le nombre de cellules émises par la source i durant la période $[0,t]$. Le nombre des arrivées pour le flot agrégé est

$$N^s(0,t) = \sum_{i=1}^{n} N_i(0,t) \tag{6.3}$$

et donc

$$\begin{aligned} \mathrm{E}[N^s(0,t)] &= \sum_{i=1}^{n} \mathrm{E}[N_i(0,t)] \\ \mathrm{var}[N^s(0,t)] &= \sum_{i=1}^{n} \mathrm{var}[N_i(0,t)] \end{aligned} \quad (6.4)$$

Si bien que

$$\begin{aligned} \lambda_s &= \sum_{i=1}^{n} \lambda_i \\ C^2_{a_s} &= \frac{1}{\lambda_s} \sum_{i=1}^{n} \lambda_i C^2_{a_i} \end{aligned} \quad (6.5)$$

où

$$\lambda_i = \lim_{t \to \infty} \frac{1}{t} \mathrm{E}[N_i(0,t)] \quad (6.6)$$

Il faut noter que la valeur maximale de $I(t)$ pour $t \geq 0$ est considérée comme une alternative à C^2_a [Cost 91] et dans de nombreux cas[1] est atteint pour $t \to \infty$. A la lumière de ces observations, C^2_a est considéré comme un candidat plausible pour la description synthétique des trafics [Cost 91].

6.3.2 Le commutateur ATM

L'évaluation de la performance d'un contrôleur d'admission d'appel dépend fortement du modèle de réseau employé et, donc inévitablement de la manière dont sont commutées les cellules. Un commutateur ATM d'un seul étage comportant des tampons en entrée et en sortie est considéré comme un modèle viable capable d'émuler de commutateurs non-bloquants [Beylot 93, Oie 92] (e.g. les réseaux d'interconnexion Batcher Banyan, de Clos, de Benes, les réseaux Omega, etc.) dès lors que la *capacité de commutation* est suffisamment élevée [Liadis 90]. La capacité de commutation, notée C, désigne le nombre maximal de cellules en contention qu'il est possible de transmettre à un même port de sortie pendant l'unité de temps (temps de réception d'une cellule). L'analyse d'un tel commutateur a été menée pour $N \to \infty$ [Hui 87, Liadis 90, Oie 92], avec un routage équiprobable et trafic Poissonnien. Les résultats obtenus démontrent que des tampons de taille modeste en sortie suffisent à augmenter le débit maximal en entrée du commutateur à sa limite théorique [Liadis 90].

Considérons un commutateur de taille $N \times N$ représenté Fig. 6.2, assurant l'acheminement des cellules par voie logique [Coudreuse 90]. Parce que la mémoire au sein du

[1] par exemple lorsque les durées des *burst* et des silences ont une distribution exponentielle, ou lorsque la durée des *burst* est déterministe et la durée des silences est de distribution exponentielle.

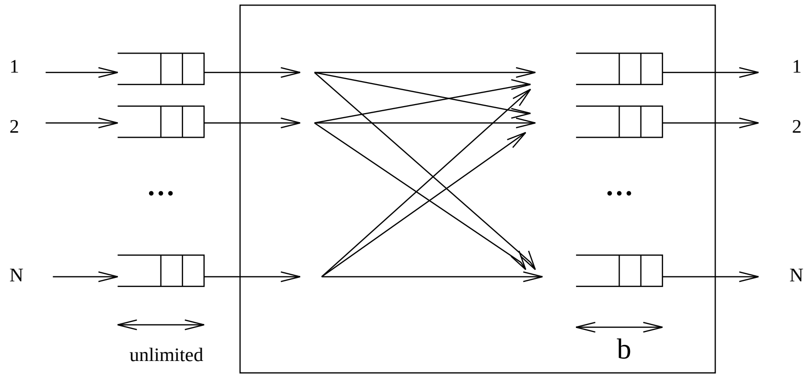

FIG. 6.2 - : Représentation d'un commutateur ATM $N \times N$ avec des tampons en entrée et en sortie.

commutateur est une ressource critique, la taille mémoire des tampons en sortie est limitée à b. En revanche, les tampons en entrée ne possèdent pas de propriété multi-port et peuvent être considérés de taille illimitée, et situés à l'extérieur du commutateur lui-même. Afin de résoudre les contentions d'accès aux ports de sortie[2] et éviter les pertes de cellules, un mécanisme retarde la transmission des cellules en entrée selon la politique du premier arrivé, premier sorti (FIFO). Lorsque $M \leq C$ cellules atteignent la tête de leurs files respectives, elles sont aussitôt envoyées vers la tampon de sortie si ce dernier est libre. Maintenant, si la file de sortie ne possède pas la mémoire requise ou si $M > C$, certaines cellules seront retardées en entrée. Dans ce cas de figure, les cellules retardées demeurent en tête de file et bloquent les cellules suivantes.

L'efficacité d'un commutateur se fonde sur deux mesures fondamentales : le taux de pertes et le débit maximal en entrée. Puisqu'il n'y a pas de pertes dans ce modèle, seule le débit et donc les délais de commutation nous intéressent. Ce dernier se décline en trois composantes

1. le temps d'attente dans le tampon d'entrée avant le placement en tête de file.
2. le temps d'attente en tête de file du à la contention de sortie (HOL) + le temps de transmission.
3. le temps d'attente dans le tampon de sortie avant la transmission.

Lliadis et Denzel [Liadis 90] ont montré que pour une valeur de N asymptotique, un routage uniforme et un trafic Poissonnien de paramètre λ dans chaque port d'entrée, le processus des arrivées à un tampon de sortie s'approxime par une processus de renouvellement de taux λ/N. La superposition de ces trafics vers un processus de Poisson de taux λ tend lorsque $N \to \infty$ d'après le théorème de Palm-Khinchine [Kleinrock 76]. En

[2]Lorsque plusieurs cellules $M \leq N$ arrivent de différentes entrées simultanément et visent une destination identique se pose un problème d'arbitrage quand $C < M$ où C représente la capacité de commutation. Le blocage qui en résulte porte le nom de *Head of the Line Blocking* (HOL).

conséquence, la gigue de commutation peut être ramenée au temps d'attente dans deux files M/G/1 et M/D/1 en cascade. La première file est associée au délai 1, tandis que la seconde représente les délais 2 et 3.

6.3.3 La modélisation neuronale des files d'attente

Alors que, dans un contexte idéal, les délais de transmission au sein d'un commutateur ATM se déterminent par l'analyse du taux d'occupation de deux files d'attente en cascade, aucune conclusion ne peut être tirée pour un trafic non-Markovien au routage non-symétrique. Néanmoins, il est légitime de supposer que le comportement réel des files dévie "raisonnablement" de celui des files M/G/1.

Le réseau de neurones destiné à modéliser le commutateur, s'adaptera aux trafics non-Markoviens, aux routages évolutifs ainsi qu'à une capacité de commutation quelconque. Dans cette perspective, il convient d'examiner quelques résultats éculés de la théorie des files d'attente pour déterminer le modèle connexioniste le plus approprié.

Le processus de génération des cellules est décrit par deux paramètres, le taux moyen des arrivées λ et le carré du coefficient de variation de la distribution des inter-arrivées, C_a^2. Soit b la taille des tampons de sortie et C_{norm}^2 une constante de normalisation. Une file d'attente peut être représentée par trois neurones associés à $\overline{N}$ (le nombre moyen de cellules en attente), ρ (la taux d'utilisation) et C_a^2 (appelés 1,2 et 3 par la suite), i.e. leur activation est interprétée comme la valeur normalisée de ces valeurs dans l'intervalle $[0, 1]$. Soulignons que $\overline{N}$ n'est pas exactement le nombre moyen de cellules en attente dans la tampon mais dans la file d'attente équivalente dont nous venons de parlé. Il s'agit en entrée des cellules en attente avant le HOL, et en sortie, des cellules en attente d'être convoyées vers cette sortie plus celles qui sont déjà dans la file. Par souci de clarté, cette légère déviation du sens intuitif sera occultée dans la suite.

Les valeurs désirées sont donc données par

$$\hat{q}_1 = \begin{cases} \frac{\overline{N}}{b} & \text{si } \overline{N} \leq b \\ 1 & \text{sinon} \end{cases} \qquad \hat{q}_2 = \rho \quad \hat{q}_3 = \begin{cases} (\frac{C_a}{C_{norm}})^2 & \text{si } C_a \leq C_{norm} \\ 1 & \text{sinon} \end{cases} \tag{6.7}$$

Nous abordons le problème de l'interaction de ces neurones de façon à refléter le plus fidèlement possible les dépendances stochastiques entre ces grandeurs. Nous sommes mus par la nécessite de trouver une connectivité judicieuse avec un nombre de connexions minimal. La formule notoire de Pollazek-Khinchine dédiée aux files M/G/1 s'écrit

$$\overline{N} = \frac{\rho - \alpha}{1 - \rho} \quad \text{où} \quad \alpha = \frac{\rho^2(1 - C_b^2)}{2} \tag{6.8}$$

où C_b^2 est le carré du coefficient de variation de la distribution des intervalles de départ. Notons que α est positif pour les files M/D/1, nul pour les files M/M/1, et sinon indéterminé. Ce résultat est remarquable en ce sens qu'il exprime le nombre moyen de cellules en attente comme une fonction rationnelle du taux d'utilisation de la file. Au vu de ce résultat, il n'est pas réaliste de vouloir représenter ces neurones avec une fonction sigmoïde. D'autant plus qu'on peut aisément définir des relations qualitatives entre les neurones des files distantes si ces dernières sont reliées entre elles (e.g. si la première file est congestionnée, la seconde l'est probablement aussi, etc.). Ces observations nous amènent à choisir le RNA avec lequel il est possible d'attribuer des interactions du type excitation et/ou inhibition.

6.4 Le Modèle RNA

Eq. 6.8 plaide en faveur d'une relation fonctionnelle rationnelle entre les neurones 1 et 2. L'analyse intuitive des influences mutuelles entre les tampons permettra en outre de sélectionner des connexions positives ou négatives en fonction des ses interactions.

Les paramètres du système dans notre étude sont déterminés par la relaxation d'équations différentielles du premier ordre du type Cohen-Grossberg

$$\tau \frac{dq_i}{dt} = -q_i + \frac{\sum_{j=1}^{N} w_{ji}^{+} q_j + \lambda_i}{\sum_{j=1}^{N} w_{ji}^{-} q_j + \Lambda_i + \sum_{j=1}^{N} w_{ij}^{+/-}} \tag{6.9}$$

Il est clair que le point fixe[3] vérifie $dq_i/dt = 0$, soit

$$q_i = \frac{\sum_{j=1}^{N} w_{ji}^{+} q_j + \lambda_i}{\sum_{j=1}^{N} w_{ji}^{-} q_j + \Lambda_i + \sum_{j=1}^{N} w_{ij}^{+/-}} \tag{6.10}$$

Le seul piège pernicieux auquel il faut porter une attention particulière est la stabilité du réseau. Le système (6.9) converge toujours vers un point fixe unique mais le réseau n'est stable que lorsque la solution vérifie $q_i < 1, \forall i$. Sinon le réseau est instable et l'amplitude des signaux λ (resp. Λ) doit être diminuée (resp. augmentée).

6.4.1 La matrice de poids

Les similarités évidentes entre (6.10) et (6.8) suffisent à légitimer l'emploi du RNA. Toutefois, il faut poursuivre l'exploration des résultats issus de la théorie des files d'attente pour établir une matrice de poids convenable. Puisque notre objectif est d'appréhender les risques de congestion, de telles situations ne se produisent que lorsque la charge du réseau a déjà atteint un seuil élevé. Il est crucial que notre modèle se comporte correctement dans cette situation extrême. Heureusement, une approximation facilement interprétable concerne les files G/G/1 lorsque $\rho \simeq 1$,

[3] l'existence et l'unicité du point fixe sont discutés au § 2.8.

$$W \simeq \frac{(\sigma_a^2 + \sigma_b^2)}{2\bar{t}(1-\rho)} \quad \text{et} \quad \overline{N} \simeq \lambda W + \rho = \frac{C_a^2}{2(1-\rho)} + \frac{\lambda \sigma_b^2}{2\bar{t}(1-\rho)} + \rho \tag{6.11}$$

où W est le temps d'attente moyen, σ_a^2 (resp. σ_b^2), la variance de la durée des inter-arrivées (resp. le temps de service), et $\bar{t}$ le temps moyen de la période d'inter-arrivée. En d'autres termes, la dépendance de $\overline{N}$ à C_a^2 devient linéaire lorsque ρ est élevé, via un facteur qui dépend de ρ.

De (6.7), (6.8) et (6.11), on déduit les relations suivantes

$$\hat{q}_2 = \frac{b\hat{q}_1 + \alpha}{b\hat{q}_1 + 1} \quad \text{et} \quad \frac{\partial \hat{q}_1}{\partial \hat{q}_3} \simeq \frac{C_{norm}^2}{2b(1-\rho)} \tag{6.12}$$

Si on compare (6.12) avec (6.10), il vient que les neurones 1 et 2 doivent être reliés par deux connexions de poids b et que $\Lambda_1 = \alpha$ pour la file M/D/1 pour laquelle $C_b = 0$. Cependant α peut être négatif pour une file M/G/1, c.a.d. $C_b > 1$, si le temps de service est plus irrégulier que ne l'est un temps de service exponentiel pour lequel $C_b = 1$. Comme les signaux ne peuvent être négatifs, la valeur négative de α doit être compensée par un signal opposé Λ_1 au dénominateur dans (6.10). D'après (6.10) et (6.12), les poids initiaux vérifient

$$w_{12}^+ = w_{12}^- = b\,, \quad w_{31}^+ = \frac{C_{norm}^2}{(1-\rho)} \tag{6.13}$$

Une difficulté se pose du fait que w_{13}^+ et Λ_1 (ou λ_1) dépendent du taux d'occupation ρ, lequel fluctue au gré du trafic en entrée du tampon. Toutefois, nous allons montrer que les variations de ρ sont relativement continues de sorte que l'algorithme d'apprentissage puisse adapter les poids selon l'évolution des files. En effet, le cycle d'apprentissage (inférieur à la seconde) est petit au regard de la durée moyenne d'établissement des connexions ($\simeq$ 100s) si bien que la proportion des nouveaux appels dans le trafic multiplexé a une influence négligeable. Par ailleurs, nous supposons que les flots individuels des données émises en rafale se compensent en moyenne dans le multiplexeur lorsqu'on considère le trafic agrégé. Une valeur arbitraire $\rho = 0.8$ est choisie dans (6.13) pour déterminer les paramètres initiaux.

Bien que cette représentation neuronale s'accorde avec les files M/G/1 dans une certaine mesure, il faut observer que plus C_b s'écarte de 1, et moins bon sera le modèle car α devra être estimé.

Interconnexions entre les files - Nous pouvons intuitivement étendre cette analyse à l'ensemble du commutateur. Chaque neurone 2 des files en entrée émet un signal excitatoire aux neurones 1 des files de sorties O proportionnel à la probabilité que les cellules transitent par ce chemin. Aussi, fixons leur poids tels que

$$\sum_{j \in O} w_{2j}^{+/-} = 1 \tag{6.14}$$

Par ailleurs, les neurones 3 des files en entrée émettent un signal excitatoire à leurs vis-à-vis en sortie car le comportement sporadique des flots en entrée se répercute sur les tampons de sorties. Les tampons absorbent néanmoins ces variations d'autant plus que le trafic est intense. C'est pourquoi, des signaux inhibiteurs sont introduits des neurones 2 en entrée aux neurones 3 en sortie. Des signaux particuliers doivent refléter l'effet du mécanisme de blocage des cellules en tête de file. Cet effet a pour conséquence d'accentuer les risques de congestion dans les tampons d'entrée. Aussi, des connexions entre les neurones 1 en sortie vers les neurones 1 en entrée sont établies (cf. Fig. 6.3). Une raisonnement identique est appliqué aux files distantes reliées par des lignes de transmission à la différence que les connexions rétrogrades sont omises puisque les cellules sont émises quelque soit la charge des tampons destinataires (cf. Fig. 6.4). Une vision globale du système est donnée Fig. 6.5.

6.4.2 Des signaux d'inertie

Compte tenue de la durée négligeable d'un cycle d'apprentissage en comparaison de la durée moyenne de l'établissement des connexions, nous spéculons sur une comportement relativement régulier du trafic en entrée des tampons. Dans l'intention d'alléger la tâche dévolue au réseau de neurones, il est préférable d'apprendre à anticiper des variations plutôt que les valeurs exactes. Cette simplification se heurte tout de même à un problème : la relation rationnelle du premier ordre que nous avons exhibée n'est plus valide si l'on considère les variations $\Delta\overline{N}$ et $\Delta\rho$. En dépit de cela, il est possible d'employer un artifice en envoyant des signaux spécifiques, appelés signaux d'*inertie* par l'entremise des signaux λ et Λ.

Si $q_i(k)$ et $d_i(k)$ désignent les potentiels réels et désirés à l'instant k; le signal d'inertie transmis au neurone i à l'instant $k+1$ est défini de manière à avoir $q_i(k+1) = d_i(k)$ dès lors que les autres signaux internes et externes demeurent identiques entre k et $k+1$. L'idée qui sous-tend cette méthode est de capturer la dynamique des variations $d_i(k+1) - d_i(k)$. Le signal d'inertie $\lambda_{i,inert}$ (ou $\Lambda_{i,inert}$) à envoyer à chaque neurone dépend des potentiels précédents. Un apprentissage *teacher forcing* est employé [Hertz 91, Williams 89] si bien que tous les $q_j(k-1)$ sont remplacés par leur valeur désirée à l'instant k, $d_j(k-1)$.

Un commentaire sur les notations employées par la suite. Nous désignerons implicitement $q_i(k)$ par q_i et $d_i(k-1)$ par d_i pour ne pas alourdir inutilement les formules.

Définissons

$$q_i^\star \triangleq \frac{\displaystyle\sum_{j=1}^{n} w_{ji}^+ d_j + \lambda_i}{\displaystyle\sum_{j=1}^{n} w_{ji}^- d_j + \sum_{j=1}^{N} w_{ij}^{+/-}} \tag{6.15}$$

où λ_i est non-nul pour les neurones 1 et 3 des files en entrée uniquement. Nous sommes conduit à distinguer deux cas suivant le signe de $q_i - d_i$.

Lorsque $q_i - d_i > 0$, désignons par $\lambda_{i,inert}^+$ le signal d'inertie positif destiné au neurone i défini comme suit

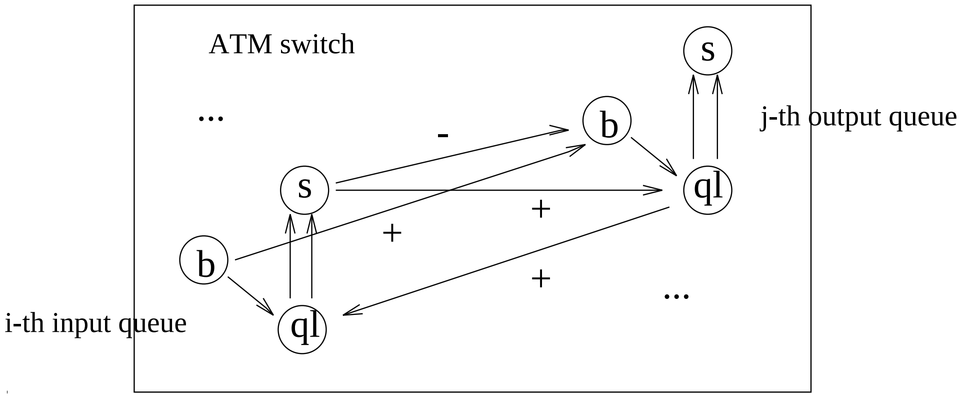

FIG. 6.3 - : Interconnexion des files d'attente modélisées au sein du commutateur.

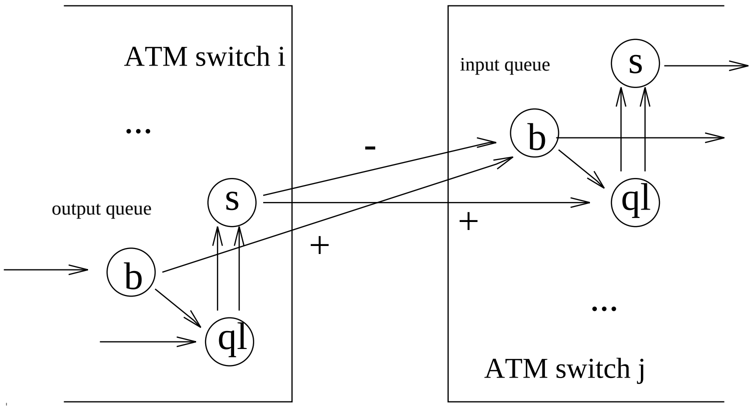

FIG. 6.4 - : Interconnexion des files d'attente modélisées entre commutateurs distants.

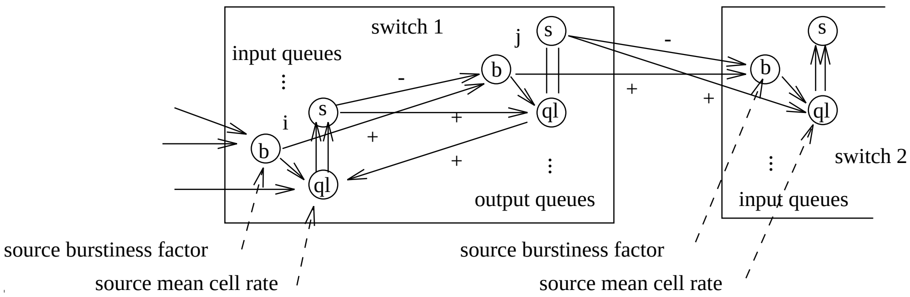

FIG. 6.5 - : Schéma d'interconnexion global.

$$\lambda^+_{i,inert} = d_i \left[\sum_{j=1}^{n} w^-_{ji} d_j + \sum_{j=1}^{N} w^{+/-}_{ij} \right] - \sum_{j=1}^{n} w^+_{ji} d_j - \lambda_i \tag{6.16}$$

de sorte que,

$$d_i = \frac{\sum_{j=1}^{n} w^+_{ji} d_j + \lambda^+_{i,inert} + \lambda_i}{\sum_{j=1}^{n} w^-_{ji} d_j + \sum_{j=1}^{N} w^{+/-}_{ij}} \tag{6.17}$$

Similairement, définissons $\lambda^-_{i,inert}$ lorsque $q_i - d_i < 0$ par

$$\lambda^-_{i,inert} = \frac{1}{d_i} \left[\sum_{j=1}^{n} w^+_{ji} d_j + \lambda_i \right] - \sum_{j=1}^{n} w^-_{ji} d_j - \sum_{j=1}^{N} w^{+/-}_{ij} \tag{6.18}$$

ainsi

$$d_i = \frac{\sum_{j=1}^{n} w^+_{ji} d_j + \lambda}{\sum_{j=1}^{n} w^-_{ji} d_j + \lambda^-_{i,inert} + \sum_{j=1}^{N} w^{+/-}_{ij}} \tag{6.19}$$

$\lambda^{+/-}_{inert}$ vise à compenser l'erreur courante $d_i(k) - q_i(k)$. Le signal est injecté au dénominateur ou au numérateur à l'instant k, en fonction du signe de $d_i(k) - q^\star_i(k)$ de la façon suivante

$$\lambda_{i,inert} = \begin{cases} \lambda^+_{i,inert} & \text{si} \quad d_i - q^\star_i > 0 \\ \lambda^-_{i,inert} & \text{si} \quad d_i - q^\star_i < 0 \end{cases} \tag{6.20}$$

Les neurones sont influencés par les variations des signaux externes au lieu de l'être par les valeurs absolues. Le réseau statique récurrent est désormais commandé par des signaux qui dépendent de l'état passé du réseau ce qui rend la rétro-propagation, telle qu'elle est définie en § 2.8.1, inappropriée.

Le signaux externes - Les trafics émanant des sources externes sont multiplexés, et ensuite expédiés vers les tampons en entrée. Ainsi, chaque tampon en entrée reçoit un trafic multiplexé en provenance d'un commutateur distant et de l'extérieur. Les paramètres statistiques, $\overline{\lambda_i}$ et $\overline{C_{a_i}}^2$, des sources alimentant une file d'entrée particulière sont regroupés selon (6.5) et normalisés. Le résultat est transmis aux neurones sous forme de signaux excitatoires; $\overline{\lambda_s}$, est transmise au neurone 1, et $\overline{C_{a_s}}^2$ est transmis au neurone 3, de sorte que

$$\lambda_{ql} = \Upsilon_{ql} \overline{\lambda_s} \quad \text{et,} \quad \lambda_b = \Upsilon_b \overline{C_{a_s}}^2 \tag{6.21}$$

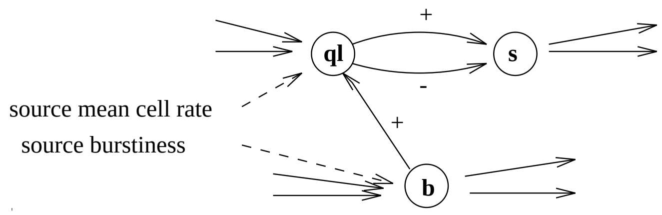

FIG. 6.6 - : Les files d'attente sont modélisées par trois neurones aléatoires. Les flèches en pointillé représentent les signaux représentatifs du trafic multiplexé en entrée des files d'entrée. Les flèches en trait plein désignent les connexion positives ou négatives internes.

Les coefficients Υ_{ql} et Υ_b sont adaptés comme le sont les poids et peuvent être considérés comme des connexions à part entière.

6.4.3 La descente du gradient distribuée

La descente du gradient est réexaminée compte tenu de la dynamique partielle dont nous avons doté le réseau. L'apprentissage des réseaux de neurones implémentés sur les commutateurs est indépendant les uns des autres. Ils ajustent leur propres connexions ainsi que celles qui leurs sont destinées mais ne propagent pas leur propre gradient d'erreur au-delà. Nous supposons que les grandeurs statistiques moyennées des tampons en entrée et en sortie sont disponibles. On pourrait objecter que les connexions reliant des neurones similaires dans les files ont un poids identique. Nous ne ferons pas cas de cette observation car cela reviendrait à simplifier excessivement notre modèle, surtout que les connexions reflètent les caractéristiques courantes du trafic dans chaque file.

Sans perte de généralité, considérons un réseau complètement bouclé et une fonction de coût quadratique

$$E = \frac{1}{2} \sum_j E_j^2 \tag{6.22}$$

où

$$E_j = \begin{cases} d_j - q_j & \text{si } j \text{ est un neurone de sortie} \\ 0 & \text{sinon} \end{cases} \tag{6.23}$$

La descente du gradient s'écrit

$$\Delta w_{uv} = -\eta \frac{\partial E}{\partial w_{uv}} = \eta \sum_i E_i \frac{\partial q_i}{\partial w_{uv}} \tag{6.24}$$

L'évaluation de $\partial q_i / \partial w_{uv}$ s'obtient par différenciation de l'équation (6.10),

$$\frac{\partial q_i}{\partial w_{uv}} = \frac{1}{D_i} \sum_j \left[w_{ji}^+ - w_{ji}^- q_i \right] \frac{\partial q_j}{\partial w_{uv}}$$

$$
\begin{aligned}
&+\delta_{iv}\frac{q_u}{D_v}\left[\mathbf{1}_{w^+} - q_v\mathbf{1}_{w^-}\right] - \delta_{iu}\frac{q_u}{D_u} \\
&+\frac{1}{D_i}\left[\mathbf{1}_{\lambda^+_{i,inert}}\frac{\partial\lambda^+_{i,inert}}{\partial w_{uv}} - \mathbf{1}_{\lambda^-_{i,inert}}q_i\frac{\partial\lambda^-_{i,inert}}{\partial w_{uv}}\right] \qquad (6.25)
\end{aligned}
$$

où

$$
\begin{aligned}
\frac{\partial\lambda^+_{i,inert}}{\partial w_{uv}} &= \delta_{vi}d_u\left[\mathbf{1}_{w^-}d_v - \mathbf{1}_{w^+}\right] + \delta_{ui}d_u \\
\frac{\partial\lambda^-_{i,inert}}{\partial w_{uv}} &= \delta_{vi}d_u\left[\mathbf{1}_{w^+}\frac{1}{d_v} - \mathbf{1}_{w^-}\right] - \delta_{ui} \qquad (6.26)
\end{aligned}
$$

La notation $\mathbf{1}_{w^+}$ est une forme compacte de $\mathbf{1}_{\left[w_{uv}=w^+_{uv}\right]}$. δ_{iu} est le symbole de Kronecker, lequel peut aussi s'écrire $1_{[i=u]}$. Cette notation permet de ne pas surcharger inutilement les formules.

Une fois les termes regroupés, on obtient

$$
\begin{aligned}
\sum_j \mathbf{L}_{ij}\frac{\partial q_j}{\partial w_{uv}} &= \delta_{iv}\frac{q_u}{D_v}\left[\mathbf{1}_{w^+} - q_v\mathbf{1}_{w^-}\right] - \delta_{iu}\frac{q_u}{D_u} \\
&\quad +\frac{1}{D_i}\left[\mathbf{1}_{\lambda^+_{i,inert}}\frac{\partial\lambda^+_{i,inert}}{\partial w_{uv}} - \mathbf{1}_{\lambda^-_{i,inert}}q_i\frac{\partial\lambda^-_{i,inert}}{\partial w_{uv}}\right] \qquad (6.27)
\end{aligned}
$$

où

$$
\mathbf{L}_{ij} = \delta_{ij} - \frac{1}{D_i}\left(w^+_{ji} - w^-_{ji}q_i\right) \qquad (6.28)
$$

Substituons (6.26) dans (6.27). L'inversion des équations linéaires donne

$$
\begin{aligned}
\frac{\partial q_j}{\partial w_{uv}} &= (\mathbf{L}^{-1})_{jv}\frac{q_u}{D_v}q_u\left[\mathbf{1}_{w^+} - q_v\mathbf{1}_{w^-}\right] - (\mathbf{L}^{-1})_{ju}\frac{q_u}{D_u} \\
&\quad +(\mathbf{L}^{-1})_{jv}\frac{d_u}{D_v}\mathbf{1}_{\lambda^+_{v,inert}}\left[\mathbf{1}_{w^-}d_v - \mathbf{1}_{w^+}\right] \\
&\quad -(\mathbf{L}^{-1})_{jv}\frac{d_u}{D_v}\mathbf{1}_{\lambda^-_{v,inert}}q_v\left[\mathbf{1}_{w^+}\frac{1}{d_v} - \mathbf{1}_{w^-}\right] \\
&\quad +\frac{(\mathbf{L}^{-1})_{ju}}{D_u}\left(\mathbf{1}_{\lambda^+_{u,inert}}d_u - \mathbf{1}_{\lambda^-_{u,inert}}q_u\right) \qquad (6.29)
\end{aligned}
$$

Soit

$$
\delta_p = \frac{Y_p}{D_p} \quad \text{et} \quad Y_p = \sum_j E_j(\mathbf{L}^{-1})_{jp} \qquad (6.30)
$$

D'après (6.29), (6.30) et (6.24), la règle delta apparaît sous la forme

$$
\begin{aligned}
\Delta w_{uv} \;=\;& \eta\left[\delta_v\left(\mathbf{1}_{w^+}-q_v\mathbf{1}_{w^-}\right)-\delta_u\right]q_u \\
&+\eta\delta_v\frac{d_u}{D_v}\mathbf{1}_{\lambda^+_{v,inert}}\left[\mathbf{1}_{w^-}d_v-\mathbf{1}_{w^+}\right] \\
&-\eta\delta_u\frac{q_v d_u}{D_v}\mathbf{1}_{\lambda^-_{v,inert}}\left[\mathbf{1}_{w^+}\frac{1}{d_v}-\mathbf{1}_{w^-}\right] \\
&+\eta\delta_u\left[\mathbf{1}_{\lambda^+_{u,inert}}d_u-\mathbf{1}_{\lambda^-_{u,inert}}q_u\right]
\end{aligned}
\tag{6.31}
$$

Remarquons l'apparition de δ_u qui n'apparaît pas dans l'expression de la règle delta familière. La présence de ce terme s'explique intuitivement par le fait que le potentiel dépende du poids des connexion sortantes par le truchement de r_i. Eq. 6.31 requiert l'inversion de la matrice $\mathbf{L}$ de taille $N \times N$ à chaque itération. Toutefois, l'astuce que l'on doit à Pineda [Pineda 87] et Almeida [Almeida 87] s'applique aussi à notre cas.

Ecrivons

$$\sum_j \mathbf{L}_{ji} Y_j = E_i \tag{6.32}$$

grâce à (6.28), on obtient

$$Y_i = \sum_j w^\star_{ji} Y_j + E_i \tag{6.33}$$

avec

$$w^\star_{ji} = \frac{\left[w^+_{ij} - q_j w^-_{ij}\right]}{D_j} \tag{6.34}$$

Eq. 6.33 est très similaire à Eq. 6.10 et peut être résolue de la même manière, en relaxant un *réseau à propagation d'erreur* ou réseau *adjoint* régi par un système d'équations dynamiques analogues à Eq. 6.9

$$\tau\frac{dY_i}{dt} = -Y_i + \sum_j w^\star_{ji} Y_j + E_i \tag{6.35}$$

Cette procédure ne porte pas préjudice à la localité de l'algorithme et convient à une implémentation parallèle. Si on considère un commutateur de taille $N \times N$, il faut $O(N^2)$ connexions pour modéliser ce dernier. Le temps nécessaire à l'intégration de (6.35) est en $O(N^4)$. Cela se traduit par des économies draconiennes pour de grandes valeurs de N s'il fallait inverser la matrice $\mathbf{L}$. On obtiendrait une complexité en $O(N^6)$.

6.5 La gestion des ressources

Cet paragraphe aborde la question de la gestion des ressources dans un réseau à intégration de services large-bande (B-ISDN). Par gestion des ressources, nous entendons les mesures prises par le contrôleur d'admission des appels pour accepter ou rejeter les appels selon la qualité de service requise par l'utilisateur et celle que le réseau peut offrir. Si le contrôleur fonctionnait de manière optimale, la congestion serait évitée dans la mesure où les trafics responsables de l'excédent de charge seraient simplement rejetés. Lorsqu'un appel est accepté, un *contrat* tacite est établi entre l'administrateur réseau et le souscripteur spécifiant les caractéristiques du flot de données et la qualité de service requise par la connexion.

Le rôle du réseau à intégration de services est de convoyer des trafics sporadiques décrits par des variables statistiques. Aussi, le contrôleur doit être suffisamment flexible pour appréhender des trafics aux fluctuations aléatoires. Afin de délivrer la qualité de service promise dans la phase d'établissement des connexions, le contrôleur doit superviser l'état du réseau en chaque nœud et estimer le surplus de charge occasionné par la transmission de trafics supplémentaires. Puisque le temps moyen de transit à travers une file d'attente se déduit du nombre moyen de client à l'état stationnaire d'après le théorème de Little, il suffit de prédire le nombre moyen de cellules en attente dans les files lorsqu'un nouveau trafic est convoyé. Pour chaque nouvelle requête émise au contrôleur, celui-ci sélectionne un ensemble de routes possibles, et pour chacune d'entre elles, estime le temps de transfert des cellules de bout-en-bout. La qualité de service sur la route au délai minimal est comparée à la qualité requise par la connexion.

L'arbitrage fait au *niveau appel* [Hui 87] doit garantir une qualité de service au niveaux *burst* et *cellule* s'il est fait correctement.

6.5.1 Le contrôle d'admission des appels

L'arbitrage des appels est effectué localement à la fin de chaque cycle d'apprentissage et les requêtes sont observées suivant leur ordre d'arrivée. Nous nous restreignons à un modèle simple dans lequel le délai de transmission est le seul critère de qualité de service retenu. Le contrôleur se fonde sur la nature de la source et la qualité de service requise pour décider ou non d'accepter ce trafic. Les paramètres retenus sont

- le couple source-destination, (s, d).
- le débit moyen de la source, λ_i.
- une estimation du degré de fluctuation du trafic entrant représenté par le coefficient de variation de la distribution des inter-arrivées, C_{a_i}.
- un facteur de qualité de service, $SQR \in [0, 1]$.

Ce facteur de qualité de service *SQR* est perçu comme une mesure *relative* parmi l'ensemble des requêtes *au même instant*. Il faut seulement garder à l'esprit qu'une valeur faible de *SQR* est associée à une bonne qualité de service et vice-versa.

Le contrôle d'admission au commutateur S est effectué au cycle k sur la base du vecteur

$$\begin{aligned} C_i &= (S, D_i, \lambda_i, C_{a_i}, f_i) \quad ,\forall i \in [0, \ldots, n_1 + n_2] \\ \mathbf{C}(k) &= [C_1, \ldots, C_{n_1}, C_{n_1+1}, \ldots, C_{n_1+n_2}] \quad \text{au cycle } k \end{aligned}$$

où n_1 est le nombre de connexions actives, et n_2 est le nombre de requêtes au cycle k. Les requêtes sont mises en mémoires au fur et à mesure de leurs arrivées et traitées selon cet ordre. L'utilisateur est tenu de respecter les clauses du contrat après acceptation. Si l'appel est rejeté, le contrat est renégocié immédiatement sur la base d'autres paramètres.

Dès lors, le réseau de neurones doit apprendre la relation fonctionnelle qui existe entre ce vecteur de paramètres décrivant la source et l'état du réseau. Cette fonction doit être modifiée de manière adéquate lorsque fluctue le volume du trafic.

6.5.2 Les délais de communication

Notre point de départ est un réseau de N nœuds de commutation. Le nombre de tampons dans le système est M. Soit T_i, le temps moyen passé dans la file i; de même, $T_{s,d}$ désigne le délai moyen de transmission du trafic originaire du nœud s et à destination du nœud d et notons π_{sd} le chemin suivi. On dira que la file i est sur le chemin π_{sd} selon la notation $i \in \pi_{sd}$.

Le taux moyen des arrivées λ_i en entrée de cette cellule est égal à la somme des taux moyens des trafics transitant par cette file,

$$\lambda_i = \sum_{(s,d)/i \in \pi_{sd}} \lambda_{sd} \tag{6.36}$$

En utilisant le théorème de Little on obtient immédiatement,

$$T_{s,d} = \sum_{i \in \pi_{sd}} T_i = \sum_{i \in \pi_{sd}} \frac{\overline{N_i}}{\lambda_i} = \sum_{i \in \pi_{sd}} \frac{1}{\mu_i} \frac{\mu_i}{\lambda_i} \overline{N}_i = \sum_{i \in \pi_{sd}} \frac{b}{\mu_i} \frac{d_i^{ql}}{d_i^s} \tag{6.37}$$

où d_i^{ql} (resp. d_i^s) représente la sortie désirée du neurone 1 (resp. 2). D'après (6.37), nous déduisons une l'estimation suivante de T_{sd}

$$\hat{T}_{s,d} = \sum_{i \in \pi_{sd}} \frac{b}{\mu_i} \frac{q_i^{ql}}{q_i^s} \tag{6.38}$$

Eq. 6.38 dépend toujours de μ_i, le temps moyen de service de la file i. Cependant, on remarquera que le temps moyen de service de la file de sortie est la capacité de la ligne de sortie. Pour les files en entrée, μ_i doit être mesuré.

Cette discussion peut être généralisée au délai moyen de l'ensemble des messages [Kleinrock 76],

$$\mathbf{T} = E\,[\text{délai du message}] \tag{6.39}$$

$\mathbf{T}$ peut être relié à $T_{s,d}$ par

$$\mathbf{T} = \sum_{s=1}^{N}\sum_{d=1}^{N}\frac{\lambda_{sd}}{\lambda}T_{s,d} \tag{6.40}$$

où λ est le trafic total convoyé par le réseau défini par

$$\lambda = \sum_{s=1}^{N}\sum_{d=1}^{N}\lambda_{sd} \tag{6.41}$$

(6.40) est la décomposition sur la base des couples origine-destination. La fraction λ_{sd}/λ du trafic total subit un délai $T_{s,d}$ en moyenne. D'après (6.40), nous avons

$$\begin{aligned}
\mathbf{T} &= \sum_{s=1}^{N}\sum_{d=1}^{N}\frac{\lambda_{sd}}{\lambda}\sum_{i/i\in\pi_{sd}}T_i \\
&= \sum_{i=1}^{M}\frac{T_i}{\lambda}\sum_{(s,d)/i\in\pi_{sd}}\lambda_{sd} \\
&= \sum_{i=1}^{M}\frac{\lambda_i}{\lambda}T_i
\end{aligned} \tag{6.42}$$

Finalement, en introduisant les potentiels des neurones, il vient

$$\mathbf{T} = \sum_{i=1}^{M}\frac{\lambda_i}{\lambda}\frac{bd_i^{ql}}{\lambda_i} = \frac{b}{\lambda}\sum_{i=1}^{M}d_i^{ql} \tag{6.43}$$

Le délai moyen de l'ensemble des messages est proportionnel à la somme des activations des neurones 1 du RNA.

Enfin, le taux d'occupation du réseau peut être évalué de façon fort simple en sommant les potentiels de tous les neurones 3 en sortie. Il est intéressant de noter que le contrôle neuronale se résume à une sommation de neurones.

6.5.3 La sélection de la route optimale

Lorsqu'un chemin doit être établi entre (s,d), un ensemble de chemins possibles sont proposés par une algorithme *ad hoc*, tenant compte de la connectivité du réseau et des capacités des lignes. La procédure que nous présentons à pour objet de sélectionner la route qui minimise le délai de bout-en-bout $T_{s,d}$.

Nous avons déjà mentionné que les connexions issues des neurones 3 (Fig. 6.3) associés au taux d'occupation des files en entrée ont des poids *proportionnels* aux probabilités de routage correspondants. Or cette probabilité varie d'un façon discontinue lorsque un nouveau trafic transite par cette voie. Par conséquent, il faut tenir compte de ce phénomène et ajuster les poids "manuellement" dans chaque commutateur. Puisque la somme des poids est théoriquement une constante, nous pouvons écrire

$$\left\{\begin{array}{lcl} w_{ik}q^s_{i,old} & = & (w_{ik}+\Delta w_{ik})q^s_{i,new} \quad \text{si} \quad k\neq j \\ \Delta w_{ij} & = & -\sum\limits_{k\neq j}\Delta w_{ik} \end{array}\right. \tag{6.44}$$

où $q^s_{i,old}$ et $q^s_{i,new}$ sont les potentiels du neurone 3 et les indices *old* et *new* indiquent s'il s'agit de l'ancien potentiel ou du nouveau encore inconnu. Soit $\lambda_i^{new}=\lambda_i^{old}+\lambda_i^{sup}$ le nouveaux taux moyen des arrivées en entre de la file. Un calcul immédiat conduit à

$$\frac{d^s_{i,new}}{d^s_{i,old}}=\frac{\rho_i^{new}}{\rho_i^{old}}=\frac{\lambda_i^{new}}{\lambda_i^{old}}=1+\frac{\lambda_i^{sup}}{\lambda_i^{old}}=1+\frac{1}{\mu_{in}}\frac{\lambda_i^{sup}}{\rho_i^{old}}=1+\frac{1}{\mu_{in}}\frac{\lambda_i^{sup}}{d^s_{i,old}} \tag{6.45}$$

Finalement, d'après (6.45) on peut réécrire (6.44) en substituant $q^s_{i,old}$ par sa valeur désirée $d^s_{i,old}$. On obtient

$$\left\{\begin{array}{lclcl} \Delta w_{ik} & = & -(1-\frac{\lambda_i^{old}}{\lambda_i^{new}})w_{ik} & = & \frac{-\lambda_i^{sup}w_{ik}}{\lambda_i^{sup}+\mu_{in}d^i_{s,old}} \quad \text{if} \quad k\neq j \\ \Delta w_{ij} & = & -\sum\limits_{k\neq j}\Delta w_{ik} & = & \frac{\lambda_i^{sup}}{\lambda_i^{sup}+\mu_{in}d^i_{s,old}}\sum\limits_{k\neq j}w_{ik} \end{array}\right. \tag{6.46}$$

Lorsque une nouvelle requête est considérée à un commutateur donné, plusieurs routes sont envisagées sur lesquels pourraient être propagé le trafic supplémentaire λ^{sup}. Les poids sont successivement ajustés selon (6.46) jusqu'au nœud de destination et le réseau de neurones est relaxé. Les potentiels des neurones 1 et 3 le long de la route sont transmis à la source s et le délai moyen est estimé. L'estimation la plus faible détermine la route optimale.

6.5.4 Le contrôle de la qualité de service

Un fois la route sélectionnée, le contrôle de qualité est initié. La requête est rejetée si le délai excède $\phi(SQR)$ où ϕ désigne une *fonction d'admission*. La fonction d'admission dépend de la manière dont SQR est défini; il peut s'agir d'un taux de perte maximum dans la mesure où ce dernier est relié par voix indirectes au délais de transmission.

Il est naturel de choisir *a priori* une fonction statique de SQR comme critère d'admission. Toutefois, une variation soudaine de la charge du réseau conduirait par une sous-estimation des délais par le réseau de neurones en raison de son inertie. C'est la cas, lorsque le réseau accepte à tort un trafic trop gourmand provoquant des surcharges locales. Le contrôleur, dans ce cas, accepte des connexions en se fondant sur une estimation trop optimiste des délais, sans prendre la mesure du problème. Ce phénomène conduit immanquablement à une saturation générale.

C'est pourquoi l'occupation réelle des files est observée sur la route considérée. Nous nous limitons, par soucis de simplicité, à une fonction linéaire dont le coefficient, appelé *facteur de charge* $lf(k)$, compense les fluctuation rapides du réseau à l'instant k. Ainsi, la requête est rejetée lorsque,

$$\hat{T}_{s,d}>lf(k)*SQR \tag{6.47}$$

lf est adaptée avant que ne soit initiée la procédure de contrôle, selon

$$lf(k+1) = lf(k) - \beta \sum_i \mathbf{1}_{\{occ_i > 0.8\}} (occ_i - 0.8)^2 \tag{6.48}$$

où occ_i est le taux d'occupation des files et β, une constante appropriée. Cela force le contrôleur a être plus sélectif lorsque le réseau est soumis à des trafics important. A chaque requêtes rejetée, le facteur de charge est adapté selon $lf(k) \leftarrow lf(k) + 0.01$ dans notre modèle. Il faut garder à l'esprit que la contrôle de qualité opère une sélection *relative* parmi les autres requêtes en attente. Il confère une priorité aux utilisateurs moins exigeants en termes de délais. Cette politique d'admission élémentaire ne prétend pas répondre au vaste problème du contrôle de qualité mais préfigure un stratégie de CAC global. Une sélection plus scrupuleuse fondée sur une contrainte temporelle absolue de bout-en-bout est à l'étude.

Finalement, chaque commutateur est le siège d'un processus de contrôle observant à chaque cycle le vecteur $[C_1, \ldots, C_{n_1}, C_{n_1+1}, \ldots, C_{n_1+n_2}]$. Pour les n_2 requêtes, une route optimale est choisie et le contrôle de qualité est effectué selon (6.48). Au terme de ce contrôle, la connexion est établie ou une notification de rejet est formulée à l'utilisateur. Le cas échéant, la valeur des poids est restaurée à leur valeur d'origine, lf est incrémenté, et une nouvelle requête est considérée.

6.6 Résultats expérimentaux

Dans ce paragraphe, une application de la méthode neuronal de contrôle d'admission de trafic en rafale est proposée. Le trafic généré est un flot corrélé, modulé par une chaîne de Markov, sans être un processus de renouvellement. Les paramètres des sources sont choisies aléatoirement ainsi que le facteur de qualité requis par les utilisateurs. Le contrôleur doit effectivement anticiper le surplus de charge dans les tampons au vu de ces paramètres et estimer les délai de transmission. De part la nature stochastique des flots générés, les paramètres déclarés ne représentent pas toujours précisément leurs caractéristiques réelles ce qui complique grandement la tâche du réseau de neurones. Notre espoir repose sur la flexibilité du modèle et sa tolérance aux erreurs dans les descripteurs de trafic.

6.6.1 Le modèle de source

Le réseau est supposé fonctionner en mode connexion et assimile à chaque appel un circuit logique pour une période durant laquelle la connexion est active. La durée d'établissement de la connexion est générée selon une loi exponentielle. Il n'existe pas de modèle universel représentant de façon adéquate tous les services proposés à ce jour. Les services multimédia sont représentés très schématiquement par un débit variant aléatoirement entre plusieurs débit fixes à mesure que sont activés ou désactivés différents média. Le trafic est donc typiquement généré en rafale : des périodes de grande activité alternent avec des périodes de repos.

Le choix d'un modèle de source est guidé par l'exigence de la simplicité analytique et de la versatilité du modèle. Un processus de Poisson modulé par une chaîne de Markov à

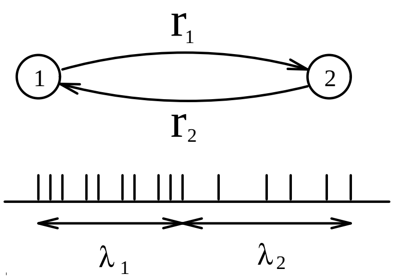

FIG. 6.7 - : Le processus MMPP à 2 états génère un trafic Poissonnien en alternance de paramètres $\lambda_1 > \lambda_2$.

deux états (2-MMPP) [Nikolaidis 92] (Fig. 6.7) répond à ces objectifs. Dans l'état i, les arrivées se font selon un processus de Poisson de taux λ_i. Les changements d'état suivent une distribution exponentielle de moyenne r_i^{-1}. Ainsi, un ensemble de quatre paramètres identifie le modèle de manière univoque le processus stochastique. Ce modèle simple, corrélé, qui de surcroît n'est pas un processus de renouvellement, est employé couramment pour approximer le flot résultant de la superposition d'une collection de sources *ON/OFF* avec un flot constant de données [Heffes86]. A partir des moments des sources individuelles, les caractéristiques du flot agrégé sont déduites en utilisant la propriété d'indépendance mutuelle. Dans un scénario réel, les descripteurs

- le temps moyen d'inter-arrivée (ou le taux moyen d'arrivée).
- le ratio variance à long terme sur moyenne du nombre d'arrivées (C_a^2).
- le ratio variance à long terme sur moyenne du nombre d'arrivées sur un intervalle $[0, t_1]$ (en général $[0, 500ms]$).
- le moment d'ordre trois du nombre d'arrivées dans l'intervalle $[0, t_2]$, (en général $[0, 500ms]$).

sont déterminés de telle sorte qu'ils correspondent à ceux du flot réel. Le trafic généré par plusieurs sources MMPP est multiplexé et injecté dans un tampon d'entrée. Ces tampons reçoivent donc du trafic d'un commutateur distant ainsi que des sources extérieures. Les descripteurs, $\overline{\lambda}$ et $C_{T_a}^2$, du trafic généré par un pocessus MMPP s'expriment en fonction de ses paramètres selon les formules

$$\begin{aligned} \lambda_s &= \frac{\lambda_1 r_2 + \lambda_2 r_1}{r_1 + r_2} \\ C_{a_s}^2 &= 1 + \frac{2(\lambda_1 - \lambda_2)^2 r_1 r_2}{(r_1 + r_2)^2(\lambda_1 r_2 + \lambda_2 r_1)} \end{aligned} \tag{6.49}$$

Les descripteurs des flots MMPP superposés sont obtenus grâce à Eq. 6.5. Ils sont normalisés, puis injectés dans les neurones correspondants (Fig. 6.6).

item	*valeurs*
capacité des lignes	$150 Mb/s$
taille des tampons	sortie : 5 entrée : ∞
capacité de commutation	3 cell/μs
nombre de nœuds	4
nombre de lignes	12
période de contrôle	$1s$

TAB. 6.1 - : Caractéristiques du réseau ATM simulé.

6.6.2 Résultats expérimentaux

Afin de valider notre modèle, une simulation à l'aide d'un simulateur de réseau ATM à événements discrets est réalisée au cours de laquelle le contrôle d'admission et le routage sont effectués au vu des prédictions neuronales. Les contrôleurs implémentés sur chaque commutateur disposent des caractéristiques individuelles des trafics 2-MMPP pour prédire la charge du réseau. Dans un scénario pratique, le flot de données en provenance de l'extérieur est constitué d'une superposition de nombreux trafics utilisateurs (e.g. réseaux locaux, terminaux multiplexés, etc.). Les propriétés statistiques de ce trafic évoluent à mesure que s'activent et se désactivent les connexions. Celles-ci sont réinitialisées aléatoirement lorsque les temps d'appel respectifs sont arrivés à échéance. A l'échelle de la seconde, l'établissement d'une nouvelle connexion a un effet qu'on peut considérer comme marginal sur le comportement du flot de données global. La continuité statistique du trafic n'est donc pas rompue. Dans l'espoir de modéliser correctement cette situation, plusieurs sources MMPP sont combinées pour former un trafic multiplexé. Le flot résultant est transmis aux tampons localisés en entrée des commutateurs. Ainsi, le trafic en entrée de chaque file d'entrée peut être représenté comme un processus MMPP à 2^n états si n désigne le nombre de sources 2-MMPP multiplexées.

La configuration du réseau ATM - Cette simulation est effectuée avec un modeste réseau composé de 4 nœuds de commutation de taille 3×3 totalement interconnectés. Tous les trafic convoyés transitent obligatoirement par 2 nœuds intermédiaires de façon à accroître les risques de blocage dans les files d'attente en entrée (cf. Fig. 6.8). Cette contrainte sur le routage ne nuit pas à la signification des résultats; bien au contraire, elle est à l'origine de la congestion dans les tampons d'entrée que nous cherchons à reproduire. Les tampons sont de taille infinie en entrée et finie en sortie (5 cellules). La capacité des liens est de $150 Mb/s$. la capacité de commutation est fixée à 3 cellules/μs, c'est-à-dire que le processus de commutation est activé 3 fois par microseconde. Table 6.1 résume les paramètres de simulation.

Chaque tampon en entrée reçoit un trafic en provenance d'un commutateur distant et de 5 processus 2-MMPP externes au réseau. La durée d'établissement d'un appel suit

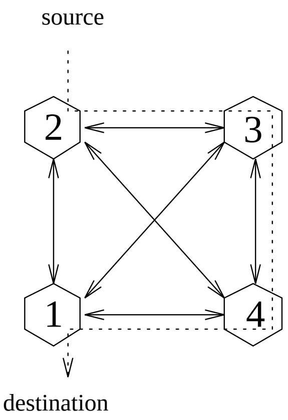

FIG. 6.8 - : Topologie du réseau ATM à quatre nœuds de commutation. Un exemple de chemin est représenté en pointillé; le détour a pour objet d'accroître les risques de blocage des cellules dans les files d'entrée.

item	*valeurs numériques*
$1/r_i, i = 1, 2$	10-50ms
$\lambda_i, i = 1, 2$	5-15Mb/s
nombre d'appels par file d'entre nombre d'appels actifs	5 $5 \times 12 = 60$
durée moyenne des appels	10s

TAB. 6.2 - : Paramètres des sources 2-MMPP injectées dans les tampons situés en entrée des commutateurs.

une loi exponentielle de durée moyenne $10s$ au terme de laquelle une nouvelle requête est formulée au contrôleur[4] (Table 6.2). Les taux Poissonniens λ_i sont choisis aléatoirement dans l'intervalle $[5, 15]Mb/s$ et les taux de transition dans l'intervalle de durée moyenne $[10, 50]ms$. Le facteur de qualité de service, SQR, est choisi aléatoirement dans $[0.3, 0.7]$ pour tenir compte des contraintes sur les délais de transmission exprimées par les utilisateurs. Avec 4 nœuds, il n'y a que 2 routes possibles (voir Fig. 6.8). Le routage se limite donc à sélectionner selon (6.38) l'un des deux chemins dont l'estimation du délai de transmission est la plus faible. La requête est rejetée si le critère (6.47) n'est pas vérifié. Dans ce cas, la requête est reconsidérée avec un nouveau vecteur de paramètres : la destination, les descripteurs de trafic et le facteur de qualité de service sont réinitialisés aléatoirement. Le processus est répété jusqu'à l'admission de la connexion. Ainsi, le nombre de trafics véhiculés par le réseau demeure identique au cours de la période de simulation.

Le contrôle d'admission est délocalisé mais synchrone; c'est-à-dire que chaque commutateur effectue son propre contrôle au même instant (à la fin du cycle d'apprentissage). Lorsque le temps d'une communication expire, une requête est générée dans le commutateur où est issue la connexion. La période d'observation est fixée à $1s$ de façon à traiter en moyenne plusieurs requêtes simultanément (entre 1 et 2 demandes d'admission par commutateur en moyenne à chaque cycle).

L'apprentissage - Les sous-réseaux de neurones implémentés sur les commutateurs ont chacun 27 neurones et 54 connexions. Soit un total de 108 neurones et 216 connexions. Les valeurs initiales des poids sont déterminés selon les résultats obtenus au § 6.4.1. Les poids des connexions entre les files d'attente sont fixés selon des heuristiques visant à obtenir $q_i = 0.5$ pour un trafic uniforme d'intensité intermédiaire.

Le pas d'apprentissage est pris assez élevé pour adapter le RNA aux variations temporelles du réseau ATM; un pas de 0.05 est suffisant. L'apprentissage est effectué durant plusieurs centaines de secondes au cours desquelles le contrôle neuronal est désactivé; un contrôle rudimentaire est effectué en limitant le taux moyen des trafic d'entrée à la moitié du taux maximum possible. Tout risque de congestion est évité mais le réseau est sous-utilisé. Une fois la convergence observée, le réseau de neurones est activé tandis que les contraintes sur le trafic sont levées progressivement.

La vitesse de convergence confirme le bien fondé de l'analyse de files d'attente. De part la nature stochastique des données en entrée du réseau de neurones, il est clair que la surface d'erreur est caractérisée par des minimums fugaces que poursuit indéfiniment la procédure d'apprentissage.

[4] La durée moyenne d'un appel est intentionnellement faible pour réduire le temps de simulation sans pour autant altérer la signification qualitative des résultats.

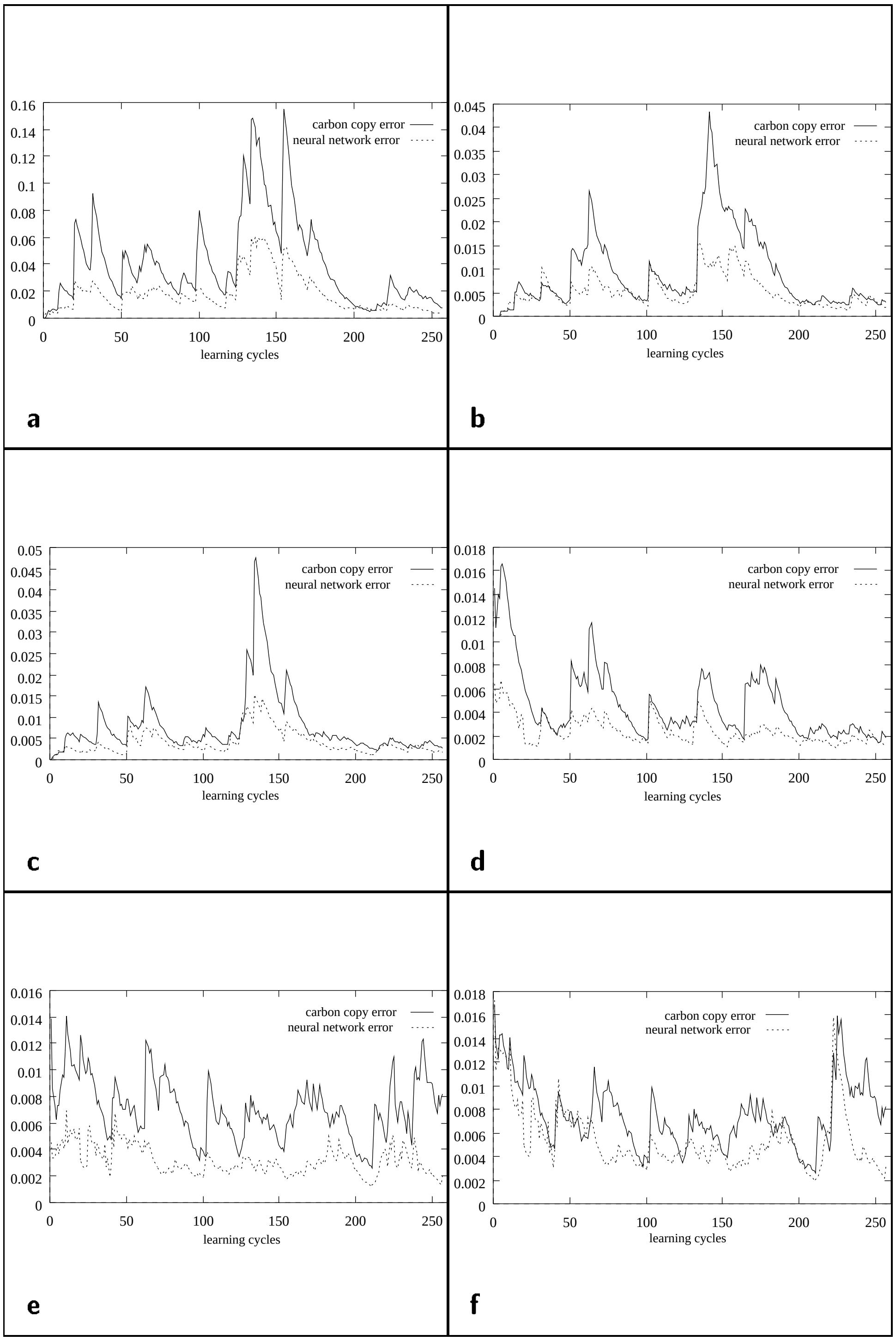

FIG. 6.9 - : Erreurs quadratiques du réseau de neurones modélisant une file d'entrée à gauche et une file de sortie à droite. En trait pointillé, le RNA, en trait plein, les prédictions "copie carbone" (i.e. $\hat{x}_t = x_{t-1}$). De haut en bas : nombre moyen de cellules en attente, taux d'utilisation de la file et carré du coefficient de variation de la distribution des inter-arrivées.

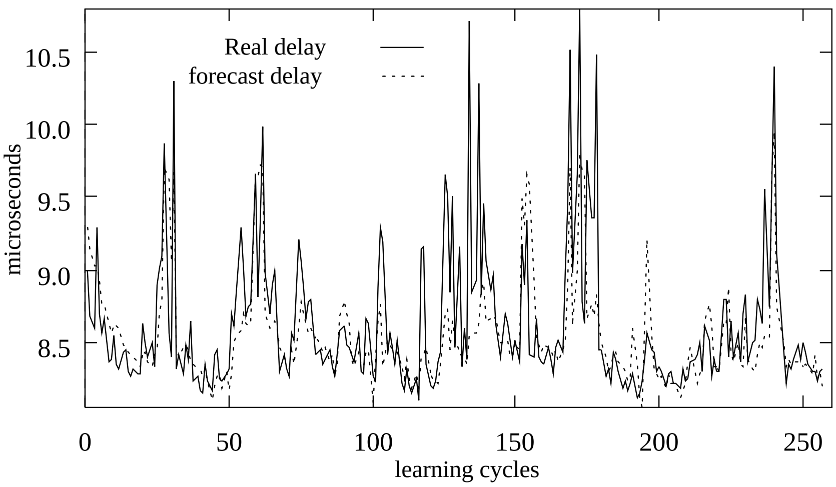

FIG. 6.10 - : Comparaison entre les prédictions du temps moyen de transmission des messages (ligne pointillée) et les valeurs obtenues par simulation (trait plein), au cours des cycles d'apprentissage.

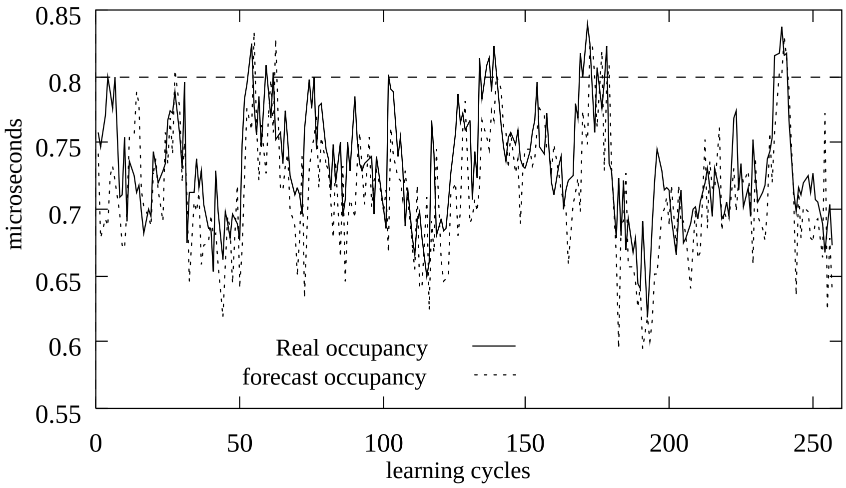

FIG. 6.11 - : Comparaison entre les prédictions du taux d'occupation du réseau (en ligne pointillée) et les valeurs observées par simulation (en trait plein), au cours des cycles d'apprentissage.

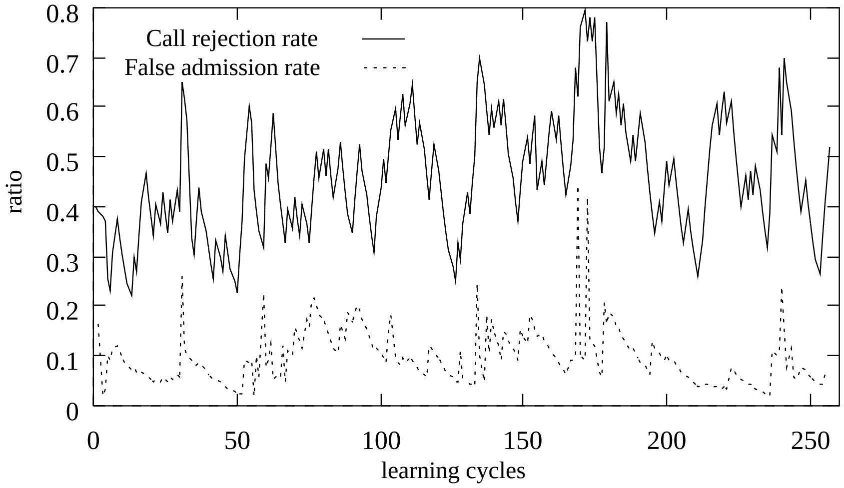

FIG. 6.12 - : Fraction des appels admis par rapport au nombre de requêtes à chaque cycle de contrôle (en trait plein). Fraction du nombre de connexions pour lesquelles la qualité de service requise n'a pas été fournie (en pointillé).

Fig. 6.9 illustre le comportement typique de l'erreur quadratique (non-normalisée) sur 250 cycles de contrôle après convergence du réseau de neurones. Deux files d'attente ont été arbitrairement sélectionnées sur un commutateur et chaque neurone est observé au cours de cette période. L'erreur "copie carbone" ($x_t = \hat{x}_t$) est représentée comme un index de performance.

On observe que la courbe d'erreur du réseau de neurones est maintenue en-deçà de celle de la copie carbone. L'apprentissage a donc permis d'anticiper les variations statistiques des files d'attente à l'échelle de la seconde. Les tampons en entrée se prêtent mieux à la prédiction pour la simple raison que le trafic à leur entrée est mieux appréhendé. Les tampons en sortie sont sujets aux variations de routages et aux blocages intempestifs des cellules en tête de files dûs à la congestion momentanée des tampons de sorties. L'effet des arrivées en rafale sur l'occupation des tampons en entrée est clairement mis en évidence par un pic abrupte commun aux neurones d'entrée.

Le contrôle - Les graphes du délai moyen de transmission et d'occupation du réseau sont représentés Fig. 6.10 et 6.11. Le temps moyen de transmission exhibe des variations sporadiques correctement anticipées par le contrôleur exceptés les pics abruptes manifestement imprévisibles. La raison est que les files en entrée sont infinies alors que le contrôleur les considère finie (de taille b); si tôt que sature un neurone représentant le nombre moyen de cellules dans une file d'entrée, le contrôleur estime le nombre moyen de cellules à 5 alors que ce nombre est souvent supérieur. Du reste, la charge élevée de cette file d'entrée contribue sans aucun doute à la hausse substantielle observée à la même époque (cf. Fig. 6.10).

Le taux d'occupation du réseau (cf. Fig. 6.11) est maintenu en-deçà de 0.8, l'occupation maximale que nous nous sommes fixés dans Eq. 6.48. L'adaptation du RNA à la dynamique du trafic permet la gestion efficace des ressources du réseau aux limites des possibilités offertes. Enfin, le taux de rejet des demandes d'appel est tracé Fig. 6.12; environ 50% des requêtes sont rejetées avec des pointes lorsque la charge du réseau est élevée. Mais, ce taux est entièrement dépendant des intervalles dans lesquels sont choisis les descripteurs de trafic et de la façon dont est adapté le facteur $lf(k)$ à chaque instant.

La fraction des connexions qui n'ont pas eu la qualité de service escomptée est aussi représentée Fig. 6.12. Le calcul de cette valeur est effectué de la façon suivante : à chaque cycle k, les connexion admisent à l'itération précédente $k-1$ sont scrutées. L'ancienne valeur, $lf(k-1)$, est utilisée dans Eq. 6.47 avec la vraie valeur de T_{sd} mesurée. Si le critère d'admission n'est plus vérifié, alors la connexion a été admise a tort. La simulation révèle que 15% du trafic est admis à tort en moyenne. Il faut noter que ce taux est aussi dépendant des descripteurs de trafic. Compte tenu de la politique de sélection relative des requêtes, un contrôle de qualité à long terme n'est pas envisageable dans le cas présent. Le but de cette expérimentation n'est pas de résoudre tous les problèmes épineux du contrôle en une seule fois, mais plutôt d'examiner la sélectivité d'un contrôleur opérant à l'aide d'un RNA délocalisé sous des conditions de trafic extrêmes.

6.6.3 Discussion

La gestion des ressources dans un réseau ATM est un problème extrêmement dense comme en témoigne la quantité d'articles qui lui sont consacrés dans la littérature spécialisée. Le dilemme qui se pose à la communauté scientifique est de trouver un ensemble très restreint de descripteurs de flots de données multimédia qui se prête à la modélisation. Une représentation exacte des propriétés stochastiques du trafic est impossible.

Les objectifs majeurs de la gestion des ressources sont au nombre de trois : 1) minimiser les risques de congestion, 2) maximiser l'utilisation du réseau, et 3) fournir la qualité de service (délai, taux de perte, etc.) en accord avec la nature des données véhiculées (son, image, données). Dans notre étude, nous avons opté pour une approche très ambitieuse puisque nous avons cherché, dans un sens, à répondre à tous les défis posés. S'il est vrai que la gestion des ressources nécessite la modélisation globale du réseau ATM, cette approche neuronale présente en revanche un certain nombre de faiblesses et de problèmes ouverts qu'il nous semble important de souligner :

- Le contrôle de qualité a été réalisé ici d'une façon très sommaire. Les tentatives menées avec des contraintes temporelles fixes sur les délais d'acheminement sont restées infructueuses car les délais estimés sont souvent biaisés par rapports aux délais observés sous l'effet d'un début de congestion locale résorbée par la suite (c'est le cas lorsque la taille moyenne de la file d'entrée est supérieure à 5). Ensuite, il faudrait vérifier pour chaque requête si le trafic supplémentaire acheminé ne porte pas préjudice à la qualité de service des connexions actives. Le nombre de vérifications à opérer devient exorbitant.

- Un accord sur les mécanismes de CAC au sein des réseaux à intégration de services opérant en mode ATM n'est pas encore disponible. En optant pour une modélisation globale du réseau, il fût impossible de trouver une base de comparaison rigoureuse avec d'autres méthodes proposées dans la littérature. Les méthodes de contrôle d'admission reposant sur les réseaux connexionistes [Hiramatsu 90, Tran-Gia 91] formulent le CAC comme un problème de mémoire associative. Un ensemble de M types de sources distinctes aux propriétés stochastiques prédéfinies est sensé décrire l'éventail des services possibles. Un vecteur d'appel constitué du nombre d'appels actifs dans chaque classe est additionné au nombre de requêtes puis présenté au réseau sous la forme $[n_1(t) + \Delta n_1(t), \ldots, n_M(t) + \Delta n_M(t)]$. La réponse du réseau de neurones statique non-bouclé est binaire; elle indique si oui ou non le contrôleur accepte le trafic global. Cette formulation consiste donc à scinder l'espace multidimensionnel en deux régions distinctes, une région d'acceptation et l'autre de refus et ce, en vue de maintenir la probabilité de perte dans le multiplexeur en dessous d'une valeur prédéfinie. Bien que le succès de cette approche soit clairement établi, la tâche est limitée à la gestion d'un seul tampon de taille limitée, celui du multiplexeur en entrée du réseau. En aucun cas, le problème de la congestion à l'échelle du réseau n'est abordé, ni même envisagé. Un certain nombre d'heuristiques (e.g. équivalent bandwith method (EBL), weight variance method WV, etc. [Cost 91]) estiment analytiquement la bande passante "équivalente" dans un multiplexeur. Les restrictions concernant l'application de ces méthodes sont les mêmes que précédemment. De plus, seules les sources aux propriétés statistiques très similaires se prêtent bien à cette approximation.

- Le commutateur ATM à deux étages comportant des tampons en entrée et en sortie est commode pour la modélisation. On se dirige toutefois vers l'emploi d'un réseau d'interconnexion multi-étages non-bloquant (e.g. Banyan, Clos, Benes networks, etc.) doté d'une capacité de routage multi-chemins. Il est aussi question de positionner les files d'attente aux points de connexion interne. Cela revient à mettre en place une file d'attente pour chaque couple d'entrée/sortie [Beylot 93]. Or, il est irréaliste de vouloir modéliser, dans ce cas, toutes les files par un réseau de neurones. Il reste à examiner dans quelles mesures il est possible de faire un rapprochement entre ce commutateur et un modèle équivalent ne possédant que des files en entrée et en sortie.

- Nous avons supposé que les statistiques des tampons à l'échelle de la seconde étaient disponibles pour effectuer l'apprentissage. Cette hypothèse est idéaliste dans la mesure où le nombre d'opérations nécessaires au calcul exacte de la moyenne et de la variance du nombre moyen de cellules en attente est proportionnel au nombre de cellules qui ont transitées par cette file. Une telle situation est clairement indésirable et surtout techniquement impossible à réaliser compte tenu des débits envisagés. A défaut de valeurs exactes, il faut se contenter d'approximations fournies par un mécanisme *ad hoc*. Il reste à examiner la qualité de ces estimations et la robustesse du réseau de neurones soumis à des valeurs désirées erronées.

- Enfin, nous avons limité nos simulations à un réseau de 4 nœuds de commutation pour ramener le temps de simulation à des proportions raisonnables. Dans un scénario pratique, le nombre d'entrées/sorties N du commutateur peut dépasser la centaine. Le réseau de neurones implémenté sur le commutateur comporte un nombre de connexions de l'ordre de N^2; la complexité temporelle de la relaxation du réseau de neurones est donc $O(N^4)$. Toutefois, en vertu de l'implémentation parallèle, le temps nécessaire à la relaxation du réseau pour chaque requête examinée peut être réduit considérablement.

6.7 Conclusion

Dans ce chapitre, nous nous sommes intéressés à l'application des réseaux récurrents à un problème de contrôle adaptatif très spécifique : le contrôle d'admission des appels dans un réseau ATM.

Le contrôleur d'admission doit assurer un niveau de service négocié dans la phase d'établissement des appels et écarter tout risque de congestion. Aussi, la faculté d'adaptation est la qualité indispensable dont il a fallu doter le contrôleur pour appréhender efficacement l'impact des trafics, décrits par des grandeurs statistiques sommaires, sur le réseau en activité. C'est pourquoi il était légitime de se tourner vers des modèles capables d'apprendre continuellement par l'exemple tels que les réseaux connexionistes.

Dans le cadre d'un contrôle global, la structure du réseau de communication doit être connue du réseau de neurones ainsi que la charge des commutateurs. Nous avons examiné la modélisation de chaque file d'attente dans un nœud de commutation à l'aide de trois neurones artificiels. Les résultats notoires de la théorie de files d'attente suggèrent une interaction à forme rationnelle du premier ordre entre les neurones. Aussi, chaque commutateur a été représenté par un modèle RNA communiquant avec ses voisins par les lignes de communication et dont l'apprentissage est strictement autonome. Grâce aux descripteurs de trafic, le RNA est capable d'appréhender l'impact des nouveaux flots de données sur la charge du réseau et donc, d'estimer la qualité de service, en terme de délai de transmission.

Une simulation a été réalisée au cours de laquelle le neuro-contrôleur sélectionne pour chaque requête d'établissement d'appel, le chemin au délai de transmission minimal. Ce délai est ensuite comparé à la qualité de service exigée par l'utilisateur sous forme d'un facteur de qualité relatif aux utilisateurs. Le modèle affiche des performances encourageantes et se montre capable de maintenir un taux d'utilisation important du réseau tout en évitant soigneusement les risques de congestion.

Chapitre 7

CONCLUSION ET PERSPECTIVES

L'objet de cette dissertation a été de présenter les aspects théoriques et des applications pratiques des réseaux récurrents munis d'une mémoire interne dans le contexte de la prédiction, de la modélisation et du contrôle. Les fondations théoriques de ces réseaux ont été mises en perspective avec les techniques prédictives linéaires et non-linéaires traditionnelles. En outre, nous avons montré que la modélisation des systèmes dynamiques non-linéaires dont l'observation ne permet pas de reconstruire le vecteur d'état sous-jacent, requiert l'usage d'un réseau de neurones récurrent muni d'une dynamique propre.

Le modèle de réseau de neurones récurrent et dynamique (DRNN) développé au Chapitre 3 s'inscrit dans le sillage des extensions apportées à ce jour au réseau statique à couches. Le DRNN est au réseau statique récurrent ce que le réseau FIR [Wan 93a] est au réseau statique à couches, c'est-à-dire une extension *temporelle*. La modélisation des synapses par des filtres linéaires auto-régressifs est particulièrement bien adaptée pour préserver une mémoire de l'état du réseau. Le DRNN se prête donc bien au traitement temporel, dans la mesure où la mémoire interne dont dispose le modèle lui permet de trouver une représentation interne du vecteur d'état du système étudié. De surcroît, la présentation d'une fenêtre temporelle en entrée du réseau n'est plus nécessaire.

L'accent a été mis sur la recherche d'un algorithme d'apprentissage de descente du gradient à faible coût. L'analyse de la convergence du gradient rétro-propagé à travers un réseau adjoint (le transposé du réseau original), met en évidence la décroissance *exponentielle* du gradient d'erreur vers l'origine, sous certaines conditions relatives à la matrice de poids. Nous arguons qu'un nombre limité de passes à travers l'adjoint suffit à approximer le gradient d'erreur exact avec une précision arbitraire. Au reste, cette étude confirme l'inadéquation de la descente du gradient aux problèmes affichant de longue dépendances temporelles.

Notre choix s'est porté sur une procédure à propagation récurrente arrière appelée rétro-propagation temporelle récurrente (TRBP) tirant profit de l'évanouissement du gradient dans l'adjoint. A la croisée de la rétro-propagation récurrente [Almeida 87, Pineda 87, Rohwer 87] et de la rétro-propagation dans le temps [Minsky 69, Rumelhart 86], l'algorithme TRBP bénéficie des avantages de l'un et de l'autre sans en présenter les inconvénients. L'observation du gradient propagé dans l'adjoint en cours d'apprentissage permet de stopper la duplication itérative du réseau dans le temps au moment opportun. Ainsi, le nombre d'opérations requises se limite au stricte minimum et les opération non locales, ou superflues, sont bannies. La complexité temporelle de la procédure est en $O(D^2N^2)$ où N est le nombre de connexions et D est le retard maximal dans la propagation des signaux synaptiques. En contraste, $O(D^2N^4)$ opérations sont nécessaires avec une stratégie du type RTRL [Williams 89].

Les qualités prédictives du modèle DRNN sont illustrées aux Chapitres 4 et 5. Une analyse exhaustive de la suite des *sunspots* a été menée dans la double perspective de souligner les propriétés nouvelles du DRNN et de valider les résultats théoriques obtenus au Chapitre 3. Un modèle composé de seulement deux neurones cachés et alimenté par la dernière valeur de la suite, est capable de rivaliser avec un modèle non-bouclé nécessitant une fenêtre temporelle de taille 12 à son entrée. Au vu de ce résultat surprenant, la suite des sunspots pourrait être générée par deux équations différentielles du second ordre.

Nous avons ensuite abordé le problème de la modélisation des processus chaotiques déterministes qui par définition ne sont pas prédictibles au-delà de quelques itérations. Le DRNN est capable, dans tous les cas étudiés, de reproduire fidèlement le comportement du processus chaotique ne disposant que d'un nombre restreint de variables d'état. Une fois collectées les réponses successives du réseau opérant en bouclage fermé, l'attracteur fractal du processus chaotique est reconstruit. Ce dernier est quasiment identique à l'original. Au reste, le DRNN dispose d'un nombre plus faible de poids qu'il n'en faut à son homologue non-bouclé (le réseau FIR) pour accomplir cette tâche. Enfin, de meilleures prédictions à court terme ont été accomplies avec le DRNN.

Le Chapitre 5 fut consacré à la prédiction des conditions environnementales des télescopes terrestres avec le modèle DRNN. L'objet de cette étude fût de fournir une évaluation préliminaire des performances prédictives de la future *Astronomical Weather Station* dont l'installation est envisagée sur le site de Paranal à la fin de ce siècle. En outre, nous avons posé les premiers jalons d'une stratégie opérationnelle vouée à la gestion des programmes d'observations astronomiques. Les qualités prédictives du DRNN ont été évaluées grâce aux résultats obtenus avec d'autres méthodes neuronales et statistiques, en particulier la technique de reconstruction locale du *k-nearest neighbor*.

Aux antipodes de la prédiction des suites temporelles sans connaissances *a priori* du système sous-jacent, le Chapitre 6 est dédié à l'application des réseaux récurrents à la modélisation d'un réseau ATM. Notre attention s'est portée sur le vaste problème de la gestion des ressources grâce au contrôle d'admission des appels et au routage optimal.

La faculté d'adaptation est la qualité indispensable dont il faut doter le contrôleur ATM pour appréhender efficacement l'impact des trafics, décrits par des grandeurs statistiques sommaires, sur le réseau en perpétuelle évolution. Aussi, nous avons examiné comment modéliser chaque tampon dans les nœud de commutation à l'aide de trois neurones artificiels, à la lumière de la théorie de files d'attentes. Notre choix s'est porté sur le Réseau de Neurones Aléatoires (RNA) pour bénéficier de la forme rationnelle de la fonction d'activation du neurone aléatoire.

Une simulation est réalisée au cours de laquelle le neuro-contrôleur sélectionne pour chaque requête d'établissement d'appel, un chemin au délai de transmission minimal au vu de l'occupation moyenne estimée par le RNA des tampons sur ce chemin. Ce délai est ensuite comparé à la qualité de service souhaitée par l'utilisateur sous la forme d'un facteur de qualité relatif. Le modèle affiche des performances honorables et se montre capable de maintenir un taux d'utilisation important tout en évitant soigneusement les risques de congestion.

Bibliographie

[Akaike 74] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control* **AC-19**, pp. 716-722, 1974.

[Almeida 87] L.B. Almeida, "Backpropagation in perceptrons with feedback," *Neural Computers*, Eds. R. Eckmiller et Ch. von der Malsburg, Springer-Verlag, Neuss, pp. 199-208, 1987.

[Atalay 92] V. Atalay, E. Gelenbe et N. Yalabik, "The random neural network model for texture generation," *International Journal of Pattern Recognition and Artificial Intelligence* **6**, No. 1, pp. 131-144, 1992.

[Atalay 92] V. Atalay et E. Gelenbe, "Parallel algorithm for coulour texture generation using the random neural network model," *International Journal of Pattern Recognition and Artificial Intelligence* **6**, No. 2 & 3, pp. 437-446, 1992.

[Aussem 92a] A. Aussem, "Application of genetic algorithms as adaptive models for water demands," rapport interne, Institut National des Télécommunications, Evry, France, 1992.

[Aussem 92b] A. Aussem, "Financial time series prediction with recurrent neural networks," mémoire de stage de fin d'études, Institut National des Télécommunications, Evry, France, 1992.

[Aussem 94] A. Aussem, "Call admission control in ATM networks with the Random Neural Network," *Proceedings of the IEEE International Conference on Neural Networks* **4**, pp. 2482-2487, 1994.

[Aussem 95a] A. Aussem, F. Murtagh et M. Sarazin, "Dynamical recurrent neural networks and pattern recognition methods for time series prediction: application to seeing and temperature forecasting in the context of ESO's VLT Astronomical Weather Station," *Vistas in Astronomy* **38**, pp. 352-374, 1995.

[Aussem 95b] A. Aussem, F. Murtagh et M. Sarazin, "Fuzzy astronomical seeing nowcasts with a dynamical and recurrent connectionist network," à paraître dans *Neurocomputing*, 1995.

[Aussem 95c] A. Aussem, F. Murtagh et M. Sarazin, "Dynamical recurrent neural networks - towards environmental time series prediction," à paraître dans *International Journal on Neural Systems*, 1995.

[Aussem 95d] A. Aussem, "Training dynamical recurrent neural networks with the temporal recurrent back-propagation algorithm: Application to time series and characterization," soumis à *IEEE Transactions on Neural Networks*, 1994.

[Aussem 95e] A. Aussem, "Call admission control and optimal routing in ATM networks with distributed Random Neural Networks" soumis à *IEEE Transactions on Neural Networks*, 1995.

[Azencott 86] R. Azencott et D. Dacunha-Castelle, *Series of Irregular Observations*. Springer-Verlag, New York, 1986.

[Baldi 95] P. Baldi, "Gradient descent learning algorithm overview: A general dynamical systems perspective," *IEEE Trans. on Neural Networks* **6**, pp. 182-195, 1995.

[Baum 89] E.B. Baum et D. Haussler, "What size gives generalization?," *Neural Computation* **1**, pp. 151-160, 1989.

[Becker 92] M. Becker, A.L. Beylot et Z. Choueri, "A Markov model of sources for ATM traffic: Superposition of on-off voice, data and video sources," *Int. Symp. on Comp. and Information Sciences VII*, Eds. E. Gelenbe, U. Halici and N. Yalabik, pp. 335-342, 1992.

[Becker 93] M. Becker et P. Douillet, "Hierarchical simulation for rare events," *Proceedings IASTED International Conference, Modelling and Simulation Conference*, Pittsburgh, 1993.

[Bengio 94] Y. Bengio, P. Simard et P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. on Neural Networks* **5**, pp. 157-166, 1994.

[Beylot 93] A.L. Beylot, *Modèles de Trafics et de Commutateurs pour l'Evaluation de la Perte et du Délai*. Thèse de Doctorat, Université Paris VI, Paris, 1993.

[Beylot 94] A.L. Beylot, I. Kohlenberg et M. Becker, "Performance analysis of an ATM switch based on three-stage Clos interconnection network under non-uniform traffic patterns," *Proceedings of Broadband Communications, Conference IFIP TC6*, Paris, 1994.

[Bouleau 88] N. Bouleau, *Processus Stochastiques et Applications.* Hermann, Paris, 1988.

[Box 76] G.E. Box et G.M. Jenkins, *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco, 1976.

[Bryson 69] A.E. Bryson et Y.-C. Ho, *Applied Optimal Control.* New York: Blaisdell, 1969.

[Bylinskym 93] G. Bylinskym, "Computers that learn by doing," *Fortune*, pp. 96-102, Septembre, 1993.

[Casdagli 89] M. Casdagli, "Nonlinear prediction of chaotic time series,", *Physica D* **35**, pp. 335-356, 1989.

[Casdagli 92] M. Casdagli et S. Eubanks, *Nonlinear Modeling and Forecasting*, Proceedings Vol. XII, Santa Fe Institue Studies in the Science of Complexity, Addison-Wesley Publishing Company, 1992.

[Charbit 90] M. Charbit, *Eléments de Théorie du Signal: Les Signaux Aléatoires.* Ellipses, Collection pédagogiques des télécomminications, 1990.

[Chien 91] X. Chien, I.M. Leslie, "A neural network approach towards adaptive congestion control in broadband ATM networks," *Proc. IEEE Globecom*, pp. 115-119, 1991.

[Cohen 83] M.A. Cohen et S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Transactions on Systems, Man and Cybernetics* **13**, pp. 815-826, 1993.

[Connor 94] J.T. Connor, R.D. Martin, an L.E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Trans. on Neural Networks* **5**, pp. 240-254, 1994.

[Cost 91] COST 224 Committee, "COST 224 Final Report: Performance Evaluation and Design of Multiservice Networks," Eds. J. Roberts, Paris, 1991.

[Coudreuse 90] J.P. Coudreuse, P. Boyer, D. Tranchier, "Contrôle de trafic et gestion des ressources," *L'Echo des Recherches*, No. 144, 1991.

[Cover 65] T.M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers* **14**, pp. 326-334, 1965.

[Cox 66] D.R. Cox et P.A. Lewis, *The Statistical Analysis of Series of Events*. London: Methuen 1966.

[Cragg 55] B.G. Cragg et H.N.V. Temperley, "Memory: the analogy with ferromagnetic hysteresis," *Brain* **78**, II, pp. 304-316, 1955.

[Cybenko 89] G. Cybenko, "Continuous value neural networks with two hidden layers are sufficient," *Math. Control Signals and Systems* **2**, pp. 303-314, 1989.

[Davalo 89] E.Davalo et P. Naïm, *Les Réseaux de Neurones*. Eyrolles, Paris, 1989.

[Debar 92] H. Debar, M. Becker and D. Siboni, "Hyperview and intelligent security supervisor," *ICIN'92 Second International Conference on Intelligence in Networks*, pp. 177-182, 1992.

[Delmas 92] J.P. Delmas, *Introduction aux Probabilités*. Cours, Institut National des Télécommunications, Evry, France, 1992.

[Elman 90] J.L. Elman, "Finding structure in time," *Cognitive Science* **14**, pp. 179-211, 1990.

[Farmer 88] J.D. Farmer et J. Sidorowich, "Exploiting chaos to predict the future and reduce noise," Y.C. Lee, Ed., *Evolution, Learning, and Cognition*, World Scientific Press, pp. 277, 1988.

[Frasconi 92] P. Frasconi, M. Gori, et G. Soda, "Local feedback multilayered networks," *Neural Computation* **4**, pp. 120-130, 1992.

[Gelenbe 89a] E. Gelenbe, "Réseaux stochastiques ouverts avec clients négatifs et positifs", *Comptes Rendus Acad. Sci. Paris*, t. 309, Série II, pp. 979-982, 1989.

[Gelenbe 89b] E. Gelenbe, "Random Neural Networks with negative and positive signals and product form solution," *Neural Computation* **I**, No. 4, pp. 502-510, 1989.

[Gelenbe 90a] E. Gelenbe, "Réseaux neuronaux aléatoires stables," *Comptes Rendus Acad. Sci. Paris*, t. 310, Série II, pp. 177-180, 1990.

[Gelenbe 90b] E. Gelenbe, "Sable Random Neural Networks," *Neural Computation* **2**, No. 2, pp. 239-247, 1990.

[Gelenbe 91a] E. Gelenbe, "Queueing networks with negative and positive customers," *Journal of Applied Probability* **28**, pp. 656-663, 1991.

[Gelenbe 91b] E. Gelenbe, P. Glynn et K. Sigman, "Queues with negative arrivals," *Journal of Applied Probability* **28**, pp. 245-250, 1991.

[Gelenbe 91c] E. Gelenbe, *Neural Networks: Advances and Applications I*, North-Holland Pub. Co., Amsterdam, 1991.

[Gelenbe 92a] E. Gelenbe, *Neural Networks: Advances and Applications II*, North-Holland Pub. Co., Amsterdam, 1992.

[Gelenbe 92b] E. Gelenbe et M. Schassberger, "Stability of Product form G-Networks," *Probability in the Engineering and Informational Sciences* **6**, pp. 271-276, 1992.

[Gelenbe 93a] E. Gelenbe, "G-Networks with instantaneous customer movement,"*Journal of Applied Probability*, Amsterdam 30 (3), pp. 742-748, 1993.

[Gelenbe 93b] E. Gelenbe, "G-Networks with signals and batch removal," *Probability in the Engineering and Informational Sciences* **7**, pp. 335-342, 1993.

[Gelenbe 93c] E. Gelenbe, "Learning in the recurrent Random Neural Network model," *Neural Computation* **V**, pp. 154-164, 1993.

[Gelenbe 94] E. Gelenbe, "G-Networks: An unifying model for queueing networks and neural networks," *Annals of Operations Research* **48**, pp. 443-461, 1994.

[Geman 92] S. Geman, E. Bienenstock et R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation* **4**, no. 1, pp. 1-58, 1992.

[Gershenfeld 93] N.A. Gershenfeld et A.S. Weigend, "The future of time series: learning and understanding," *Time Series Prediction: Forecasting the Future and Understanding the Past*, Eds. A.S. Weigend et N.A. Gershenfeld, Redwood City, CA: Addison-Wesley, pp. 1-70, 1993.

[Hammel 85] S. Hammel, C. Jones et J.V. Moloney, "Global dynamical systems, and bifurcations of vector fields," *J. Opt. Soc. Am.* **B-2**, pp. 552, 1985.

[Hammerstrom 93] D. Hammerstrom, "Neural networks at work," *IEEE Spectrum*, pp. 26-32, Juin 1993.

[Haykin 93] S. Haykin, *Neural Networks.* MacMillan Publishing Company, New York, 1993.

[Hebb 49] D.O. Hebb, *The Organization of Behavior.* New York: Wiley, 1949.

[Heffes86] H. Heffes, D.M. Lucantoni, "Markov Modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4, No. 6, 1986, pp. 856-867.

[Hertz 91] J. Hertz, A. Krogh et R. Palmer, *An Introduction to the Theory of Neural Computation.* Redwood City, CA: Addison-Wesley, 1991.

[Hinton 86] G.E. Hinton et T.J. Sejnowski, "Learning and relearning in Boltzman machines," in *Parallel Distributed Processing* **1**, chap. 7. Voir [Rumelhart 86].

[Hiramatsu 90] A. Hiramatsu, "ATM communications network control by neural network," *IEEE Transactions on Neural Networks* **I**, No. 1, 1990, pp. 122-130.

[Hodgkin 52] A.L. Hodgkin et A.F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, Londres 117:500-554, 1952.

[Hopfield 82] J.J. Hopfield, "Neural networks and physical systems with emergent collective abilities," *Proceedings of the National Academy of Sciences, USA* **81**, pp. 3088-3092, 1982.

[Hopfield 84] J.J. Hopfield, "Neurons with graded responses have collective computational properties like those of two-state neurons,"*Proceedings of the National Academy of Sciences, USA* **84**, pp. 8429-8433, 1984.

[Hubner 93] U.W. Hubner, C.O. Weiss, N.B. Abraham et D. Tang, "Lorenz-like chaos in NH_3-FIR Lasers," in *Time Series Prediction: Forecasting the Future and Understanding the*

Past, Eds. A.S. Weigend et N.A. Gershenfeld, Redwood City, CA: Addison-Wesley, pp. 243-265, 1993.

[Hui 87] J. Hui, E. Arthurs, "A broadband packet switch for integrated transport," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-5, No. 8, 1987, pp.1264-1273.

[Intel 93] Intel *Pentium^{TM} Processor User's Manual.* Intel Corporation Literature Sales, Mt. Prospects, IL, 1993.

[Jordan 86] M. Jordan, "Attractor dynamics and parallelism in a connectionist sequential machine," *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pp. 531-546, 1986.

[Jorgenson 91] M.B. Jorgenson, G.J.M. Aitken et E.K. Hege, "Evidence of a chaotic attractor in star-wander data," *Optics Letters* **16**, pp. 64-66, 1991.

[Kailath 80] T. Kailath, *Linear Systems.* Prentice-Hall, Englewood Cliffs, New York, 1980.

[Kalman 61] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng. Ser. D*, **82**, pp. 35-45, 1961.

[Kandel 87] E.R. Kandel et J.H. Schwartz, *Principle of Neural Science.* New York: Elsevier/North Holland, 1981.

[Karol 87] M. Karol, M. Hluchij, S. Morgan, "Input versus output queueing on a space-division packet switch," *IEEE Transactions on Communications*, Vol. COM-35, No. 12, 1987, pp. 1347-1356.

[Kleinrock 76] L. Kleinrock, *Queueing Systems.* Vol. I & II, John Wiley & Sons, New York 1976.

[Lang 88] K. Lang et G. Hinton, "Time-Delay Neural Network architecture for speech recognition," rapport technique CMU-CS-88-152, Carnegie-Mellon University, Pittsburgh, 1988.

[Lapedes 87] A. Lapedes et Farber, "Non linear signal prediction using neural networks: prediction and system modelling, Preprint," Los Alamos National Laboratory LA-UR-87-2662, 1987.

[Le Cun 85] Y. Le Cun, "Une procédure d'apprentissage pour réseau à seuil assymétrique," in *Cognitiva 85: A la Frontière de* *l'Intelligence Artificielle des Sciences de la Connaissance des Neuronsciences*, CESTA, Paris, pp. 599-604, 1985.

[Leighton 91] R. Leighton et B. Conrad, "The autoregressive backpropagation algorithm," *Proceedings of the International Joint Conference on Neural Networks*, 1991.

[Liadis 90] I. Liadis, W. Denzel, "Performance of packets switches with input and output queues," *Proc. IEEE Infocom*, pp. 747-753, 1990.

[Lorenz 63] E.N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, pp. 130-141, 1963.

[MacGregor 87] R. MacGregor, *Neural and Brain Modeling*, Academic Press, New York 1987.

[Mackey 77] M. Mackey et L. Glass, "Oscillations and chaos in phsysiological control systems," *Science*, pp. 197-287, 1977.

[McCulloch 43] W.S. McCulloch et W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull Math. Biophysics,* 5:115-113, 1943.

[Minsky 67] M.L. Minsky, *Computation: Finite and Infinite Machines.* Englewood Cliffs: Prentice-Hall, 1967.

[Minsky 69] M.L. Minsky et S.A. Papert, *Perceptrons.* Cambridge: MIT Press, 1969.

[Miyamoto 90] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis.* Dordrecht, Kluwer, 1990.

[Mokhtari 94] M. Mokhtari, *Réseau Neuronal Aleatoire: Applications à l'Apprentissage et à la Reconnaissance d'Images.* Thèse de Doctorat, Université René Descartes, Paris, 1994.

[Moore 91] C. Moore, "Generalized Shifts: Unpredictability and undecitaptibility in dynamical systems," *Nonlinearity* **4**, pp. 199-230, 1991.

[Mozer 93] M.C. Mozer, "Neural Net architectures for temporal sequence processing," *Time Series Prediction: Forecasting the Future and Understanding the Past*, Eds. A.S. Weigend et N.A. Gershenfeld, Redwood City, CA: Addison-Wesley, pp. 243-265, 1993.

[Murtagh 92] F. Murtagh, M. Sarazin et H.M. Adorf, "Statistical prediction of astronomical seeing and of telescope thermal environment," *ESO Conference on Progress in Telescope and Instrumentation Technologies*, Ed. M.H. Ulrich, pp. 377. Garching, Germany, 1992.

[Murtagh 93] F. Murtagh et M. Sarazin, "Nowcasting astronomical seeing: a study of ESO La Silla and Paranal," *Publications of the Astronomical Society of the Pacific* **105**, 932-939, 1993

[Murtagh 94] F. Murtagh, A. Aussem et M. Sarazin, "Nowcasting astronomical seeing: towards an operational approach," à paraître dans *Publications of the Astronomical Society of the Pacific*, 1995.

[Narendra 91] K.S. Narendra et K. Parthasarathy, "Gradient methods for the optimization of linear control systems," *IEEE Trans. on Neural Networks*, pp. 252-262, 1991.

[Nichols 92] J. Nichols, R. Martin et B. Wallice, *From Neuron to Brain.* Sinauer Associates, Inc. Publishers, 1992.

[Nikolaidis 92] I. Nikolaidis, I. Akyildiz, "Source characterization and statistical multiplexing in ATM networks," rapport de recherche, College of Computing, Georgia Institute of Technology, 1992.

[Oie 92] Y. Oie, M. Murata, K. Kubota, H. Miyahara, "Performance analysis of nonblocking packet switch with input and output buffers," *IEEE Transactions on Communications* **40**, No. 8, August 1992, pp. 1294-1297.

[Ott 93] E. Ott, *Chaos in Dynamical Systems.* Cambridge University Press, 1993.

[Packard 80] N.H. Packard, J.P. Crutchfeld, J.D. Farmer et R.S. Shaw, "Geometry from a time series," *Physical Review Letter* **45(9)**, pp. 712-716, 1980.

[Parker 85] D.B. Parker, "Learning logic," Rapport technique, Center for Computational Research in Economics and Management Science, M.I.T., Cambridge, MA, 1985.

[Peretto 84] P. Peretto, "Collective properties of neural networks: a statistical physics approach," *Biological Cybernetics* **50**, pp. 51-62, 1984.

[Piche 94] S.W. Piche, "Steepest descent algorithms for neural network controllers and filters," *IEEE Trans. on Neural Networks* **5**, pp. 198-212, 1994.

[Pineda 87] F.J. Pineda, "Generalization of back-propagation to recurrent neural networks," *Physical Review Letters* 59, pp. 2229-2232, 1987.

[Pineda 89] F.J. Pineda, "Recurrent back-propagation and the dynamical approach to adaptive neural computation," *Neural Computation* **1**, pp.161-172, 1989.

[Press 88] W.H. Press, B.P. Flannery, S.A. Teukolsky et W.T. Vetterling, *Numerical Recipes in C.* Cambridge University Press, 1988.

[Priestley 81] M.B. Priestley, *Spectral Analysis and Time Series.* Academic Press, 1981.

[Priestley 93] M.B. Priestley, "A short history of time series," *International Conference on Applications of Time Series Analysis in Astronomy and Meteorology*, Padua, Septembre 1993.

[Qian 88] N. Quian et T. Sejnowski, "Leraning to solve random-dot stereograms of dense transparent surfaces with recurrent back-propagation," in *Proceedings of the 1988 Connection Models Summer School*, (Pittsburgh 1988), Eds. D. Touretzky, G. Hinton et T. Sejnowski, San Mateo: Morgan Kaufmann, pp. 435-443, 1989.

[Rohwer 87] R. Rohwer 87 et B. Forrest, "Training time-dependence in neural networks," *IEEE First International Conference on Neural Networks*, San Diego, pp. 701-708, 1987.

[Rosenblatt 62] F. Rosenblatt, *Principles of Neurodynamics.* New York: Spartan, 1962.

[Ruelle 85] D. Ruelle et J.P. Eckmann, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**, pp. 617-656, 1985.

[Rumelhart 86] D.E. Rumelhart, G.E. Hinton et R.J. Williams, "Learning internal representations by error propagation," *Parallel Distributed Processing Explorations in the Microstructure of Cognition*, Eds. D.E. Rumelhart et J.L. McClelland, Cambridge, MA: MIT Press, Bradfords Books, vol. 1, pp. 318-362, 1986.

[Sarazin 92] M. Sarazin, *Caractérisation des Propriétes Optiques de la Turbulence Atmospherique : Apllication au Choix du Site du VLT de l'Observatoire Europeen Austral (ESO).* Thèse de Doctorat, Université D'Aix-Marseille III, Marseille, 1992.

[Sastry 94] P.S. Sastry, G. Santharam et K.P. Unnikrishnan, "Memory neuron networks for identification and control of dynamical systems," *IEEE Trans. on Neural Networks* **5**, pp. 306-319, 1994.

[Sauer 93] T. Sauer, "Time Series Prediction by using Delay Coordinate Embedding," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, Eds. A.S. Weigend et N.A. Gershenfeld, Redwood City, CA: Addison-Wesley, pp. 175-193, 1993.

[Schmidhuber 92] J. Schmidhuber, "A fixed size storage $O(n^3)$ time complexity learning algorithm for fully recurrent continually running networks," *Neural Computation* **4**, No. 2, pp. 243-248, 1992.

[Schwartz 90] D.B. Schwartz, V.K. Salamam, S.A. Solla et J.S. Denker, "Exhaustive learning," *Neural Computation* **2**, pp. 371-382, 1990.

[Schwartz 92] E. Schwartz et J. Treece, "Smart programs go to work: how applied intelligence sofware makes decisions for the real world," *Business Week*, pp. 97-105, Mars 1992.

[Sejnowski 86] T.J. Sejnowski et C.R. Rosenberg, "NETTALK: a parallel network that learns to read loud," Rapport technique, The John Hopkins University Electrical Engineering and Computer Science, JHU/EECS-86/01, 1986.

[Simard 89] P.Y. Simard, M.B. Ottaway et D.H. Ballard, "Analysis of recurrent back-propagation," in *Proceedings of the 1988 Connection Models Summer School*, (Pittsburgh 1988), Eds. D. Touretzky, G. Hinton et T. Sejnowski, San Mateo: Morgan Kaufmann, pp. 103-112, 1989.

[Srinivasan 94] B. Srinivasan, U.R. Prasad et N.J. Rao, "Back propagation through adjoints for the identification of nonlinear dynamic systems using recurrent neural networks," *IEEE Trans. on Neural Networks* **5**, pp. 213-228, 1994.

[Subba Rao 84] T. Subba Rao and M.M. Gabr, "An introduction to bispectral analysis and bilinear time series models," *Lecture Notes in Statistics* **24**, Springer, New York, 1984.

[Tank 87] D. Tank et J. Hopfield, "Concentrating information in time: Analog neural networks with applications to speech recognition," *First International Conference on Neural Networks* **4**, pp. 455-468, 1987.

[Takens 80] F. Takens, "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, Eds. D.A. Rand et L.S. Young, *Lecture Notes in Mathematics*, No. 898, Springer-Verlag, Berlin, 1980, pp. 336-381.

[Tong 80] H. Tong et K.S. Lim, "Threshold autoregression, limit cycles and cyclical data," *Journal Royal Statistical Society B* **42**, pp 245, 1980.

[Tong 90] H. Tong, *Non Linear Time Series.* Clarendon Press Oxford, 1990.

[Tran-Gia 91] P. Tran-Gia et O. Gropp, "Structure and performance of neural nets in broadband system admission control," rapport de recherche, Institute of Computer Science, University of Würzburg, 1991.

[Tsoi 94] C.T. Tsoi et A.D. Back, "Locally recurrent globally feedforward networks: a critical review of architectures," *IEEE Trans. on Neural Networks* **5**, pp. 229-239, 1994.

[Vapnik 71] V.N. Vapnik et A.Y. Chervonenkis, "On the union convergence of relative frequencies of events to their probabaility," *Theory of Probability and Its Applications* **16**, pp. 262-280, 1971.

[Volterra 59] V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations.* New York:Doover, 1959.

[Von Neumann 56] J. Von Neumann, "Probabilistics Logics and the synthesis of reliable organisms from unreliable components," in *Automata Studies*, Eds. C.E. Shannon et J. McCarthy, Princeton University Press, pp. 43-98, 1956.

[de Vries 92] B. de Vries et J. Principe, "The Gamma Model - A new neural model for temporal processing," *Neural Computation* **5**, pp. 565-576, 1992.

[Wan 93a] E.A. Wan, "Time series prediction by using a connectionist network with delay lines," *Time Series Prediction: Forecasting the Future and Understanding the Past*: Eds. A.S. Weigend et N.A. Gershenfeld, Redwood City, CA: Addison-Wesley, pp. 195-207, 1993.

[Wan 93b] E.A. Wan, *Finite Impusle Response Neural Networks with Applications in Time Series Prediction.* Thèse de Doctorat, Stanford University, Novembre, 1993.

[Weigend 90] A.S. Weigend, D.E. Rumelhart et B.A. Huberman, "Predicting the future: a connectionist approach," *International Journal of Neural Systems* **1**, pp. 195-220, 1990.

[Werbos 74] P. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in Behavioral Sciences.* Thèse de Doctorat, Harvard University, Cambridge, MA, Août 1974.

[Widrow 60] B. Widrow et H. Hoff, in *IRE Western Electric Show and Convention Record*, pp. 96-104, 1960.

[Widrow 62] B. Widrow, "Generalization and information storage in networks of "Adaline" neurons," in *Self-Organizing Systems*, Eds. M.C. Yovits, G.T. Jacobi et G.D. Goldstein, Washington, 435-461, 1962.

[Widrow 85] B. Widrow et S.D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, New York, Prentice Hall, 1985.

[Williams 89] R.J. Williams et D. Zipser, "Experimental analysis of the real-time recurrent learning algorithm," *Connection Science* **1**, pp. 87-111, 1989.

[Wilkinson 65] J. H. Wilkinson, *The Algebraic Eigenvalue Problem.* Eds. L. Fox et J. Walsh, Oxford University Press, 1965.

[Yule 27] G.U. Yule, "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers," *Philos. Trans. Roy. Soc. London Ser. A*, vol. 226, pp. 267, 1927.