



HAL
open science

Intégration de Réseaux de Neurones pour la Télémétrie Laser

Laurent Gatet

► **To cite this version:**

Laurent Gatet. Intégration de Réseaux de Neurones pour la Télémétrie Laser. Micro et nanotechnologies/Microélectronique. Institut National Polytechnique (Toulouse), 2007. Français. NNT : 2007INPT027H . tel-04622074

HAL Id: tel-04622074

<https://ut3-toulouseinp.hal.science/tel-04622074v1>

Submitted on 24 Jun 2024

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

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

DOCTORAT

de l' UNIVERSITE DE TOULOUSE

*dé*livré par

l' INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE

Ecole Doctorale : Génie Electrique, Electronique, Télécommunications

Spécialité : Conception des Circuits Microélectroniques et Microsystèmes

par

M. Laurent GATET

<p>Intégration de</p> <p>Réseaux de Neurones</p> <p>pour la Télémétrie Laser</p>
--

Soutenue le 27 Septembre 2007 devant le jury composé de :

M.	Pascal	FOUILLAT	Président
M.	Gérard	DREYFUS	Rapporteur
M ^{me}	Fan	YANG SONG	Rapporteur
M.	Marc	LESCURE	Examineur
M.	Daniel	ROVIRAS	Examineur
M.	Luc	LAPIERRE	Examineur
M ^{me}	Hélène	TAP-BETEILLE	Directrice de thèse

Thèse préparée au Laboratoire d'Optoélectronique pour les Systèmes Embarqués
INP – ENSEEIHT

Remerciements

Je souhaite d'abord remercier du fond du cœur M^{me} Hélène TAP-BETEILLE pour son encadrement, ses conseils et son soutien sans faille durant cette thèse. Ses nombreuses qualités, telles que sa motivation inaltérable et sa grande facilité de communication, m'ont permis de mettre en valeur jour après jour l'ensemble des travaux effectués.

Je tiens aussi à exprimer ma profonde gratitude à M. Marc LESCURE pour les conseils qu'il m'a prodigués au cours de ces trois années. La pertinence de ses critiques et ses nombreuses connaissances dans les domaines de l'optoélectronique et de la conception de circuits analogiques m'ont permis d'explorer de nombreuses voies et d'étoffer considérablement mes travaux de recherche.

Je souhaite remercier M. Daniel ROVIRAS pour l'expérience qu'il a su me transmettre tout au long de cette thèse dans les domaines liés à l'utilisation et au paramétrage des réseaux de neurones.

Je tiens à témoigner toute ma reconnaissance à M. Pascal FOUILLAT, président du jury de soutenance, à M. Gérard DREYFUS et M^{me} Fan YANG SONG, rapporteurs de ce manuscrit, et enfin à M. Luc LAPIERRE pour avoir accepté de faire partie de mon jury de thèse.

Je souhaite aussi remercier M. Francis BONY pour les compétences qu'il a mises en œuvre lors de la conception du réseau de neurones numérique et pour la bonne humeur qu'il affiche jour après jour au sein du laboratoire LOSE.

Un très grand merci à M. Francis JAYAT pour l'aide qu'il m'a apportée lors de la réalisation des circuits imprimés de test des différents réseaux de neurones intégrés, et à M. Clément TRONCHE pour l'ensemble du matériel qu'il a mis à ma disposition au moment des tests de caractérisation des réseaux.

Je tiens à exprimer ma profonde gratitude à M. Thierry BOSCH, directeur du laboratoire LOSE, pour tous les conseils qu'il a su me prodiguer au cours de ces trois années. Les diverses connaissances acquises me seront particulièrement bénéfiques pour l'ensemble de mes travaux à venir.

Je souhaite aussi témoigner toute ma reconnaissance à M. Pascal TANNOU, M^{me} Danielle ANDREU, à M^{me} Nicole CAUSERO et M^{me} Catherine MONTELS pour le soutien permanent qu'ils m'ont apporté tout au long de ces trois années.

Je remercie aussi M. Michel CATTOEN, M. Han CHENG, M. Julien PERCHOUX, M. Olivier BERNAL, M. Denis STANDAROWSKI, M. Pierre LAQUERRE, M^{elle} Caroline BES,

Remerciements

l'ensemble des doctorants (M. Joseph EL ASSAD, M. Saroj PULLTEAP, M^{me} Maha SULLEMAN, M. Usman ZABIT, M. Emmanuel MOUTAYE, ...), et autres personnels du laboratoire pour le sérieux et la bonne humeur qu'ils amènent ou ont amené au jour le jour.

Je souhaite enfin témoigner toute ma gratitude à l'ensemble des personnes qui m'ont soutenu tout au long de cette thèse. Une grande partie de la qualité des résultats obtenus leur incombe.

Mots Clés

- Réseaux de neurones
- Technologie CMOS basse tension
- Classification de surfaces
- Mesure de distance
- Circuits intégrés analogiques (ASIC)
- Télémétrie laser

Résumé

Grandes lignes : Un réseau de neurones est une architecture paramétrable composée de plusieurs modules appelés neurones. Ils peuvent être utilisés pour compenser des variations non souhaitées de certains phénomènes physiques ou pour effectuer des tâches de discrimination.

Un réseau de neurones a été intégré en technologie CMOS basse tension pour être implanté au sein d'un télémètre laser par déphasage. Deux études ont été menées en parallèle. La première consiste à lever l'indétermination sur la mesure de distance déduite de la mesure de déphasage. La seconde étude permet la classification de différents types de surfaces à partir de deux signaux issus du télémètre.

Résumé détaillé : Un réseau de neurones a la faculté de pouvoir être entraîné afin d'accomplir une tâche d'approximation de fonction ou de classification à partir d'un nombre limité de données sur un intervalle bien défini. L'objectif de cette thèse est de montrer l'intérêt d'adapter les réseaux de neurones à un type de système optoélectronique de mesure de distance, la télémétrie laser par déphasage.

La première partie de ce manuscrit développe de manière succincte leurs diverses propriétés et aptitudes, en particulier leur reconfigurabilité par l'intermédiaire de leurs paramètres et leur capacité à être intégré directement au sein de l'application. La technique de mesure par télémétrie laser par déphasage est développée dans le deuxième chapitre et comparée à d'autres techniques télémétriques.

Le troisième chapitre montre qu'un réseau de neurones permet d'améliorer nettement le fonctionnement du télémètre. Une première étude met en valeur sa capacité à accroître la plage de mesure de distance sans modifier la résolution. Elle est réalisée à partir de mesures expérimentales afin de prouver le réel intérêt de la méthode comportementale développée. La deuxième étude ouvre une nouvelle perspective relative à l'utilisation d'un télémètre laser par déphasage, celle d'effectuer la classification de différents types de surfaces sur des plages de distances et d'angles d'incidence variables.

Pour valider expérimentalement ces deux études, les cellules de base du neurone de type perceptron multi-couches ont été simulées puis implantées de manière analogique. Les phases de simulation, de conception et de test du neurone analogique sont détaillées dans le quatrième chapitre. Un démonstrateur du réseau de neurones global a été réalisé à partir de neurones élémentaires intégrés mis en parallèle. Une étude de la conception des mêmes cellules en numérique est détaillée succinctement dans le cinquième chapitre afin de justifier les avantages associés à chaque type d'intégration.

Le dernier chapitre présente les phases d'entraînement et de validation expérimentales du réseau intégré pour les deux applications souhaitées. Ces phases de calibrage sont effectuées extérieurement à l'ASIC, par l'intermédiaire de l'équation de transfert déterminée après caractérisation expérimentale et qualification du réseau de neurones global. Les résultats expérimentaux issus de la première étude montrent qu'il est possible d'obtenir à partir des signaux de sorties du télémètre et du réseau de neurones, une mesure de distance de précision ($50\mu\text{m}$) sur un intervalle de mesure 3 fois plus important que celui limité à la mesure du déphasage. Concernant l'application de discrimination de surfaces, le réseau de neurones analogique implanté est capable de classer quatre types de cibles sur l'intervalle $[0.5\text{m} ; 1.25\text{m}]$ pour un angle d'incidence pouvant varier de $-\pi/6$ à $\pi/6$.

Keywords

- Neural networks
- Low-voltage CMOS technology
- Surface classification
- Distance measurement
- Analog integrated circuits (ASIC)
- Laser rangefinding systems

Abstract

Outline : A neural network is a trainable structure composed by modules called neurons. They may be used in order to compensate adverse variations of physical phenomenon or to achieve discrimination tasks.

Two studies were held in order to integrate a neural network in low voltage CMOS technology in a phase-shift laser rangefinder. The first one consists in raising the indecision on distance measurement deduced from the phase-shift measurement. The aim of the second study is to classify different kinds of surfaces using two signals issued from the rangefinder.

Detailed abstract : A neural network has the capability to be trained in order to approximate functions or to achieve classification from a limited number of data on a well defined interval. The first part of the manuscript develops succinctly their various properties and aptitudes, particularly their reconfigurability through they parameters and their capability to be integrated directly in the application.

The aim of this thesis is to demonstrate the interest of adapting neural networks to a type of distance measurement optoelectronic system, the phase-shift laser rangefinding. This measurement technique is developed in the second chapter and compared to other rangefinding techniques.

The third chapter demonstrates that a neural network allows to improve considerably the rangefinder functioning. A first study highlights its capability to increase the distance measurement range without modifying the resolution. It is achieved from experimental measurements, in order to prove the real interest of the developed behavioural method. In a

second study, the same neural network structure is used in order to show its capability to discriminate different types of surfaces on variable distance and incidence angle ranges.

The main cells of the multi-layer perceptron-type neuron were simulated then implanted in analog. A conception study of the same cells in digital were achieved in order to justify the advantages associated to each type of integration. The simulation, conception and test stages are detailed in the fourth chapter. The whole neural network were achieved from elementary integrated neurons in parallel. The digital version of the neuron is succinctly detailed then compared to the analog structure in the fifth chapter.

The last part of the thesis presents the behavioural and test training and validation phases of the integrated network for the two developed applications. These calibrage phases are achieved off-chip through the transfer equation issued from experimental characterisation and qualification of the whole neural network.

Thus, by combining the signals provided by the phasemeter and the neural network outputs, it is possible to reach a distance measurement with high resolution ($50\mu\text{m}$) on a measurement range three times wider than the one limited by the phase-shift measurement. Concerning the surfaces discrimination application, the implanted analog neural network is capable of classifying four types of targets on the interval $[0.5\text{m} ; 1.25\text{m}]$ for a incidence angle varying between $-\pi/6$ and $\pi/6$.

Glossaire

- a_{1ij} : Tension de sortie continue mesurée au niveau du buffer de sortie du multiplieur Mult1ij situé entre l'entrée V_{inNN_i} et le $j^{\text{ème}}$ neurone caché. Si le réseau n'utilise qu'une entrée, l'indice i est supprimé. Si le neurone est étudié seul, l'indice j est supprimé.
- a_{2j} : Tension de sortie continue mesurée au niveau du buffer de sortie de l'additionneur Add1j situé sur le $j^{\text{ème}}$ neurone caché. Si le neurone est étudié seul, l'indice j est supprimé.
- a_{3j} : Tension de sortie continue mesurée au niveau du buffer de sortie de la fonction tangente hyperbolique située sur le $j^{\text{ème}}$ neurone caché. Si le neurone est étudié seul, l'indice j est supprimé.
- a_{4j} : Tension de sortie continue mesurée au niveau du buffer de sortie du multiplieur Mult2j situé entre le $j^{\text{ème}}$ neurone caché et le neurone de sortie. Si le neurone est étudié seul, l'indice j est supprimé.
- $A_d(jf)$: Réponse en fréquence du gain différentiel en boucle ouverte de l'amplificateur transimpédance.
- A_R : Surface de la pupille de l'optique de photoréception.
- a_{VCO} : Vitesse de variation de la fréquence du VCO.
- b_{1j} : Biais relatif à l'additionneur Add1j situé sur le $j^{\text{ème}}$ neurone caché. Si le neurone est étudié seul, l'indice j est supprimé.
- b_2 : Biais relatif à l'additionneur Add2 situé sur le neurone de sortie.
- B_n : Bande passante équivalente de bruit.
- c : Célérité de la vitesse dans le vide ($c = 3.10^8 \text{ m.s}^{-1}$).
- C_{DB} : Capacité drain-substrat du transistor MOS.
- C_{DS} : Capacité drain-source du transistor MOS.
- C_{GB} : Capacité grille-substrat du transistor MOS.
- C_{GD} : Capacité grille-drain du transistor MOS.
- C_{GS} : Capacité grille-source du transistor MOS.
- C_j : Capacité de jonction de la photodiode.
- C_{ox} : Capacité de l'oxyde de grille par unité de surface.
- C_{SB} : Capacité source-substrat du transistor MOS.
- D : Distance cible – capteur.
- D_{min} : Distance minimale expérimentale choisie pour l'application considérée.
- D_{0min} : Distance minimale mesurable avec une modulation linéaire idéale du VCO.
- D_{1appr} : Valeur de distance qui suit D_{min} dans le jeu d'apprentissage.
- e : Valeur de l'erreur issue de la comparaison entre les tensions idéale $V_{outNNid}$ et réelle V_{outNN} en sortie du réseau.
- e_{bin} : Critère permettant de vérifier facilement si l'erreur e est bien inférieure à e_{max} pour toute valeur relative aux différents ensembles de validation.
- e_{max} : Critère de convergence relatif à l'erreur maximale à ne pas dépasser sur un nombre donné d'itérations.
- e_{MEAN} : Ecart moyen constaté entre la caractéristique de test et l'équation du transfert s'en rapprochant le plus.
- f_i : Fréquence de coupure haute associée à la constante de temps τ_i .
- f_{HF} : Haute fréquence, égale à la somme des fréquences f_{RF} et f_{LO} .
- f_{IF} : Fréquence intermédiaire, égale à la différence entre les fréquences f_{RF} et f_{LO} .
- f_{LO} : Fréquence imposée par l'oscillateur local.
- f_{RF} : Fréquence de modulation du signal modulant la diode laser.

- G_{add} : Pente de la fonction linéaire se rapprochant le plus de la caractéristique de transfert de l'additionneur à une précision relative donnée.
- G_{diff} : Pente de la fonction linéaire se rapprochant le plus de la caractéristique de transfert du circuit différentiel à une précision relative donnée.
- G_{dM} : Pente associée à la caractéristique $\Delta I_D(v_d)$ passant par (0;0) et $(I_{dM};v_{dM})$.
- g_{ds} : Conductance petit signal drain-source du transistor MOS.
- G_L : Conductance de sortie.
- G_M : Transconductance de la fonction (pente à l'origine de la caractéristique $\Delta I_D(v_d)$).
- g_M : Transconductance petit signal de la grille du transistor MOS.
- g_m : $m^{ème}$ fonction de régression non linéaire associé à la $m^{ième}$ sortie y_m du réseau de neurones.
- g_{Mb} : Transconductance petit signal du substrat du transistor MOS.
- G_{Mth} : Transconductance associée au circuit différentiel.
- g_{out} : Conductance de sortie petit signal.
- I_0 : Courant de polarisation de la cellule.
- I_D : Courant drain quasi-statique du transistor MOS.
- i_D : Courant drain dynamique petit signal du transistor MOS.
- I_{DL} : Amplitude du courant moyen de modulation de la diode laser.
- I_{dM} : Courant différentiel de sortie quasi-statique dont la valeur correspond à un pourcentage donné du courant de polarisation I_0 .
- I_{in} : Courant d'entrée quasi-statique.
- i_{in} : Courant d'entrée petit signal.
- I_{out} : Courant de sortie quasi-statique.
- i_{out} : Courant de sortie petit signal.
- I_{th} : Courant de polarisation de la fonction tangente hyperbolique.
- i_{pD} : Courant photoélectrique traversant la photodiode (PIN ou APD).
- I_{th} : Courant de seuil de la diode laser.
- K : Entier relatif.
- K_0 : Coefficient de proportionnalité entre la tension de sortie idéale du réseau et la distance.
- K_{01} : Coefficient de proportionnalité entre V_{outNN} et l'inverse de la racine carrée de la V_{inNN} .
- K_1 : Coefficient de proportionnalité relatif à l'émetteur, au transfert optique et à la réponse de la photodiode.
- K'_1 : Coefficient de proportionnalité relatif à la liaison optoélectronique et aux fonctions de transfert du mélangeur et du filtre passe-bande.
- K''_1 : Coefficient de proportionnalité relatif à la liaison optoélectronique et aux fonctions de transfert du mélangeur, du filtre passe-bande et du convertisseur a.c-d.c.
- K_{a1i} : Coefficient de transfert relatif au multiplieur $Mult1ij$ situé entre l'entrée V_{inNN_i} et le $j^{ème}$ neurone caché.. Si le réseau n'utilise qu'une entrée, l'indice i est supprimé.
- K_{a2} : Coefficient de transfert relatif à l'additionneur $Add1j$ situé sur le $j^{ème}$ neurone caché.
- K_{a3} : Coefficient de transfert relatif à la sortie de la fonction tangente hyperbolique.
- K_{a4} : Coefficient de transfert relatif au multiplieur $Mult2j$ situé entre le $j^{ème}$ neurone caché et le neurone de sortie.
- K_{b2} : Coefficient de transfert relatif à l'additionneur $Add2$ situé sur le neurone de sortie.
- K_{sum} : Coefficient de transfert relatif à la somme des sorties des multiplieurs $Mult2j$.
- K_{th} : Coefficient de transfert relatif à l'entrée de la fonction tangente hyperbolique.
- L : Longueur du canal du transistor MOS.
- M : Nombre d'itérations de l'algorithme de mise à jour des poids et biais (méthode de rétropropagation de l'erreur).

n : Nombre d'entrées du réseau de neurones.
 N : Nombre de valeurs d'entrées choisies parmi les valeurs appartenant à l'espace des entrées.
 N_A : Nombre d'exemples de l'ensemble d'apprentissage.
 n_{Dappr} : Nombre de distances choisies pour la phase d'apprentissage.
 n_{Dval} : Nombre de distances choisies pour la phase de validation.
 N_j : Nombre de neurones cachés.
 N_m : Nombre de neurones de sortie.
 n_{ptse} : Nombre d'itérations choisi par l'utilisateur pour vérifier si l'erreur ne dépasse pas la valeur maximale e_{max} .
 n_{Sappr} : Nombre de surfaces choisies pour la phase d'apprentissage.
 n_{Sval} : Nombre de surfaces choisies pour la phase de validation.
 N_T : Nombre d'objets d'un système à discriminer.
 N_V : Nombre d'exemples de l'ensemble de validation.
 n_{\thetaappr} : Nombre d'angles d'incidence choisis pour la phase d'apprentissage.
 n_{\thetaval} : Nombre d'angles d'incidence choisis pour la phase de validation.
 P_{DL} : Puissance crête de l'onde laser transmise.
 R_f : Résistance de contre-réaction de l'amplificateur transimpédance.
 R_L : Résistance de sortie.
 S_D : Sensibilité de la mesure.
 S_i ($i \in \{1;2;3;4\}$) : Ensemble des surfaces à discriminer (S_1 : surface plastifiée, S_2 : papier glacé, S_3 : mur peint et S_4 : surface poreuse).
 S_λ : Réponse spectrale de la photodiode.
 T_r : Période de modulation du VCO.
 T_λ : Coefficient de transmission de l'optique d'émission pour une longueur d'onde donnée.
 T_R : Coefficient de réflexion de l'optique du réception.
 t_{retard} : Temps de retard associé à une ligne à retard.
 V_{BS} : Tension substrat-source du transistor MOS.
 V_{CC} : Tension d'alimentation.
 V_{DS} : Tension drain-source du transistor MOS.
 v_d : Tension d'entrée différentielle d'un circuit différentiel.
 v_{dM} : Tension différentielle d'entrée correspondant à un courant différentiel de sortie quasi-statique égal à I_{dM} .
 V_{GS} : Tension grille-source du transistor MOS.
 V_{inNN_i} : $i^{ème}$ tension d'entrée du réseau de neurones. Si le réseau n'utilise qu'une entrée, l'indice i est supprimé.
 v_M : Tension de mode commun d'un circuit différentiel.
 v_{Mmini} : Tension de mode commun minimale admissible d'un circuit différentiel.
 v_{Mmaxi} : Tension de mode commun maximale admissible d'un circuit différentiel.
 V_{outBP} : Signal issu de la multiplication des signaux issus de l'oscillateur local et du montage transimpédance, dont le signal haute fréquence a été filtré.
 V_{out_i} : Tension de sortie du $i^{ème}$ détecteur de seuil.
 V_{outNN} : Tension de sortie du réseau de neurones.
 $V_{outNNid}$: Tension idéale souhaitée en sortie du réseau de neurones.
 $V_{outRANG}$: Signal de sortie du phasemètre, issu de la multiplication des signaux V_{outBP} et V_{ref} .
 V_{outS_i} : Amplitude du signal de sortie de l'amplificateur transimpédance correspondant au test de la surface S_i .
 $V_{outTRANS}$: Amplitude du signal de sortie de l'amplificateur transimpédance.
 V_{ref} : Signal issu de la multiplication des signaux issus des oscillateurs local et RF, dont le signal haute fréquence a été filtré.

V_{th} : Tension de seuil du transistor MOS.

W : Largeur du canal du transistor MOS.

W_0 : Ensemble des paramètres mis en jeu au sein du réseau de neurones.

w_{1ij} : Poids relatif au multiplieur Mult1ij situé entre l'entrée V_{inNN_i} et le $j^{ème}$ neurone caché. Si le réseau n'utilise qu'une entrée, l'indice i est supprimé. Si le neurone est étudié seul, l'indice j est supprimé.

w_{2j} : Poids relatif au multiplieur Mult2j situé entre le $j^{ème}$ neurone caché et le neurone de sortie. Si le neurone est étudié seul, l'indice j est supprimé.

x_i : $i^{ème}$ entrée du réseau de neurones, relative à la $i^{ème}$ variable sélectionnée dans le système.

$x_{i\alpha}$: $\alpha^{ème}$ valeur d'un ensemble présentée à la $i^{ème}$ entrée du réseau de neurones.

y_m : $m^{ème}$ sortie du réseau de neurones.

$z_m(x_1, x_2, \dots, x_n)$: $m^{ème}$ donnée à approcher du système, fonction des variables x_1, x_2, \dots, x_n .

Z_t : Fonction de transfert de l'amplificateur transimpédance.

$$\beta : \frac{\mu C_{ox}}{2} \left(\frac{W}{L} \right)$$

Λ : Intervalle maximum de mesure sans indécision.

δD : Résolution de la mesure de distance par le télémètre.

Δf : Excursion en fréquence du VCO.

Δf_i : Largeur de la bande de fréquence associé à un amplificateur.

$\Delta \varphi$: Différence de phase entre l'onde émise et l'onde reçue par le télémètre laser par déphasage.

$\delta \varphi$: Résolution de la mesure du déphasage.

ΔV_G : Différence entre V_{GS} et V_{th} .

λ : Facteur de modulation de la longueur du canal.

λ_0 : Longueur d'onde du faisceau laser.

μ : Mobilité des porteurs.

ρ : Coefficient de réflexion diffusant apparent de la cible non-coopérative.

ρ_{diff} : Coefficient de réflexion diffusant de la cible non-coopérative.

θ_0 : Angle d'incidence du faisceau entre la capteur et la cible.

θ_{0max} : Limite supérieure de l'intervalle associé à la mesure de l'angle d'incidence.

ρ_i : Coordonnées du centre de la $i^{ème}$ gaussienne.

σ : Ecart-type de la gaussienne.

τ_D : Temps de vol.

τ_i : Constante de temps localisée $\tau_i = R_i C_i$.

Liste des publications

REVUES INTERNATIONALES A COMITE DE LECTURE

Publications acceptées :

GATET L., TAP-BETEILLE H., LESCURE M.,
"Real Time Surface Discrimination using an Analog Neural Network implemented in a Phase-Shift Laser Rangefinder",
IEEE Sensors Journal,
Vol. 7, Issue 10, Oct 2007, Pages: 1381-1387.

GATET L., TAP-BETEILLE H., LESCURE M.,
"Surface Detection using Phase-Shift Laser Rangefinder and Neural Network",
Optical Engineering, Optical System Design,
Vol. 46, Issue 7, July 2007, Pages: 073002 1-11.

GATET L., TAP-BETEILLE H., LESCURE M.,
"Design of an analog MLP-type neural network to increase the distance measurement range of a phase-shift laser rangefinder",
WSEAS Transactions on Circuits and Systems,
Vol. 5, Issue 6, Juin 2006, Pages: 821-828.

Publications en 1^{ère} lecture à la date de la soutenance :

GATET L., TAP-BETEILLE H., LESCURE M.,
"Test of a CMOS Analog Neural Network for Real-Time Surface Detection with a Laser Rangefinder",
IMTC 2007 Sp. Issue of IEEE Tr. on Instrumentation and Measurement.

GATET L., TAP-BETEILLE H., LESCURE M.,
"Measurement Range Increase of a Phase-Shift Laser Rangefinder using a CMOS Analog Neural Network",
IEEE Transactions on Instrumentation and Measurement.

GATET L., TAP-BETEILLE H., LESCURE M., ROVIRAS D., MALLET A.,
"Functional tests of a 0.6µm CMOS MLP Analog Neural Network for fast on-board signal processing",
Analog Integrated Circuits and Signal Processing, Springer Science & Business Media.

COMMUNICATIONS AVEC ACTES : CONFERENCES INTERNATIONALES

GATET L., TAP-BETEILLE H., LESCURE M.,
"Analog Neural Network Design for Real-Time Surface Detection with a Laser Rangefinder",
2007 IEEE Instrumentation and Measurement Technology Conference (IMTC 2007),
Warsaw, Pologne, 1-3 Mai 2007, Pages: 1-6.

Présentation orale : L. Gatet

Cet article a reçu le prix "IEEE IMTC 2007 student travel award".

GATET L., TAP-BETEILLE H., LESCURE M., ROVIRAS D., MALLET A.,
"Design and Test of a CMOS MLP Analog Neural Network for Fast On-Board Signal Processing",

L. Gatet, H. Tap-Béteille, M. Lescure, D. Roviras, A. Mallet,
13th IEEE Int. Conference on Electronics, Circuits and Systems (ICECS 2006),
Nice, France, 11-13 Déc. 2006, Pages: 922-925.

Poster : L. Gatet

GATET L., TAP-BETEILLE H., LESCURE M.,
"Raising of indecision of a Phase-Shift Laser Rangefinder thanks to feedforward MLP Neural Network",

7th WSEAS International Conference on Neural Networks (NN'06),
Cavtat, Croatie, 12-14 Juin 2006, Pages: 67-72.

Présentation orale : L. Gatet

COMMUNICATIONS AVEC ACTES : CONFERENCES NATIONALES

GATET L.,

"Réalisation de réseaux de neurones analogiques appliqués à la télémétrie laser",
Journée de l'Ecole Doctorale GEET,
ENSAT - INP Toulouse - 8 Mars 2007.

Présentation orale : L. Gatet

TAP-BETEILLE H., GATET L., LESCURE M.,

"Conception et test d'un ASIC analogique pour la prédistorsion de signaux par réseau de neurones",

Rencontre de Technologie Spatiale - Télécommunications, Programme CNES de Recherche et Technologie des systèmes orbitaux,
CNES - 11 Mai 2006.

Présentation orale : H. Tap-Béteille

RAPPORTS DE CONTRAT

TAP-BETEILLE H., GATET L., LESCURE M.,

"Test et validation de l'ASIC réseau de neurones en technologie CMOS",
Rapport d'avancement d'un contrat CNES – LEN7 (du 20/09/04 au 30/06/05).

TAP-BETEILLE H., LESCURE M., GATET L., JOVE R., ROVIRAS D.,

"Conception et intégration d'un réseau de neurones analogique en technologie CMOS",
Rapport d'avancement d'un contrat CNES – LEN7 – IRIT (du 01/04/03 au 31/07/04).

Table des Matières

Résumé.....	A
Abstract.....	C
Glossaire.....	E
Liste des publications.....	I
Introduction générale.....	1
I. Chapitre I : Les réseaux de neurones.....	3
I.1. INTRODUCTION.....	3
<i>I.1.1. Le neurone.....</i>	<i>3</i>
<i>I.1.2. Les réseaux de neurones.....</i>	<i>5</i>
I.1.2.1. Les réseaux de neurones non bouclés.....	5
I.1.2.2. Les réseaux de neurones bouclés.....	6
<i>I.1.3. Apprentissage d'un réseau de neurones.....</i>	<i>7</i>
I.2. RESEAU DE NEURONES NON BOUCLE A APPRENTISSAGE SUPERVISE.....	8
<i>I.2.1. Propriété fondamentale des réseaux de neurones non bouclés à apprentissage supervisé.....</i>	<i>8</i>
<i>I.2.2. Modélisation statique.....</i>	<i>9</i>
I.3. CALIBRAGE D'UN RESEAU DE NEURONES.....	11
I.4. APPLICATIONS.....	15
I.5. CONCEPTION ET INTEGRATION DE RESEAUX DE NEURONES.....	18
<i>I.5.1. Avantages et inconvénients de l'intégration.....</i>	<i>18</i>
<i>I.5.2. Différents types d'intégration des réseaux de neurones.....</i>	<i>18</i>
I.5.2.1. Intégration numérique.....	18
I.5.2.2. Intégration analogique.....	20
I.5.2.3. Intégration électronique mixte.....	23
I.5.2.4. Intégration optoélectronique.....	23
I.6. CONCLUSION.....	24
I.7. REFERENCES.....	25

II. Chapitre II : Télémétrie laser.....	29
II.1. INTRODUCTION.....	29
II.2. MESURE DE DISTANCE PAR TEMPS DE VOL.....	32
II.2.1. <i>Technique impulsionnelle</i>	32
II.2.2. <i>Régime sinusoïdal permanent</i>	33
II.2.2.1. Télémétrie par modulation de fréquence (pseudo FMCW).....	34
II.2.2.2. Télémétrie laser par mesure de déphasage.....	35
II.3. CONCLUSIONS ET PROBLEMATIQUES.....	39
II.3.1. <i>Conclusion</i>	39
II.3.2. <i>Problématiques</i>	40
II.3.2.1. Amplitude du signal photoélectrique reçu.....	40
II.3.2.2. Mesure de distance.....	40
II.3.2.3. Classification de surfaces.....	41
II.3.2.4. Bilan.....	45
II.4. REFERENCES.....	46
III. Chapitre III : Utilisation de réseaux de neurones dans le domaine de la télémétrie laser par déphasage.....	51
III.1. INTRODUCTION.....	51
III.1.1. <i>Mesure de distance</i>	51
III.1.2. <i>Classification de surfaces</i>	53
III.2. MESURE DE DISTANCE.....	55
III.2.1. <i>Principe</i>	55
III.2.2. <i>Simulations comportementales sous Matlab</i>	58
III.3. CLASSIFICATION DE SURFACES.....	66
III.3.1. <i>Principe</i>	66
III.3.2. <i>Simulations comportementales sous Matlab</i>	71
III.4. CONCLUSION.....	79
III.5. REFERENCES.....	80

IV. Chapitre IV : Conception du réseau de neurones analogique appliqué à la télémétrie laser par déphasage.....	81
IV.1.INTRODUCTION.....	81
IV.2.CELLULES ANALOGIQUES EN TECHNOLOGIE CMOS BASSE TENSION....	82
<i>IV.2.1.Modèle niveau 1.....</i>	<i>82</i>
<i>IV.2.2.Schéma équivalent petit signal.....</i>	<i>83</i>
<i>IV.2.3.Transfert de courant par miroir.....</i>	<i>85</i>
IV.2.3.1.Introduction.....	85
IV.2.3.2.Miroir de courant simple.....	85
IV.2.3.3.Miroir de courant cascodé.....	87
IV.2.3.4.Discussion.....	89
<i>IV.2.4.Circuit différentiel à sources couplées.....</i>	<i>90</i>
IV.2.4.1.Introduction.....	90
IV.2.4.2.Transfert du circuit différentiel.....	91
IV.2.4.3.Domaine de linéarité du mode différentiel.....	93
IV.2.4.4.Domaine admissible de la tension de mode commun.....	95
IV.2.4.5.Circuit mono-tension.....	96
<i>IV.2.5.Multiplication par un poids référencé à la masse.....</i>	<i>97</i>
IV.2.5.1.Principe.....	97
IV.2.5.2.Application au multiplieur quatre quadrants.....	99
<i>IV.2.6.Génération d'un courant de décalage additionnel contrôlé par la tension de biais</i>	<i>100</i>
<i>IV.2.7.L'additionneur.....</i>	<i>102</i>
<i>IV.2.8.Tangente hyperbolique.....</i>	<i>102</i>
IV.3.CONCEPTION DE L'ASIC.....	106
IV.4.VALIDATION EXPERIMENTALE.....	108
<i>IV.4.1.Premier multiplieur.....</i>	<i>108</i>
<i>IV.4.2.Fonction tangente hyperbolique.....</i>	<i>110</i>
<i>IV.4.3.Multiplieur de sortie – Neurone élémentaire.....</i>	<i>112</i>
<i>IV.4.4.Réalisation et test du réseau de neurones complet.....</i>	<i>114</i>
IV.5.CONCLUSION.....	118
IV.6.REFERENCES.....	120

V. Chapitre V : Exploitation des résultats.....	121
V.1. LEVEE D'INDETERMINATION SUR LA MESURE DE DISTANCE.....	121
V.1.1. <i>Entraînement du réseau de neurones.....</i>	<i>121</i>
V.1.2. <i>Validation comportementale.....</i>	<i>125</i>
V.1.3. <i>Calibrage du réseau de neurones sur PCB – Tests.....</i>	<i>126</i>
V.2. CLASSIFICATION DE SURFACES.....	128
V.2.1. <i>Entraînement du réseau de neurones.....</i>	<i>128</i>
V.2.2. <i>Validation comportementale.....</i>	<i>134</i>
V.2.3. <i>Calibrage du réseau de neurones sur PCB – Tests.....</i>	<i>136</i>
V.3. COMPARAISON AVEC UN RESEAU DE NEURONES NUMERIQUE.....	138
V.3.1. <i>Conception numérique.....</i>	<i>138</i>
V.3.2. <i>Avantages liés à chaque méthode d'intégration.....</i>	<i>143</i>
V.4. PERFORMANCES – LIMITATIONS.....	145
V.5. REFERENCES.....	146
Conclusion générale.....	147
Annexe A : Modélisation statique.....	151
Annexe B : Calibrage d'un réseau de neurone.....	155
Annexe C : Exemple d'intégration d'un réseau de neurones analogique - numérique dans un système de télécommunication embarqué.....	161
a. Description du système.....	161
b. Description du réseau de neurones choisi.....	163
c. Description des fonctions implémentées et testées.....	165
Annexe D : Conception de l'ASIC.....	171
Références des annexes.....	174
Liste des figures et tableaux	i

Introduction générale

La mesure de distance, la classification et la détection de formes sont devenus prépondérants dans divers domaines d'applications tels que la robotique terrestre ou spatiale, l'industrie, la sécurité, les transports, le médical, etc. Pour répondre aux contraintes liées à l'environnement, chaque système de mesure doit répondre à un cahier des charges spécifique. Hormis le coût de production, les principaux critères de conception sont la plage de mesure, la résolution, la consommation, l'encombrement, la fiabilité et la reconfigurabilité. Les performances relatives à chacun de ces critères dépendent de l'application considérée. Néanmoins, quelles que soient les performances requises, un traitement approprié des signaux issus du système est nécessaire, aussi simple soit-il.

Les réseaux de neurones sont une alternative permettant de traiter différents signaux issus d'un système de mesure. En effet, ces architectures, composées de plusieurs modules interconnectés appelés neurones, peuvent être entraînés à partir d'un nombre limité de données expérimentales, de manière à approcher une fonction ou effectuer des opérations de classification sur un intervalle bien défini. De plus, leur caractère paramétrable les rendent aisément reconfigurables en fonction des conditions d'expérimentation. Enfin, ils offrent la possibilité d'être implantés directement au sein d'une application.

La première partie de ce manuscrit présente de manière succincte les diverses propriétés et aptitudes des réseaux de neurones au sein de différentes applications et à travers plusieurs types d'intégration. L'objectif de cette thèse est de montrer l'intérêt d'utiliser un réseau de neurones au sein d'un système optoélectronique de mesure de précision décrit dans le deuxième chapitre : la télémétrie laser par déphasage. Pour cela, deux études ont été menées en parallèle.

La première consiste à lever l'indétermination sur la mesure de distance déduite de la mesure du déphasage. Une démonstration comportementale de la capacité d'un réseau de neurones à accroître l'intervalle de mesure de distance sans modifier la résolution est développée dans le troisième chapitre.

La seconde étude proposée dans ce chapitre montre la possibilité d'effectuer une classification de différents types de surfaces sur des plages de distances et d'angles d'incidence variables à partir d'une structure neuronale quasi identique.

Ces deux études, ne faisant appel à aucune modification structurelle du télémètre déjà conçu, sont réalisées à partir de valeurs expérimentales afin de prouver le réel intérêt des deux méthodes développées.

Une même architecture neuronale de type perceptron multi-couches a été conçue de manière à valider expérimentalement la faculté du réseau à répondre aux deux applications proposées. Pour cela, un neurone élémentaire a été simulé et implanté dans un ASIC analogique en technologie CMOS 0.35 μ m. Les phases de simulation, de conception et de test du neurone analogique sont détaillées dans le quatrième chapitre. Un démonstrateur du réseau de neurones global, réalisé à partir de neurones élémentaires intégrés mis en parallèle, a été caractérisé expérimentalement et qualifié.

Le cinquième chapitre présente les phases d'entraînement, de validation et de test du réseau intégré pour les deux applications développées. Les phases de calibrage sont effectuées extérieurement à l'ASIC à partir de l'équation de transfert globale du réseau de neurones déterminée lors de la phase de caractérisation expérimentale.

Une version numérique du neurone intégré est ensuite présentée succinctement, puis comparée à la structure analogique afin de justifier les avantages associés à chaque type d'intégration.

Enfin, un bilan des performances obtenues est présenté en fin de chapitre, de manière à justifier les avantages d'une intégration neuronale au sein d'un télémètre laser par déphasage, n'ayant entraîné aucune modification de la structure télémétrique.

Chapitre I

Les Réseaux de Neurones

Ce chapitre présente des généralités sur les réseaux de neurones. Une partie des domaines d'applications y sont explicités, ainsi que certaines applications associées. Enfin, un état de l'art sur les techniques de conception et d'intégration des réseaux de neurones embarqués est détaillée en fin de partie.

I.1. INTRODUCTION

I.1.1. Le neurone

Par définition, un neurone est une fonction non linéaire, paramétrée, à valeurs bornées. On appelle entrées les variables sur lesquelles le neurone opère, et sortie le résultat de l'opération. La figure I-a est une représentation graphique d'un neurone formel ([Mc Culloch, 1943], [Minsky, 1969]). Elle s'inspire de la représentation d'un neurone biologique, dont le rôle est de transmettre le signal électrique issu des dendrites vers la terminaison axonique du neurone.

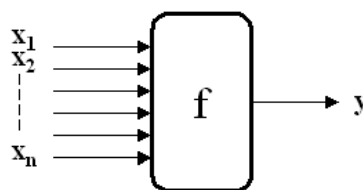


Figure I-a : Un neurone réalise une fonction non linéaire bornée $y = f(x_1, x_2, \dots, x_n)$ où les $\{x_i\}$ sont les entrées et y la sortie.

Parmi les structures de neurones, deux se dégagent en terme de fréquence d'utilisation :

- Dans le premier cas, la sortie y du neurone est une fonction non linéaire d'une combinaison des entrées $\{x_i\}$ pondérées par les paramètres $\{w_i\}$, désignés sous le nom de "poids" ou "poids synaptiques", toujours en relation avec la biologie. A la somme pondérée des entrées d'un neurone s'ajoute un terme constant b_0 appelé "biais". La fonction responsable de la non-linéarité est appelée fonction d'activation. Cette fonction est généralement une

fonction sigmoïdale, telle que la tangente hyperbolique ou la fonction arctangente (figure I-b). Ce type de neurones est l'élément de base des réseaux de neurones les plus utilisés à l'heure actuelle, les Perceptrons Multi-Couches (PMC, ou MLP pour *Multi-Layer Perceptrons*).

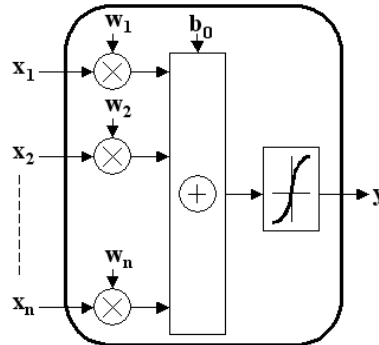


Figure I-b : Neurone de base du perceptron multi-couches.

La sortie du neurone représenté sur la figure I-b a pour équation :

$$y = f(x_1, x_2, \dots, x_n) = \text{th}(b_0 + \sum_{i=1}^n w_i x_i) \quad (\text{I-1})$$

- Dans le deuxième cas, les paramètres interviennent directement au sein de la non-linéarité de la fonction f . Cette fonction non-linéaire peut être une fonction radiale (RBF en anglais pour Radial Basis Function) ([Powell, 1987]), ou encore une ondelette ([Mallat, 1989]). Par exemple, la sortie d'un neurone RBF à non-linéarité gaussienne (figure I-c) a pour équation :

$$y = \exp \frac{(\sum_{k=1}^n x_k - \rho_k)^2}{2\sigma^2} \quad (\text{I-2})$$

où les paramètres $\{\rho_k\}$ sont les coordonnées du centre de la gaussienne et σ son écart-type.

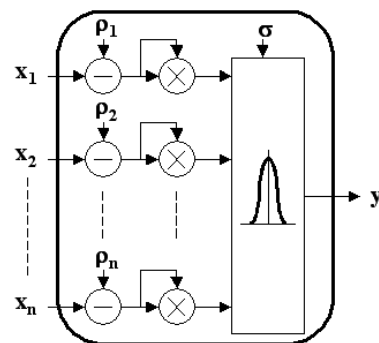


Figure I-c : Neurone RBF à non-linéarité gaussienne.

Le choix d'un de ces deux types de neurones dépend du type d'approximation que l'on souhaite faire. Les RBF ou les ondelettes par exemple ont des non-linéarités locales, qui tendent vers zéro dans toutes les directions de l'espace des entrées sauf pour des valeurs

proches des centres des gaussiennes. Ces types de neurones seront donc utilisés pour faire de la discrimination discrète ou de l'approximation dans des zones localisées de l'espace. En revanche, les neurones à fonction d'activation sigmoïde seront utilisés pour effectuer des approximations globales et continues sur l'intégralité de l'espace des entrées.

Ainsi, un neurone réalise simplement une fonction non linéaire de ses variables d'entrée à l'aide de paramètres externes. L'intérêt des neurones réside dans les propriétés qui résultent de leur association en réseaux.

I.1.2. Les réseaux de neurones

Un réseau de neurones est constitué de plusieurs neurones organisés d'une certaine manière de façon à ce que les sorties $\{y_m\}$ réalisent l'approximation de la fonction souhaitée en fonction des entrées $\{x_i\}$. En d'autres termes, un réseau de neurones est une composition de fonctions non linéaires réalisées par chacun des neurones qui le constitue. On distingue deux types de réseaux de neurones : les réseaux bouclés et non bouclés.

I.1.2.1. Les réseaux de neurones non bouclés

Un réseau de neurones non bouclé (*feedforward neural network*) peut se représenter graphiquement par un ensemble de neurones connectés de manière à ce que l'information ne transite que des entrées vers les sorties, sans "retour en arrière" ([Dreyfus, 2004]). Les neurones qui effectuent les dernières opérations sont appelés neurones de sortie. Les autres neurones, placés entre les entrées et les neurones de sortie, sont en nombre variable. Ils peuvent s'apparenter à la partie "processeur", effectuant des calculs intermédiaires en parallèle. En conséquence, leur nombre est relatif à la puissance de calcul, ou encore à la capacité à approximer plus ou moins bien la fonction désirée. Ces neurones sont appelés neurones cachés (figure I-d).

Ces neurones sont très souvent organisés par couches. Les neurones de sortie appartiennent à la couche de sortie, et les neurones cachés s'organisent en plusieurs couches appelées couches cachées (figure I-d). Dans un grand nombre d'applications, le nombre de couches cachées se limite à une, voire deux, pour des questions de convergence de l'algorithme relatif à la mise à jour des poids et des biais que nous verrons plus tard.

Le réseau de neurones représenté sur la figure I-d contient N_m sorties sur sa couche de sortie. Il réalise donc N_m fonctions algébriques des variables d'entrées $\{x_i\}$ du réseau.

Chacune de ces fonctions est la composée de la fonction relative au neurone de sortie correspondant et des fonctions non linéaires réalisées par les neurones cachés. Ainsi, un réseau de neurones non bouclé à n entrées, N_j neurones cachés et N_m neurones de sortie réalise N_m fonctions non linéaires de ses n variables d'entrée par composition des N_j fonctions réalisées par ses neurones cachés ([Dreyfus, 2004]).

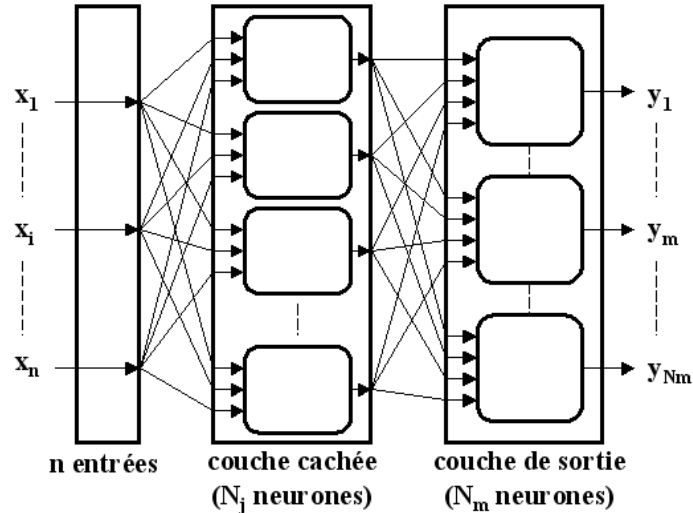


Figure I-d : Représentation d'un réseau de neurones à n entrées, une couche de N_j neurones cachés et N_m neurones sur la couche de sortie.

Le temps ne joue aucun rôle fonctionnel dans un réseau de neurones non bouclé. Si les entrées sont constantes, les sorties le sont également. Le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone est considéré comme négligeable. Pour cette raison, les réseaux non bouclés sont souvent appelés "réseaux statiques", par opposition aux réseaux bouclés ou "dynamiques" ([Dreyfus, 2004]).

I.1.2.2. Les réseaux de neurones bouclés

Il existe plusieurs types de réseaux de neurones bouclés. Le point commun entre tous les réseaux de neurones bouclés est la présence d'au moins un chemin qui part et revient au même neurone. Tout réseau de neurones bouclé peut être mis sous la forme d'un réseau de neurones non bouclé avec une contre-réaction de la sortie d'un de ses neurones vers un neurone d'une couche antérieure. Ainsi, quel que soit le type de réseau de neurones réalisé, l'étude du réseau de neurones non bouclé est fondamentale. Avant d'en indiquer les propriétés et les applications, nous allons définir la notion d'apprentissage.

I.1.3. Apprentissage d'un réseau de neurones

L'apprentissage d'un réseau de neurones peut être défini comme sa phase de calibrage, durant laquelle les divers paramètres le caractérisant sont remis à jour jusqu'à ce qu'ils permettent au réseau d'approximer au mieux la fonction qu'il a à réaliser. Selon l'application dans laquelle le réseau va être intégré, la fonction à approcher peut être connue ou inconnue analytiquement. En effet, le phénomène physique, chimique, économique, biologique, etc. qui nécessite l'emploi d'un réseau de neurones peut être parfaitement modélisé par l'intermédiaire d'équations analytiques ou simplement observables, mettant en jeu un nombre fini de valeurs expérimentales de sorties fonction de valeurs d'entrées bien choisies, sans que les relations qui les relient ne soient connues.

On appelle apprentissage supervisé cette forme de calibrage qui consiste à minimiser l'erreur entre un jeu de valeurs expérimentales et le jeu de valeurs de sortie du réseau de neurones associé, calculé algébriquement en fonction de ses paramètres. Les réseaux de neurones à apprentissage supervisé sont ainsi utilisés pour effectuer des approximations de fonctions connues analytiquement (dont les sorties sont parfaitement connues en fonction des entrées), ou de la modélisation statistique ou encore de la régression de fonctions dont la forme analytique est inconnue expérimentalement.

Il existe aussi des réseaux à apprentissage non supervisé, pour lesquels on dispose d'un ensemble de données, représentées par des vecteurs de grande dimension, que l'on cherche à regrouper selon des critères de ressemblance qui sont inconnus a priori. Les réseaux à apprentissage non supervisé les plus étudiés et utilisés, notamment dans le domaine de la visualisation ou de l'analyse de données, sont les "cartes auto-organisatrices" ou "cartes de Kohonen" ([Kohonen, 1984], [Dreyfus, 2004]).

I.2. RESEAU DE NEURONES NON BOUCLE A APPRENTISSAGE SUPERVISE

Les réseaux de neurones à apprentissage supervisé ont pour propriété d'être des approximateurs universels. Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire ([Hornik, 1989], [Hornik, 1990], [Hornik, 1991]). Cette partie décrit dans quel domaine les réseaux de neurones sont plus aptes à être utilisés que d'autres types d'approximateurs universels.

1.2.1. Propriété fondamentale des réseaux de neurones non bouclés à apprentissage supervisé

Un des critères de choix d'un approximateur parmi d'autres est son aptitude à approcher la fonction désirée avec la structure la plus simple et un nombre de paramètres mis en jeu le plus faible possible. L'un des principaux atouts d'un réseau de neurones est son caractère d'approximateur parcimonieux. En effet, un réseau de neurones, entraîné pour un nombre fini de couples {entrées ; sorties}, a la capacité d'extrapoler les valeurs intermédiaires, et ainsi réaliser une fonction continue à partir de données discrètes.

Dans le cas d'approximateurs linéaires par rapport à leur paramètres, comme les polynômes ou les réseaux de neurones RBF à centres et écarts-types fixés ou encore d'ondelettes à centres et dilatations fixés, le nombre de paramètres ajustables pour une précision donnée croît exponentiellement avec le nombre de variables. En revanche, dans le cas d'approximateurs non-linéaires par rapport à leurs paramètres, comme les réseaux de neurones à fonction d'activation sigmoïde, le nombre de paramètres ajustables croît linéairement avec le nombre de variables ([Barron, 1993]) Ainsi, plus le nombre de variables d'entrée sera important, plus le nombre de paramètres ajustables sera grand. Ces derniers seront d'autant plus nombreux si l'approximateur choisi est linéaire par rapport à ses paramètres.

Pour un modèle à une ou deux entrées, on peut utiliser indifféremment un modèle linéaire par rapport à ses paramètres (polynôme par exemple) ou un modèle non linéaire par rapport à ses paramètres (réseau de neurones par exemple). Dans ce cas, des arguments pratiques peuvent justifier une préférence pour un type d'approximateur plutôt qu'un autre, tels que des connaissances a priori sur le type de non-linéarité souhaitable, le caractère localisé ou non de la fonction, la rapidité de calcul, la facilité d'initialisation de l'apprentissage, la facilité de

réalisation en circuit spécialisé, etc. En revanche, au delà de deux entrées, les approximateurs non-linéaires tels que les réseaux de neurones à fonction d'activation sigmoïde sont à privilégier de manière à limiter le nombre de paramètres à introduire ([Dreyfus, 2004]).

1.2.2. Modélisation statique

Dans la pratique, il est rare que l'on cherche à réaliser une approximation uniforme d'une fonction connue. Le plus souvent, le problème concret qui se pose est le suivant : on dispose d'un ensemble de N vecteurs de N_m données $\{z_m(x_1, x_2, \dots, x_n), m \in [1 ; N_m]\}$. Chaque donnée z_m dépend de n variables x_i du système $\{x_i, i \in [1 ; n]\}$ (figure A-b, annexe B). Ces données sont rattachées à un processus de nature quelconque (physique, chimique, économique, financier...). On suppose qu'il existe une relation entre les entrées et les sorties du système, et on cherche à déterminer une forme mathématique de cette relation, valable dans le domaine où les données ont été acquises, sachant que ces données sont en nombre fini et qu'elles sont certainement entachées de bruit. Ces données sont appelés "exemples".

Ainsi, l'objectif est de modéliser le phénomène étudié à l'aide des paramètres ajustables du système, pour une précision donnée, à partir des exemples disponibles. Ce type de modélisation est appelé "modélisation boîte noire". Il s'oppose au "modèle de connaissance" ou "modèle de comportement interne", qui désignent un modèle mathématique établi à partir d'une analyse physique (ou chimique, physico-chimique, économique, etc.) du processus que l'on étudie, dont on tire un nombre limité de paramètres physiques ajustables, capables de régir l'évolution du système à eux seuls ([Dreyfus, 2004]).

Les réseaux de neurones sont utilisés pour l'élaboration de modèles "boîtes noires", mais aussi de modèles "boîtes grises", intermédiaires entre les modèles "boîtes noires" et les "modèles de connaissance". L'annexe A démontre cette aptitude des réseaux de neurones à la modélisation pour un cas simple, puis dans le cas général et enfin au travers d'un exemple.

Au delà du fait qu'un réseau de neurones approche de manière relativement satisfaisante une fonction pour diverses valeurs expérimentales, on note que le réseau de neurones permet aussi de donner une allure du phénomène expérimenté pour des points intermédiaires n'ayant pas été mesurés. Cette faculté et propriété des réseaux de neurones est désignée sous le terme de "généralisation".

L'estimation de la capacité de généralisation d'un réseau de neurones est au cœur de la problématique de l'apprentissage. En effet, cette généralisation est nécessairement limitée,

puisqu'elle ne peut s'étendre au delà de l'intervalle de mesures pour lequel le réseau a été entraîné.

Les réseaux de neurones non bouclés à apprentissage supervisé restent néanmoins de très bons approximateurs sur un espace d'entrées bien défini, notamment pour toutes les applications ou processus où il faut modéliser une tendance alors que les conditions de mesures sont difficiles, ou que les temps d'acquisition sont importants, ou encore lorsque les données et paramètres accessibles du système sont en nombre limité.

Les réseaux de neurones à apprentissage supervisé ayant la propriété d'approcher, de manière parcimonieuse, toute fonction de régression non linéaire suffisamment régulière, il peut donc être avantageux de les utiliser pour toute application nécessitant de trouver, par apprentissage, une relation non linéaire entre des données. Pour cela, il faut nécessairement remplir les conditions suivantes ([Dreyfus, 2004]) :

- Disposer d'échantillons de taille suffisamment grande, et bien représentatifs,
- S'assurer de l'intérêt réel d'un modèle non linéaire pour l'application considérée,
- S'assurer de l'utilité d'un réseau de neurones plutôt qu'un autre approximateur de fonctions non linéaire, les polynômes par exemple. Rappelons que, pour une précision donnée, le nombre de paramètres de la première couche de connexions varie linéairement avec le nombre de variables, alors qu'il varie exponentiellement dans le cas, par exemple, d'une approximation polynomiale,
- S'enquérir de la faisabilité du réseau de neurones choisi (d'un point de vue conception) s'il est amené à être implémenté à bord du système où il doit approcher la fonction désirée.

Il peut être nécessaire de mettre en œuvre un pré-traitement des données si celles-ci ne sont pas exploitables directement par le réseau de neurones. Dans le cas d'un pré-traitement des données, il est nécessaire de réfléchir sur les complexités de structure et sur les temps de calculs relatifs au système de pré-traitement et au réseau de neurones.

Si toutes ces conditions sont validées et que l'utilité d'une structure neuronale est justifiée, deux questions centrales se posent :

- comment, dans une famille de réseaux de neurones donnée, trouver celui qui minimise le mieux la fonction de coût des au sens des moindres carrés ?
- une fois que celui-ci a été trouvé, comment juger s'il est "satisfaisant"?

I.3. CALIBRAGE D'UN RESEAU DE NEURONES

Pour valider l'utilisation d'un réseau de neurones au sein d'une application, il est nécessaire de :

- déterminer le type de réseau en fonction de l'application,
- définir sa structure,
- calculer ses paramètres à partir d'entrées pertinentes, avec pour critère la minimisation de la fonction de coût des au sens des moindres carrés,
- valider la structure choisie et le jeu de paramètres associés à l'aide d'un autre jeu de valeurs d'entrées, de manière à constater la faculté de généralisation du réseau pour l'application à laquelle il est destiné.

Le premier choix à effectuer se porte donc sur le type de réseau. Une connaissance a priori du système (modèle "boîte grise") permet de se tourner vers le choix d'une approximation sur un ensemble continu de valeurs ou de manière discrète sur une somme d'intervalles disjoints appartenant à l'espace des entrées. Dans le cadre d'une approximation de fonction continue sur un intervalle, le choix se tournera plus vers un réseau de neurones à fonction d'activation sigmoïdale; en revanche, dans le cadre d'une approximation de fonction sur une somme d'intervalles disjoints, les RBF ou les ondelettes, présentant des non-linéarités locales, seront préférés (paragraphe I.1.1.). D'un point de vue conception, les réseaux de neurones à fonction d'activation sigmoïdale sont souvent les plus simples à mettre en œuvre (paragraphe I.5.2.).

Une fois le type de réseau de neurones choisi, il faut déterminer sa structure, soit , en d'autres termes, le nombre d'entrées, le nombre de neurones de sortie, le nombre de couches intermédiaires dites "cachées" et enfin le nombre de neurones par couche cachée.

Pour déterminer le nombre d'entrées d'un réseau de neurones, on commence par réunir l'ensemble des données accessibles du système susceptibles d'influencer les sorties que l'on souhaite modéliser. Ensuite, l'objectif est de limiter le nombre d'entrées de manière à limiter la complexité du réseau, notamment au niveau du nombre de connexions entre les entrées et les neurones de la première couche cachée. On détermine donc parmi l'ensemble des paramètres ceux qui sont les plus pertinents, en supprimant ceux qui ont une influence négligeable ou ceux qui sont fonctions d'autres paramètres mesurables. Il est à noter que le nombre d'entrées du réseau de neurones peut être encore diminué en combinant certains paramètres pertinents entre eux en amont du réseau durant la phase de pré-traitement. Une connaissance a priori du

système aide grandement à cette opération. Il est important de pouvoir avoir accès à un certain nombre de valeurs associées à chacune des entrées, de manière à pouvoir générer une séquence d'apprentissage suffisamment grande et faire en sorte de conserver d'autres valeurs de manière à valider le bon fonctionnement du réseau une fois la phase d'apprentissage terminée. Enfin, si le réseau de neurones est amené à être conçu afin d'être intégré au sein de l'application, le caractère accessible et mesurable des paramètres d'entrée revêt une importance primordiale dans l'objectif même de la réalisation de celui-ci.

Le nombre de neurones de sortie est assez simple à déterminer. En effet, il dépend du nombre de fonctions de régression non-linéaire que l'on souhaite réaliser au sein du système.

La détermination du nombre de couches cachées et du nombre de neurones par couches sont intimement liés à la capacité d'approximation de la structure et de ses paramètres associés. En effet, ce n'est qu'une fois le résultat de l'approximation jugé satisfaisant que la structure du réseau et ses paramètres sont figés. L'annexe B détaille une méthode de détermination du nombre de couches cachées, du nombre de neurones sur chaque couche et des paramètres du réseau de neurones, ainsi qu'un critère de satisfaction associé à la finesse d'approximation de la structure neuronale testée.

La structure finale du réseau de neurones va donc dépendre en particulier de l'arrangement et du nombre de neurones sur sa ou ses couches cachées, et des paramètres associés. Ainsi, structure et calibrage sont deux phases corrélées et indissociables. Pour justifier cette corrélation, prenons l'exemple de l'apprentissage d'un réseau de neurones de type perceptron multi-couches à une entrée, un neurone de sortie et une couche cachée contenant N_j neurones (figure I-e). Seul le $j^{\text{ième}}$ des N_j neurones de la couche cachée a été représenté. Chaque neurone de la couche cachée, constitué d'un additionneur de biais b_{1j} et d'une fonction sigmoïdale, reçoit une valeur d'entrée x_1 du réseau pondérée par un poids w_{1j} et transmet au neurone de sortie une valeur pondérée par un autre poids w_{2j} . Les sorties du premier multiplieur, de l'additionneur, de la fonction sigmoïdale et du deuxième multiplieur appartenant au $j^{\text{ième}}$ neurone de la couche cachée sont respectivement appelés a_{1j} , a_{2j} , a_{3j} et a_{4j} . L'ensemble des sorties pondérées des neurones de la couche cachée sont sommées au niveau du neurone de la couche de sortie, constituée d'un additionneur de biais b_2 . L'ensemble des paramètres W_0 du réseau sont donc $\{w_{1j}, b_{1j}, w_{2j}, b_2, j \in [1 ; N_i]\}$.

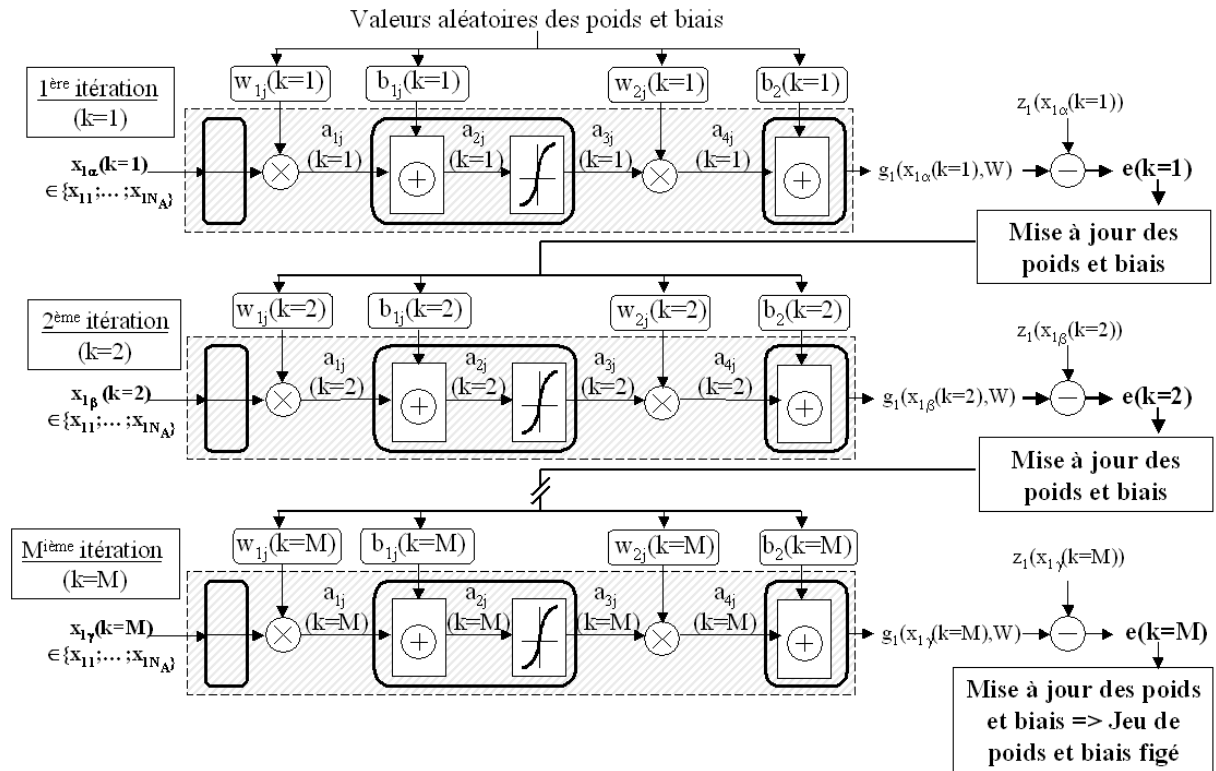


Figure I-e : Méthode itérative de rétropropagation de l'erreur appliquée à un perceptron multi-couches de manière à figer ses poids et biais.

On cherche à réaliser la fonction de régression d'un phénomène mettant en jeu N données à l'aide de ce réseau (représenté en hachuré sur la figure I-e). On choisit, parmi ces N données collectées, N_A données $z_1(x_{1\alpha})$, relatives à N_A valeurs du paramètre $x_{1\alpha}$ (avec $\alpha \in [1 ; N_A]$) qui vont constituer l'ensemble d'apprentissage. Les différents paramètres W_0 de départ sont choisis aléatoirement avec une distribution uniformément répartie autour de 0.

- A la première itération ($k=1$), on choisit aléatoirement une valeur $x_{1\alpha}$ parmi les N_A valeurs d'entrée. Le réseau de neurones effectue les différentes opérations intrinsèques à sa structure (multiplication de l'entrée par N_j poids w_{1j} , puis addition de N_j biais b_{1j} , puis application de la non-linéarité relative à la fonction sigmoïdale, puis multiplication par N_j poids w_{2j} et enfin addition de l'ensemble des N_j sorties des multiplieurs à un biais b_2 au niveau du neurone de sortie) pour donner en sortie la valeur $g_1(x_{1\alpha}, W_0)$ associée. Cette valeur est comparée à la donnée $z_1(x_{1\alpha})$ relative à la même entrée $x_{1\alpha}$. De la différence entre ces deux valeurs résulte une erreur e qui permet de calculer le gradient de chaque paramètre en fonction de cette erreur et ainsi de mettre à jour l'ensemble des paramètres par la méthode usuelle de "rétropropagation de l'erreur" ([Rumelhart, 1986], [Werbos, 1974]). Selon l'application, l'algorithme associé à la méthode de rétropropagation ne fait appel qu'au gradient simple ([Haykin, 1994]), ou au gradient de second ordre, comme la méthode de

Broyden-Flecher-Goldfarb-Shanno (algorithme BFGS [Broyden, 1970]) ou l'algorithme de Levenberg-Marquardt ([Levenberg, 1944], [Marquardt, 1963]).

- Une fois la mise à jour effectuée, la deuxième itération commence ($k=2$). Une nouvelle valeur aléatoire $x_{1\beta}$ est choisie parmi les N_A valeurs du jeu d'apprentissage. La nouvelle valeur de sortie $g_1(x_{1\beta}, W_0)$ du réseau relative à cette nouvelle entrée $x_{1\beta}$ et au nouveau jeu de poids et biais est comparée à la donnée $z_1(x_{1\beta})$ correspondante. La différence entre les deux valeurs conduit à une nouvelle erreur qui va permettre une nouvelle mise à jour des poids et biais.

- Les itérations se succèdent jusqu'à ce que le jeu de poids et biais permette de trouver une erreur inférieure à une valeur donnée, et ce pour n'importe quelle valeur de l'ensemble d'apprentissage. A ce moment là, le jeu de paramètres W_0 correspondant à l'ensemble des poids et biais est figé de manière à tester le réseau de neurones avec d'autres valeurs d'entrée, appartenant au jeu de validation.

- Si au bout d'un nombre important d'itérations l'erreur ne descend pas en dessous de l'erreur fixée par l'utilisateur pour toute valeur de l'ensemble d'apprentissage, celui-ci juge que la qualité d'approximation n'est pas satisfaisante. Un autre neurone est alors ajouté sur la couche cachée et la procédure de calcul des paramètres W_0 du réseau est relancée.

I.4. APPLICATIONS

Les applications des réseaux de neurones sont très nombreuses. On peut ainsi les trouver pour des applications de reconnaissance de forme (lecture automatique de codes postaux ([Le Cun, 1991], [Mozayyani, 1998], [Knerr, 1992]), reconnaissance de visages ([Yang, 2003], [Hui, 2004]), détection de formes ([Bishop, 1995], [De Almeida, 2000]), etc.), de contrôle non destructif (détection de défauts dans les rails par courant de Foucault ([Oukhellou, 1997])), de prévision et de prédiction ([Santos, 2004], [Perrone, 1999], [Kim, 1991]), de fouille de données (filtrage de documents ([Stricker, 2001])), de robotique (contrôle et déplacement ([Yang, 1999], [Novakovic, 1994]), pilotage automatique d'un véhicule autonome ([Rivals, 1994], [Rivals, 1995]), etc.), etc. Deux grandes familles d'applications se distinguent néanmoins : l'approximation de fonctions de régression et la classification.

La modélisation de fonctions non linéaires intervient dans différents domaines tels que la recherche d'une loi générale à partir de certains paramètres d'un système afin d'en prévoir l'évolution (dans les domaines météorologiques et démographiques par exemple), le traitement des données disponibles de manière à simplifier la prise de décision ([Roussel, 2001]) ou encore la correction de transferts non linéaires ([Castanie, 2002]). Une application relative à la linéarisation de transfert non linéaires est détaillée dans l'annexe C.

Les premières applications industrielles des réseaux de neurones à apprentissage supervisé étaient relatives à la reconnaissance de formes ou de signaux. Ceci est dû au fait que tout problème de classification peut se ramener à un problème très voisin de celui de la régression non linéaire. En effet, classer un ensemble d'objets revient, en d'autres termes, à attribuer à chacun une classe (ou "catégorie") parmi plusieurs classes définies à l'avance. Un algorithme qui réalise automatiquement une classification est appelé classificateur ([Dreyfus, 2004]). La classification de données d'un système revient donc à discriminer les objets d'une classe par rapport aux autres classes par l'intermédiaire de mesures accessibles du système, ou encore à approximer différentes fonctions discriminatrices de telle sorte que les ensembles de sorties associés à un jeu de valeurs d'entrées d'une classe donnée soient disjoints les uns des autres (figure I-f).

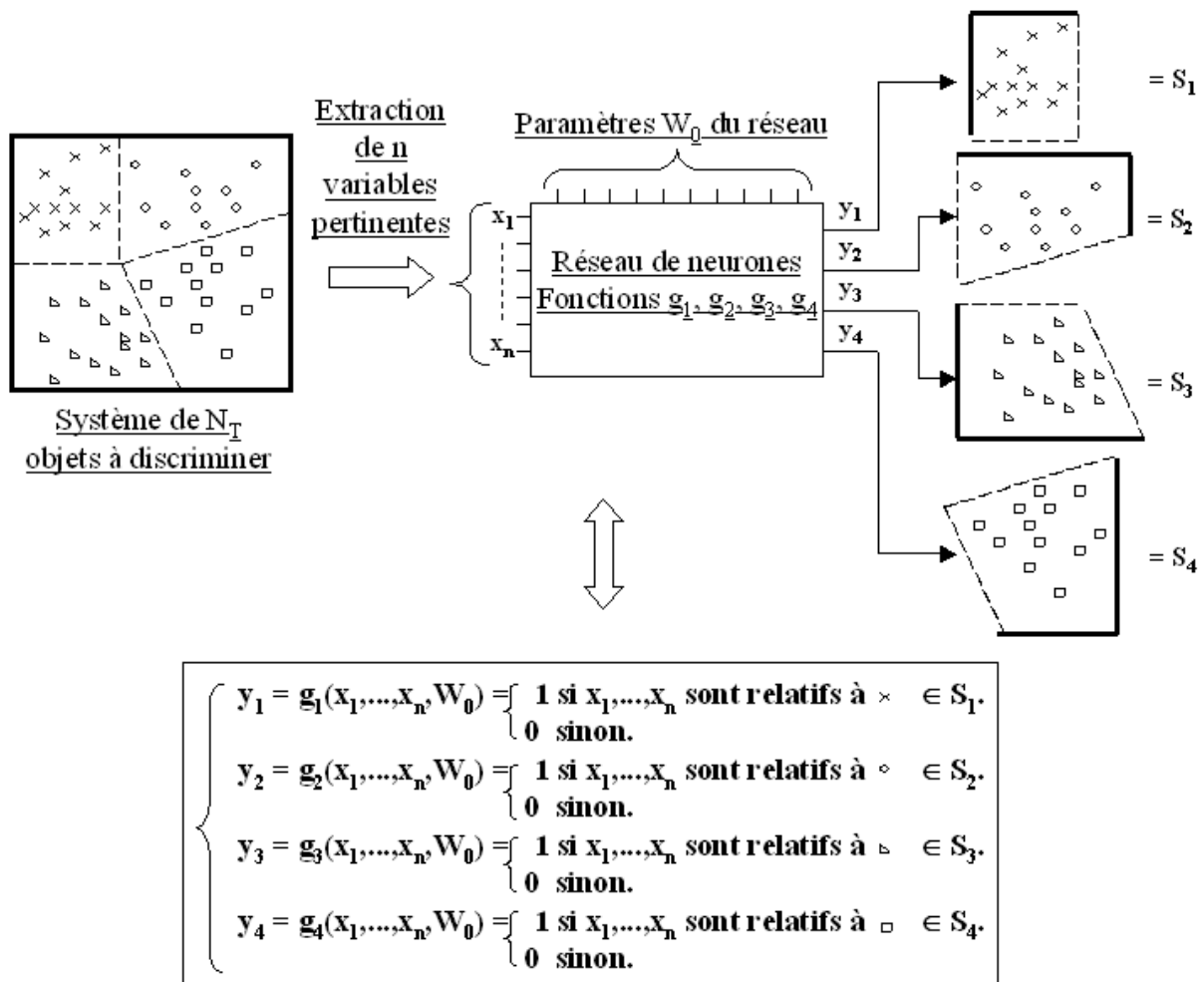


Figure I-f : Mise en œuvre de la capacité de classification du réseau de neurones par sa faculté d'approximation de fonctions non-linéaires.

Ainsi, la notion de classification peut se ramener à un problème d'approximation de N_m fonctions de régression non-linéaires g_m (avec $N_m=4$ dans le cas de la figure I-f) . Pour cette raison, les réseaux de neurones sont fréquemment utilisés comme classificateurs ou encore discriminateurs. Par la suite, nous utiliserons indifféremment le terme de "classification" ou celui de "discrimination" ([Dreyfus, 2004]).

Lorsque les grandeurs relatives au système sont plutôt bien maîtrisées et en nombre conséquent, et lorsque les frontières entre chaque classe de données collectées ne sont pas trop complexes, on utilise les réseaux de neurones pour effectuer des opérations de tri et de reconnaissance de formes, de surfaces, de caractères ou de symboles. Pour mener à bien leur tâche, il est souvent nécessaire d'effectuer des opérations sur les signaux ou les données issues du système. Par exemple, dans le cas de la lecture des codes postaux ou des montants numériques des chèques, l'opération élémentaire effectuée de manière très générale est la pixelisation des chiffres ([Le Cun, 1991], [Mozayyani, 1998], [Knerr, 1992]).

Cependant, le choix des opérations à effectuer de manière à faciliter la tâche du réseau de neurones n'est pas simple. En effet, si on souhaite simplifier au maximum le travail de discrimination du réseau, la partie relative au traitement du signal risque d'être très lourde en temps de calcul. La partie algorithmique relative au traitement des données peut donc limiter la vitesse du système, ou, en d'autres termes, sa bande passante.

Si au contraire, les opérations effectuées sur les signaux sont trop légères, la structure du réseau de neurones s'en trouve plus complexe, avec les difficultés de calibrage associées (paragraphe I.3 et annexe B). Il est donc nécessaire d'optimiser au maximum le traitement du signal en accord avec la structure du réseau de neurones.

De la même manière, il peut être nécessaire d'effectuer un post-traitement des données issues du réseau. Bien que généralement plus simple, il est tout de même important de faire en sorte que ce post-traitement ne limite pas la vitesse du système. Les pré-traitements et post-traitements doivent être faits en accord avec la structure adéquate du réseau de neurones choisi pour l'application souhaitée (exemples avec des réseaux de neurones RBF utilisés pour des applications de reconnaissance de face : [Yang, 2003], [Hui, 2004] ; exemples avec des réseaux de neurones PMC utilisés pour des applications de reconnaissances de caractères ([Mozayyani, 1998], [Knerr, 1992]) et de détection de formes ([De Almeida, 2000])).

Dans certains types d'applications, il peut être très intéressant d'effectuer le traitement de données directement au sein du système où il doit intervenir. Dans ce cas, il est important de s'intéresser aux différentes méthodes de conception de réseaux de neurones.

I.5. CONCEPTION ET INTEGRATION DE RESEAUX DE NEURONES

I.5.1. Avantages et inconvénients de l'intégration

Au delà de la possibilité d'apporter une aide au traitement des données issues d'un système par l'intermédiaire d'un logiciel (implantation "software"), il est possible d'intégrer un réseau de neurones "matériellement" (méthode "hardware"), afin de traiter les données directement au sein du système. Il faut cependant que l'application s'y prête, que les données pertinentes soient accessibles, en nombre suffisant et sous une forme exploitable. De plus, comme le réseau de neurones est embarqué, le système de pré-traitement l'est aussi, voire le système de post-traitement, de manière à ce que les données traitées soient exploitables ou tout simplement accessibles. Cela impose des contraintes de place, de consommation, de vitesse de traitement (et donc de limitation de la bande passante du système complet), de robustesse, etc. qui peuvent nuire au respect du cahier des charges de l'application considérée.

Ainsi, pour une application spatiale telle que la linéarisation d'un transfert non linéaire quelles que soient les conditions de température et d'éclairage, ou encore la détection d'objets par un robot autonome (détermination du type d'obstacle, évaluation de la distance, etc.), un réseau de neurones implanté peut avoir une très grande utilité. En revanche, compte tenu des contraintes énoncées précédemment, un soin tout particulier est à apporter à sa réalisation. Il en est de même pour tout autre système embarqué (moyen de transport, robot, etc.) sur lequel il est difficile d'intervenir en fonction de la façon dont le système a été intégré.

Ainsi, si l'application s'y prête, il existe plusieurs moyens d'intégrer "matériellement" un réseau de neurones.

I.5.2. Différents types d'intégration des réseaux de neurones

Le type d'intégration matérielle le plus commun est l'intégration électronique. Elle consiste à réaliser les différentes fonctions à l'aide de composants électroniques, de manière analogique et/ou numérique. Chacune des deux méthodes d'intégration est étudiée en détail dans les deux parties suivantes.

I.5.2.1. Intégration numérique

A l'heure actuelle, la majeure partie des réseaux de neurones intégrés existants sont numériques. Une implantation numérique a l'avantage d'être très bien maîtrisée, simple à

réaliser et parfaitement reproductible. De plus, l'intégration numérique est reconfigurable, ce qui peut être un avantage précieux lors d'une première intégration du réseau de neurones au sein d'un système. Enfin, les systèmes de mise à jour des paramètres du réseau en fonction de l'évolution de sa sortie sont en grande majorité numériques. Les signaux transitant entre le réseau de neurones et le système de mise à jour ne nécessitent donc aucune conversion (cela ne signifie pas pour autant qu'aucun traitement des signaux n'est nécessaire entre les deux structures numériques).

En revanche, ils ne réalisent pas tout à fait des opérations "en parallèle". Ces opérations sont liées à des calculateurs numériques dont une bonne partie des calculs sont en série. La puissance consommée associée est fonction de la fréquence d'horloge choisie. De plus, la simple réalisation d'une cellule de base demande un grand nombre de transistors, surtout si l'on souhaite par exemple une grande précision de traitement sur l'ensemble de la caractéristique à linéariser. Enfin, dans le cas de systèmes nécessitant le traitement de signaux analogiques, le système de pré-traitement nécessite l'intégration de convertisseurs analogiques-numériques. Il s'en suit une consommation plus importante ainsi qu'une vitesse de traitement limitée.

Les références ([Dias, 2004], [Vitabile, 2005], [Da Zhang, 2006], [Torbey, 1992], [Tang, 1997], [Wojtek Przytula, 1991], [Yang, 2003]) présentent différents types d'implantations numériques, certains étant spécifiques à une application donnée. Il existe quelques structures de réseaux RBF numériques intégrés, en particulier pour des applications où il est nécessaire d'effectuer des opérations de classification discrète directement au sein de l'application ([Yang, 2003]). La majeure partie des réseaux de neurones numériques sont néanmoins de type perceptron multi-couche.

Si les cellules de base du perceptron sont le multiplieur, l'additionneur et la fonction d'activation, le multiplieur et l'additionneur sont rarement détaillés dans la littérature, car ils sont souvent synthétisés à partir de bibliothèques déjà existantes.

La fonction d'activation est souvent le point central de l'intégration numérique. L'objectif est de concilier le caractère continu de l'approximation d'un réseau de type PMC tout en se rapprochant le plus finement possible de la caractéristique d'une fonction de type sigmoïdale. De plus, pour pouvoir bénéficier du caractère reconfigurable d'un réseau de neurones entièrement numérique avec mise à jour des paramètres intégrés, il est nécessaire d'avoir facilement accès à la dérivée de cette fonction d'activation ([Basterretxea, 2003], [Vitabile, 2005], [Da Zhang, 2006], [Torbey, 1992], [Tang, 1997]). Plusieurs types de fonctions

d'activation ont été intégrés numériquement, des plus simples (type échelon ou rampe) aux plus complexes, en passant par des approximations linéaires ou non-linéaires par morceaux ou par la fonction sinus ou encore par le développement limité de la fonction sigmoïdale à un ordre plus ou moins élevé. Plus l'approximation a besoin d'être fine, plus le nombre de circuits numériques à synthétiser est important, et plus la dérivée de la fonction est difficile à réaliser. La conséquence est une large surface de silicium occupée, un nombre de ressources mobilisées et une puissance consommée importants, une augmentation des retards et une limitation de la bande passante.

I.5.2.2. Intégration analogique

L'un des principaux avantages d'un réseau de neurones analogique est le traitement des signaux en parallèle au sein de la structure. De plus, le nombre de transistors utilisés pour synthétiser une cellule analogique est nettement plus faible. Par exemple, pour réaliser une fonction d'activation sigmoïdale, le nombre de transistors nécessaire en analogique est de l'ordre de la dizaine ; en numérique, une approximation grossière de la même fonction nécessite une centaine de blocs logiques (*slices*), ce qui conduit à une surface occupée et la puissance consommée plus importantes. Le nombre de transistors et la consommation relatifs à une cellule numérique dépendent de la précision souhaitée, et donc du nombre de bits sur lequel le signal est codé. Le caractère continu du traitement des signaux analogiques sur l'intégralité d'une caractéristique de transfert peut être primordial en fonction de l'application (notamment si l'on souhaite effectuer une régression linéaire continue et précise d'une fonction à l'aide d'un réseau de neurones de type perceptron multi-couches). Le choix d'une implantation analogique permet de diminuer la place occupée et la consommation pour une précision identique.

L'inconvénient majeur de ce type d'implantation est la non-reconfigurabilité de la structure. Ainsi, une fois le circuit intégré ou le circuit imprimé conçu, il n'est pas possible d'effectuer d'importantes modifications sur la structure du réseau de neurones. De plus, la réalisation d'un circuit intégré implique une phase de dessin des masques des circuits (phase de layout) et une phase de fabrication longue et onéreuse. La phase de layout demande une connaissance certaine de la technologie choisie et des règles de dessin, de sorte que le comportement de l'ensemble des circuits dessinés soit identique à celui des circuits utilisés en simulation, avec leur modèle associé. Néanmoins, des erreurs de tension de décalage et de caractéristique de fonctionnement peuvent apparaître, même avec un layout soigné. Le temps relatif aux étapes

de layout et de conception de la puce est de l'ordre de six mois. Le coût associé est important, du fait de la génération d'un jeu de masques uniquement pour concevoir un nombre de puces limité. La phase de conception analogique (structure choisie, placement et taille des composants, etc.) est donc primordiale. Une structure numérique a l'avantage de pouvoir être reconfigurée, en arrangeant différemment les connexions relatives aux différents transistors impliqués. De plus, le transfert de la phase de simulation des circuits à celle de réalisation "hardware" de la puce est assez simple et bien maîtrisée, du fait du grand nombre de types d'intégrations numériques existantes.

La première version commerciale disponible à grande échelle d'une implémentation entièrement analogique est le circuit 80170NX d'Intel, plus connu sous le nom d'ETANN (*Electrically Trainable Analog Neural Network*). Ce réseau non bouclé de type perceptron multi-couches est constitué de 64 neurones. D'autres réseaux de neurones analogiques ont été implantés soit pour la démonstration, soit pour des applications bien précises ([Morie, 1994], [Al-Nsour, 2001]). Certaines de ces structures présentent un apprentissage analogique implanté avec le réseau de neurones, afin de l'entraîner directement au sein de l'application dans lequel il est intégré. L'algorithme de mise à jour réalisé à l'aide de fonctions analogiques permet de minimiser les erreurs de tension de décalage ainsi que la consommation et la taille de la puce. En revanche, les méthodes de stockage des valeurs des paramètres du réseau sont nettement moins robustes que les méthodes numériques. C'est la raison pour laquelle la plupart des réseaux de neurones analogiques nécessitant une mise à jour de leurs paramètres sont des réseaux de neurones mixtes, avec une partie algorithmique numérique. Ces types de réseaux sont présentés dans le paragraphe suivant (I.5.2.3).

Enfin, certains réseaux de neurones analogiques sont entraînés par l'intermédiaire d'un logiciel annexe. Celui-ci détermine les valeurs des paramètres mis en jeu en tenant compte des imperfections de la structure neuronale implantée. L'avantage de cette méthode est la simplification du système, puisque la mise à jour ne se fait pas directement sur le réseau de neurones. En revanche, cette méthode est moins autonome et demande un temps de calibrage de la structure neuronale assez important, notamment au niveau de la détermination des fonctions de transfert de chacune des cellules et des transferts entre cellules.

Bien qu'il existe des réseaux de neurones analogiques de type RBF ([Dogaru, 1996]), le réseau de neurones analogique intégré le plus fréquent est le perceptron multi-couches, pour son aptitude à réaliser des fonctions de régression non linéaires continues. Ses trois fonctions de base - le multiplieur, l'additionneur et la fonction sigmoïdale - ont respectivement pour rôle

d'amplifier le signal d'entrée par un jeu de poids, de modifier la valeur continue des signaux résultants par un jeu de biais, puis de leur appliquer une non-linéarité de manière à limiter leur dynamique.

Plusieurs structures analogiques de multiplieurs ont déjà été développées et souvent intégrées en technologie CMOS ([Qin, 1987], [Song, 1990], [Khachab, 1991], [Choi, 1992], [Saxena, 1994], [Coue, 1996], [Abutalebi, 1998]). Tandis que les premiers multiplieurs analogiques ont été optimisés notamment en termes de dynamique pour des applications conventionnelles telles que la modulation-démodulation, la transposition en fréquence, etc., les systèmes neuronaux nécessitent des multiplieurs basse consommation, de petite taille et paramétrables. La structure du multiplieur choisi dépend de l'application à laquelle le réseau de neurones est dédiée. La plupart de ces structures dérivent de la cellule de Gilbert ([Gray, 2001]). Un exemple de multiplieur par un poids est présenté dans l'annexe C au sein d'une application neuronale (figure C-e).

La structure de l'additionneur est rarement décrite dans les différentes publications relatives aux intégrations de cellules analogiques de réseaux de neurones. Cela peut se comprendre par le fait que le principe de l'addition est assez simple et bien connu, que ce soit pour des sommes de tension par l'intermédiaire d'amplificateurs bouclés ou pour des sommes de courant. Un exemple d'addition par un biais est présenté dans l'annexe C au sein d'une application neuronale (figure C-e).

La fonction non linéaire est plus ou moins décrite. La structure choisie dépend de l'application et de la complexité, de la précision et de la maîtrise du circuit souhaitées ([Lu, 2000], [Masmoudi, 1999]). Ces paramètres vont conditionner la facilité à calibrer le réseau au sein de l'application ou par l'intermédiaire d'un logiciel annexe, tout en étant aussi responsable de l'aptitude du réseau à réaliser la fonction désirée. Pour certaines applications, une simple rampe bornée à deux valeurs suffit, alors que pour d'autres applications, la courbure de la fonction non linéaire aux alentours de la saturation a toute son importance, notamment au niveau de la continuité de sa dérivée intervenant dans l'algorithme de mise à jour. Un exemple de fonction non linéaire est présenté dans l'annexe C au sein d'une application neuronale (figure C-d).

I.5.2.3. Intégration électronique mixte

Certains réseaux de neurones combinent les avantages des deux types d'implémentation analogiques et numériques (*mixed-mode* ou *hybrid* ([Dias, 2004])). Certaines structures neuronales non bouclées et sans apprentissage intégré font appel aux deux techniques de manière à profiter des avantages de l'implantation analogique en terme de surface de silicium occupée et de ceux de l'implantation numérique pour améliorer le rapport signal sur bruit des signaux d'entrée et de sortie ([Malhotra, 1993], [Pan, 2003]).

D'autres structures avec apprentissage intégré font appel à des cellules analogiques pour réaliser le cœur du réseau (c'est-à-dire la partie non bouclée) et à la partie numérique pour stocker l'algorithme associé à la rétropropagation de l'erreur et les valeurs des paramètres mis à jour ([Gowda, 1993], [Satyanarayana, 1992]). Ainsi, la taille, le temps de transit de l'information sur la chaîne directe et la consommation sont optimisés grâce à la conception du réseau de neurones "feedforward" à l'aide de composants analogiques, et la structure reste reconfigurable et re-paramétrable directement au sein de l'application par la partie numérique de mise à jour des paramètres intervenant au sein du réseau. Le système décrit dans l'annexe C présente une application permettant de tirer profit des avantages des deux méthodes d'intégration.

I.5.2.4. Intégration optoélectronique

Même si les intégrations électroniques de réseaux de neurones sont les plus développées, il existe aussi une autre forme d'intégration : l'intégration optoélectronique ([Burns, 1996], [Ruiz-Llata, 2005]). Ces réseaux de neurones ont en particulier l'avantage de minimiser le temps de traitement de l'information au sein du réseau, puisque les informations circulent à la vitesse de la lumière. Dans les deux références citées ci-dessus, les premières couches du réseau sont optiques tandis que la couche de sortie est analogique. L'interface entre les différents types de couche est réalisée par des capteurs. L'utilité d'une couche de sortie analogique est d'éviter toute conversion analogique-numérique, ce qui se traduit par un gain en consommation, en temps de calcul et en taille de la puce.

I.6. CONCLUSION

Un réseau de neurones est une composition de fonctions non linéaires réalisées par chacun des neurones qui le constitue. Les deux principaux types de réseaux de neurones sont les réseaux RBF, dont les neurones élémentaires présentent des non-linéarités locales, et les PMC, dont les neurones permettent d'effectuer des approximations globales et continues sur l'intégralité de l'espace des entrées. Ces réseaux peuvent être bouclés ou non bouclés suivant l'application.

Les réseaux de neurones à apprentissage supervisé ont pour propriété d'être des approximateurs universels. Pour effectuer des tâches de régression de fonctions non-linéaires et/ou de classification, les réseaux de neurones à apprentissage supervisé présentent une forme de calibrage qui consiste à minimiser l'erreur entre un jeu de valeurs expérimentales et le jeu de valeurs de sortie du réseau de neurones associé, calculé algébriquement en fonction de ses paramètres.

Le type de réseau, sa structure et ses paramètres sont d'abord déterminés à partir de quelques couples de vecteurs d'entrées-sorties durant une phase dite d'apprentissage puis validés pour l'ensemble des couples de vecteurs d'entrées-sorties durant une phase dite de validation, de manière à constater la faculté de généralisation du réseau pour l'application à laquelle il est destiné. La détermination des paramètres se fait par itérations successives. A chaque début d'itération, des couples d'entrées-sorties appartenant au jeu d'apprentissage sont choisis aléatoirement. L'ensemble des paramètres est mis à jour à la fin de chaque itération par la méthode usuelle de "rétropropagation de l'erreur", en calculant le gradient de chaque paramètre en fonction l'erreur commise.

Si l'application s'y prête, il existe plusieurs moyens d'intégrer un réseau de neurones, de manière à effectuer des tâches de régression de fonctions non-linéaires et/ou de classification directement au sein du système. Le choix du type d'intégration dépend en particulier de la nature des signaux issus de l'application, des contraintes de consommation, de surface disponible et de vitesse de traitement des signaux.

Après cette brève présentation des réseaux de neurones, abordons la problématique qui fait l'objet de cette thèse, c'est-à-dire la réalisation et l'intégration de réseaux de neurones pour la télémétrie laser.

I.7. REFERENCES

- [Abutalebi, 1998] : ABUTALEBI A.H., FAKHRAIE S.M., A submicron analog neural network with adjustable level output unit, *Proc. of the 10th Int. Conf. on Microelectronics*, pp. 294–297, Dec 1998.
- [Al-Nsour, 2001] : AL-NSOUR M, ABDEL-ATY-ZOHDY H.S., MOS Fully Reinforcement Neural Network Chip, *Circuits and Systems, ISCAS 2001, 2001 IEEE International Symposium on*, Volume 3, 6-9 May 2001 Page(s):237 - 240 vol. 2.
- [Barron, 1993] : BARRON A., Universal approximation bounds for superposition of a sigmoidal function, *IEEE Transactions on Information Theory*, 39, p. 930-945.
- [Basterretxea, 2003] : BASTERRETXEIA K., TARELA J.M., DEL CAMPO I., Approximation of sigmoid function and the derivative for hardware implementation of artificial neurons, *IEE Proceedings on Circuits, Devices and Systems*, Vol. 151, Issue 1, 5 Feb. 2004, Page(s):18 - 24
- [Bishop, 1995] : BISHOP C.M., *Neural Network for pattern recognition*, Oxford University Press, 1995.
- [Broyden, 1970] : BROYDEN C.G., *The convergence of a class of double-rank minimization algorithms 2 : the new algorithm*, *Journal of the institute of Mathematics and its applications*, 6, p. 222-231.
- [Burns, 1996] : BURNS D.C., UNDERWOOD I., MURRAY A.F., VASS D.G., Design and characterisation of an optoelectronic neural chip, *Elsevier, Optics communications*, Vol. 127, n°4, 15 June 1996, pp. 313-324.
- [Castanie, 2002] : CASTANIE F., ROVIRAS D., Neural networks in space communications, *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on Volume 1*, Page(s):3 - 7 vol.1 (2002).
- [Choi, 1992] : CHOI J., SHEU B.J., VLSI design of compact and high-precision analog neural network processors, *Neural Networks, 1992. IJCNN., International Joint Conference on Volume 2*, 7-11 June 1992, Page(s):637 - 641 vol.2.
- [Coue, 1996] : COUE D., WILSON G., A four-quadrant subthreshold mode multiplier for analog neural-network applications, *IEEE Transactions on Neural Networks*, Volume 7, Issue 5, Sept. 1996 Page(s):1212 – 1219.
- [Da Zhang, 2006] : DA ZHANG, HUI LI, A low cost digital implementation of feed-forward neural networks applied to a variable-speed wind turbine system, *37th IEE Power Electronics Specialists Conference*, June 2006, Page(s):1 – 6.
- [De Almeida, 2000] : DE ALMEIDA A.R., FREIRE E.O., RONNOW C.A., VIANNA J.E.S., ROSI R.M., Neural network-based geometric references recognition applied to ultrasound echo signals, *Circuits and Systems, 2000. Proceedings of the 43rd IEEE Midwest Symposium*, Volume 3, Page(s):1344 - 1347 vol.3 (8-11 Aug. 2000).
- [Dias, 2004] : DIAS F.M., ANTUNES A., MOTA A.M., Artificial neural networks, a review of commercial hardware, *Engineering Applications of Artificial Intelligence*, Vol.17, Issue 8, Page(s):945-952, Dec. 2004.

- [Dogaru, 1996] : DOGARU R., MURGAN A.T., ORTMANN S., GLESNER M., A modified RBF neural network for efficient current-mode VLSI implementation, *Proceedings of Fifth International Conference on Microelectronics for Neural Networks*, 12-14 Feb. 1996, Page(s):265 – 270
- [Dreyfus, 2004] : DREYFUS G., MARTINEZ J.M., SAMUELIDES M., GORDON M.B., BADRAN F., THIRIA S., HERAULT L., *Réseaux de neurones, méthodologie et applications, sous la direction de Gérard Dreyfus*, 2^{ème} édition, Eyrolles, avril 2004.
- [Gowda, 1993] : GOWDA S. M., SHEU B.J., CHOI J., HWANG C-G., CABLE J.S., Design and Characterisation of Analog VLSI Neural Network Modules, *IEEE Journal of Solid-State Circuits*, vol.28, n°3, pp. 301-313, March 1993.
- [Gray, 2001] : GRAY P., HURST P., LEWIS S., MEYER R., *Analysis and design of analog integrated circuits, 4th ed., John Wiley and sons*, 2001.
- [Haykin, 1994] : HAYKIN S., *Neural Network : A Comprehensive Foundation*, Prentice Hall, 1994.
- [Hornik, 1989] : HORNIK K., STINCHCOMBE M., WHITE H., Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, p. 359-366.
- [Hornik, 1990] : HORNIK K., STINCHCOMBE M., WHITE H., Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, 3, p. 551-560.
- [Hornik, 1991] : HORNIK K., Approximation capabilities of multilayer feedforward networks, *Neural Networks*, 4, p. 251-257.
- [Hui, 2004] : HUI L.E., SENG K.P., TSE K.M., RBF neural network mouth tracking for audio-visual speech recognition system, *TENCON 2004, 2004 IEEE Region 10 Conference*, Vol. 1, 84 - 87.
- [Khachab, 1991] : KHACHAB N. I., ISMAEL M., A nonlinear CMOS analog cell for VLSI signal and information processing, *IEEE J. Solid-state Circuits*, vol. 26, no. 11, pp. 1689-1699, Nov. 1991.
- [Kim, 1991] : KIM S. S., SANDERS T. H. Jr, Thermodynamic modeling of phase diagrams in binary alkali silicate systems, *Journal of the American Ceramics Society*, 74, p. 1833-1840.
- [Knerr, 1992] : KNERR S., PERSONNAZ L., DREYFUS G., Handwritten digit recognition by neural networks with Single-layer Training, *IEEE Transactions on Neural Networks*, 3, p. 962-968.
- [Kohonen, 1988] : KOHONEN T., *Self organisation and associative memory*, 2^{ème} édition, Springer Series in Information Sciences, 1988.
- [Levenberg, 1944] : LEVENBERG K., A method for the solution of certain non-linear problems in least squares, *Quarterly Journal of Applied Mathematics*, 2, p. 164-168.
- [Lu, 2000] : LU C., SHI B., Circuit Realization of a Programmable Neuron Transfer Function and its Derivative, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, Volume 4, 24-27 July 2000 Page(s):47 - 50.
- [Malhotra, 1993] : MALHOTRA R.P., SIFERD R., Evaluation of electronic artificial neural network implementations, *Proceedings of the IEEE 1993 National Aerospace and Electronics Conference*, 24-28 May 1993, Page(s):885 - 891 vol.2.
- [Mallat, 1989] : MALLAT S., A theory for multiresolution signal decomposition : the wavelet transform, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, p. 674-693.

- [Marquardt, 1963] : MARQUARDT D.W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society of Industrial and Applied Mathematics*, 11, p. 431-441.
- [Masmoudi, 1999] : MASMOUDI M., SAMET M., TAKTAK F., ALIMI A.M., A hardware implementation of neural network for the recognition of printed numerals, *11th International Conference on Microelectronics*, 22-24 Nov. 1999, Page(s):113 – 116.
- [Mc Culloch, 1943] : McCULLOCH, W. S., PITTS W., A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5, p. 115-133.
- [Minsky, 1969] : MINSKY M., PAPERT S. *Perceptrons*. MIT Press.
- [Morie, 1994] : MORIE T., AMEMIYA Y., An All-Analog Expendable Neural Network LSI with On-Chip Backpropagation Learning, *IEEE Journal of Solid-State Circuits*, vol.29, n°9, pp. 1086-1093, Sept. 1994.
- [Mozayyani, 1998] : MOZAYYANI N., BAIG A.R., VAUCHER G., A fully-neural solution for online handwritten character recognition, *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence*, Volume 1, Page(s):160 - 164 vol.1 (4-9 May 1998).
- [Novakovic, 1994] : NOVAKOVIC B.M., Feedforward neural networks for adaptive nonlinear robot control, *Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World', IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on*, Volume 1, 12-16 Sept. 1994 Page(s):486 - 493 vol.1.
- [Oukhellou, 1997] : OUKHELLOU L., AKNIN P., Modified Fourier Descriptors : A new parametrization of eddy current signatures applied to the rail defect classification, *III International workshop on advances in signal processing for non destructive evaluation of materials*.
- [Pan, 2003] : PAN D., WILAMOWSKI B.M., A VLSI implementation of mixed-signal mode bipolar neuron circuitry, *Proceedings of the International Joint Conference on Neural Networks*, Volume 2, 20-24 July 2003, Page(s):971 - 976
- [Perrone, 1999] : PERRONE, A.L., BASTI, G., A new criterion of NN structure selection for financial forecasting, *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, Volume 6, 10-16 July 1999 Page(s):3898 - 3903 vol.6.
- [Powell, 1987] : POWELL M. J. D. [1987], Radial basis functions for multivariable interpolation : a review, *Algorithms for approximation*, p. 143-167.
- [Qin, 1987] : QIN S., GEIGER R. L., A $\pm 5V$ CMOS analog multiplier, *IEEE J. Solid-State Circuits*, vol. 22, no. 6, pp. 1143-1146, Dec. 1987.
- [Rivals, 1994] : RIVALS I., CANAS D., PERSONNAZ L., DREYFUS G., Modeling and control of mobile robots and intelligent vehicles by neural networks, *Proceedings of the IEEE Conference on Intelligent Vehicles*, p. 137-142.
- [Rivals, 1995] : RIVALS I., *Modélisation et commande de processus par réseaux de neurones: application au pilotage d'un véhicule autonome*, Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris.
- [Roussel, 2001] : ROUSSEL P., MONCET F., BARRIEU B., VIOLA A., Modélisation d'un processus dynamique à l'aide de réseaux de neurones bouclés. Application à la modélisation de la relation pluie-hauteur d'eau dans un réseau d'assainissement et à la détection de défaillances de capteurs, *Innovative technologies in urban drainage*, 1, 919-926, G.R.A.I.E.

- [Ruiz-Llata, 2005] : RUIZ-LLATA M., LAMELA-RIVERA H., Image identification system based on an optical broadcast neural network processor, *Applied Optics*, Vol. 44, n°12, 20 April 2005.
- [Rumelhart, 1986] : RUMELHART D.E., HINTON G.E., WILLIAMS R.J., Learning internal representations by error backpropagation, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, p.318-362, MIT Press.
- [Santos, 2004] : SANTOS, P.J., MARTINS, A.G., PIRES, A.J., Short-term load forecasting based on ANN applied to electrical distribution substations, *Universities Power Engineering Conference, 2004. UPEC 2004. 39th International*, Volume 1, 6-8 Sept. 2004 Page(s):427 - 432 Vol. 1
- [Satyanarayana, 1992] : SATYANARAYANA S., TSIVIDIS Y.P., GRAF H.P., A Reconfigurable VLSI Neural Network, *IEEE Journal of Solid-State Circuits*, vol.27, n°1, pp. 67-81, January 1992.
- [Saxena, 1994] : SAXENA N., CLARK J.J., A four-quadrant CMOS analog multiplier for analog neural networks, *Solid-State Circuits, IEEE Journal of*, Volume 29, Issue 6, June 1994 Page(s):746 – 749.
- [Song, 1990] : SONG H.-J., KIM C.-K., An MOS four-quadrant analog multiplier, using simple two-input squaring circuits with source followers, *IEEE J. Solid-State Circuits*, vol. 25, no. 3, pp. 841-848, June 1990.
- [Stricker, 2001] : STRICKER M., VICHOT F., DREYFUS G., WOLINSKI F., Training context-sensitive neural networks with few relevant examples for the TREC-9 routing, *Proceedings of the TREC-9 Conference*.
- [Tang, 1997] : TANG C.Z.; KWAN H.K., Digital implementation of neural networks with quantized neurons, *Proceedings of 1997 IEEE International Symposium on Circuits and Systems*, Volume 1, 9-12 June 1997, Page(s):649 - 652
- [Torbey, 1992] : TORBEY E., HAROUN B., Architectural synthesis for digital neural networks, *International Joint Conference on Neural Networks*, Volume 2, 7-11 June 1992 Page(s):601 – 606.
- [Vitabile, 2005] : VITABILE S. CONTI V., GENNARO F., SORBELLO F., *Proceedings of the 8th Euromicro Conference on Digital System Design*, 30 Aug.-3 Sept. 2005, Page(s):218 – 222.
- [Werbos, 1974] : WERBOS P.J., *Beyond regression : new tools for prediction and analysis in the behavioural sciences*, Ph. D. thesis, Harvard University.
- [Wojtek Przytula, 1991] : WOJTEK PRZYTULA K.; Parallel digital implementations of neural networks, *Proceedings of the International Conference on Application Specific Array Processors*, 2-4 Sept. 1991, Page(s):162 – 176.
- [Yang, 1999] : YANG X., MENG M., Neural network application in robot motion planning, *Communications, Computers and Signal Processing, 1999 IEEE Pacific Rim Conference on*, 22-24 Aug. 1999 Page(s):611 – 614.
- [Yang, 2003] : YANG F., PAINDAVOINE M., Implementation of an RBF neural network on embedded systems: real-time face tracking and identity verification, *Neural Networks, IEEE Transactions on*, 14 (5), 1162 – 1175.

Chapitre II

Télémétrie Laser

Ce chapitre présente des dispositifs optoélectroniques appliqués à la mesure de distance et à la détection de surfaces, et plus particulièrement les applications télémétriques. Les problématiques associées sont exposées en fin de chapitre.

II.1. INTRODUCTION

Lorsqu'une onde acoustique ou électromagnétique illumine une surface quelconque, une fraction de l'énergie est réfléchiée ou rétrodiffusée dans l'espace du demi-plan de cette surface. Pour distinguer cette surface cible de tout l'environnement dans laquelle elle est placée, il est nécessaire que la majorité de la puissance de l'onde incidente soit en interaction avec la cible, afin d'obtenir une analyse quasi-ponctuelle. Si la distance à mesurer est grande devant la dimension de la cible à détecter, il est nécessaire que le faisceau incident d'analyse soit de faible section et peu divergent. Ceci est indispensable lorsque la puissance d'émission est relativement faible et qu'elle doit encore satisfaire le rapport signal sur bruit du bilan de liaison.

Une faible divergence du faisceau d'analyse est obtenue lorsque la pupille de la source est de grande section en comparaison de la longueur d'onde, ce qui donnera une faible diffraction du faisceau d'analyse. Grâce aux diodes laser, les techniques optiques dont la longueur d'onde est proche de $1\mu\text{m}$ permettent de fournir à la fois des faisceaux de faible divergence et de forte puissance.

La mesure de distance et la détection d'objets sont devenues essentielles dans de nombreux domaines tels que la détection d'obstacles pour les véhicules ou les robots autonomes ([Omura, 1999], [Tiedeke, 1990], [Hueber, 2000], [Okada, 2005], [Dietmayer, 2001]), les applications médicales ([Cinquin, 1995]), les procédés industriels et agricoles ([Maatta, 1993], [Liessem, 2002], [Rouveure, 2003]), la détection de surface des routes ([Fukui, 1997],

[Yoda, 1995], [Sarabandi, 1997]), les systèmes de surveillance et de sécurité ([Williams, 1995]), etc.

Les conditions de détection dépendent de l'application souhaitée, des types de surfaces à discriminer, de l'intervalle de distance sur lequel la détection doit se faire et de la position angulaire du capteur par rapport à la surface. Selon ces conditions, on choisit un type précis de capteur, ayant les propriétés physiques optimales pour l'application considérée. Ces capteurs peuvent être optiques (opérant dans le visible ([Omura, 1999], [Fukui, 1997], [Ozdemir, 2001]) le proche infrarouge ([Tiedeke, 1990], [Yoda, 1995], [Neelamegan, 2003]), ou l'ultraviolet ([Pauchard, 2000], [Nerguizian, 2004], [Norris, 1999])), à ultrasons ([Hueber, 2000], [Carullo, 1996], [Fiorillo, 2002], [De Almeida, 2000], [Vossiek, 1994]), radar ([Rouveure, 2003], [Joodaki, 2001], [Sarabandi, 1997], [Giubolini, 1999]), tactile ([Baglio, 2002]), etc. Les capteurs à ultrasons sont utilisés dans les environnements nécessitant de la détection d'obstacles sur des distances moyennes (de l'ordre du mètre), les capteurs radar pour les grandes distances, les capteurs tactiles pour les applications où le contact avec la cible est possible et permet l'acquisition de données discriminantes. Les capteurs optiques présentent une gamme de mesure de distance relative importante, pour des surfaces pouvant être difficiles d'accès et dont les dimensions peuvent être petites devant la distance. Le choix de la longueur d'onde du capteur optique dépend notamment de sa réponse par rapport à la cible souhaitée et de l'hostilité de l'environnement, en fonction de l'ensemble des interférences et autres phénomènes perturbateurs du milieu. Par exemple, un faisceau laser CO₂ ayant une longueur d'onde de 10.6µm peut se propager au travers d'une épaisse fumée noire produite par un incendie ([Yamada, 1993]).

Les signaux issus du système de mesure de distance ne sont pas exploitables tels quels. Un traitement de ces signaux, aussi simple soit-il, est toujours nécessaire. En tant qu'approximateurs de fonctions et classificateurs, les réseaux de neurones sont parfois utilisés de manière à extraire plus facilement une mesure de distance ou un type de surface observé ([Omura, 1999], [Rouveure, 2003], [Neelamegan, 2003], [Carullo, 1996], [Fiorillo, 2002], [Joodaki, 2001], [Ozdemir, 2001], [De Almeida, 2000]).

Les principes de mesures de distance sans contact peuvent être séparés en deux grandes méthodes. L'une, passive, se contente d' "analyser" une scène, souvent soumise à un "éclairage naturel" et uniforme. L'autre, dite active, consiste à projeter la puissance d'une onde (ultrasonique ou électromagnétique) sur la scène à étudier et à analyser la fraction de l'énergie de l'onde de retour renvoyée par l'environnement et la cible. Dans ce chapitre, nous

nous placerons dans ce deuxième cas. On distingue trois grandes familles de techniques optiques de mesure de distance par faisceau laser ([Bosch (a), 1995]) :

- les méthodes par triangulation,
- les méthodes par interférométrie,
- les méthodes par temps de vol.

Le principe de base de la mesure de déplacement par technique de triangulation est la projection sur une cible d'un faisceau de lumière collimaté et monochromatique dont la composante diffusante est rétrodiffusée vers un photorécepteur ([Dorsch, 1994], [Häusler, 1997], [Bosch, 1994]). Bien que cette technique ne nécessite pas de performances extrêmes au niveau de la source optique et de l'électronique associée, son principal inconvénient réside dans l'existence d'une "zone d'ombre" due à la non-colinéarité des faisceaux d'émission et de réception .

Le principe de base de la mesure de déplacement par technique interférométrique consiste à compter le nombre de franges d'interférence produites par deux faisceaux lumineux cohérents issus de la même source et déphasés dans un interféromètre de type Michelson par exemple ([Lang, 1997]). Bien que cette technique permette des mesures d'une précision de l'ordre d'une fraction de longueur d'onde pour des distances de l'ordre du mètre, ses principales limitations sont souvent imposées par la nature de la surface des cibles coopératives ([Williams, 1986], [Salvadé, 1997]), et la rugosité de la cible pour des cibles non-coopératives ([Bosch (a), 1995], [Bosch, 1994]).

Le paragraphe suivant aborde la troisième grande famille de techniques optiques de mesure de distance : les méthodes par temps de vol.

II.2. MESURE DE DISTANCE PAR TEMPS DE VOL

II.2.1. Introduction

La mesure de distance D est directement déduite du "temps de vol" aller-retour τ_D de la lumière entre la source et le photorécepteur, via la cible illuminée. Si les faisceaux d'émission et de réception sont co-linéaires ou quasi-parallèles et si le rayon incident illumine la cible de façon quasi-punctuelle, une fraction de l'énergie optique émise par la source sera rétrodiffusée par la surface de la cible dans une direction très voisine de celle du faisceau incident. En supposant que l'indice de groupe de l'atmosphère à la longueur d'onde λ est égal à 1, la distance D à mesurer est directement déduite de la relation :

$$D = \frac{1}{2} c \tau_D \quad (\text{II-1})$$

Ainsi, la connaissance de la vitesse de propagation de la lumière dans le milieu considéré (air) permet d'accéder directement à la valeur absolue de la distance, par opposition aux dispositifs de comptage ou de défilement qui n'indiquent que des déplacements. A priori, la technique par temps de vol est l'une des mieux adaptées pour la mesure de distance sur une plage importante et pour des cibles quelconques, du type diffusantes ou non coopératives. La figure II-a représente un exemple de tête optique d'un dispositif fonctionnant par temps de vol. La source laser émet un faisceau optique collimaté ou faiblement divergent. Le photorécepteur est positionné sur l'axe et au foyer de l'optique de collecte.

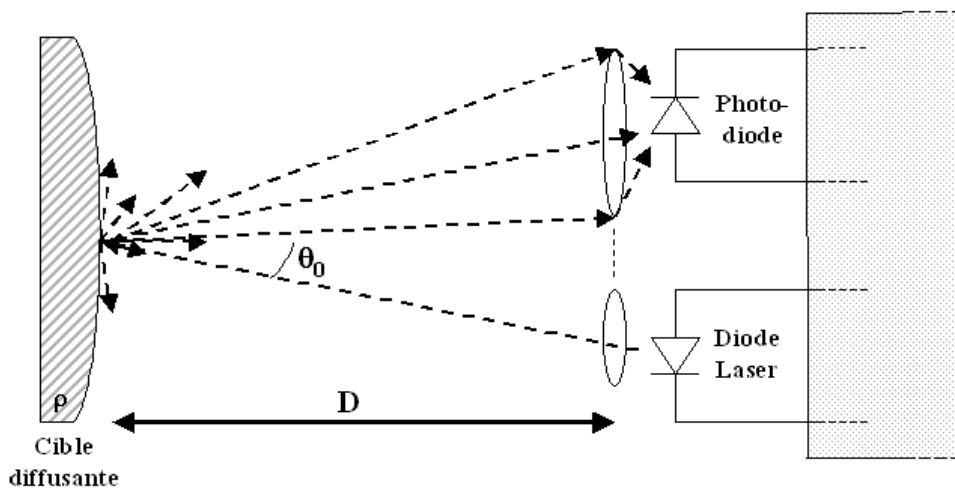


Figure II-a : Représentation de la tête optique d'un télémètre à diode laser fonctionnant par temps de vol.

La technique de mesure de distance par temps de vol est bien adaptée aux surfaces cibles diffusantes, dont la composante de rétrodiffusion lambertienne est importante. Cette propriété offre comme avantage d'avoir une orientation quelconque de la surface cible par rapport au

rayon laser d'émission. Pour accéder à la détermination de la distance par temps de vol, il apparaît principalement deux techniques "classiques" de modulation de la puissance optique émise par la source laser : le régime impulsionnel et le régime sinusoïdal permanent. Concernant la deuxième technique, deux méthodes sont possibles, l'une basée sur la modulation de fréquence (FMCW) et l'autre fonctionnant à fréquence fixe (télémétrie laser par déphasage).

II.2.2. Technique impulsionnelle

Les télémètres utilisant une technique impulsionnelle sont les plus répandus pour mesurer de grandes distances absolues ([Maatta, 1993],[Ailisto, 2002]). Le dispositif d'émission lance des impulsions optiques de fortes puissances crêtes, de $1W_{\text{crête}}$ à $100W_{\text{crête}}$, avec un temps de montée très court et une largeur de l'impulsion compatible avec le temps de montée du signal photoélectrique de la réception. La technique en régime impulsionnel permet d'accéder à la mesure de distances sur des cibles non coopératives à priori quelconques dans une plage de 3m à 300m environ, avec une résolution de quelques dizaines de cm en "mono-coup" ([Kaisto, 1983], [Koskinen, 1992]). Pour des applications représentatives, les largeurs à mi-hauteur des impulsions sont de l'ordre de 10ns et la fréquence de répétition est voisine de la dizaine de kHz. Le rapport cyclique et la puissance moyenne sont donc relativement faibles. Cette propriété est très intéressante car elle permet de concevoir des dispositifs satisfaisants les conditions d'exposition maximales permises en terme de sécurité oculaire ([CEI 825]).

II.2.3. Régime sinusoïdal permanent

La puissance optique émise par la source laser est modulée en amplitude par un signal sinusoïdal (ou carré avec un rapport cyclique proche de 50%). La méthode la plus simple est la modulation de l'intensité du courant de polarisation directe de la diode laser. Si on admet une modulation linéaire (c'est à dire du type "petit signal"), l'intensité du courant de polarisation de la diode laser est de la forme :

$$I_{DL}(t) = I_{th} + I_{DL}(1 + m \cdot \sin(2\pi f_{RF}t)) \quad (II-2)$$

avec I_{th} le courant de seuil de la diode laser, I_{DL} l'amplitude du courant moyen de modulation, m l'indice de modulation (dans la pratique, $m \approx 1$), et f_{RF} la fréquence du signal de modulation. La puissance optique émise est alors de la même forme :

$$P_{DL}(t) \cong P_{DL}(1 + m \sin(2\pi f_{RF}t)) \text{ avec } m \approx 1 \quad (II-3)$$

Jusqu'à des fréquences de modulation supérieures à plusieurs centaines de MHz, on pourra considérer que la diode laser est aperiodique en fonction du courant alternatif de polarisation. On peut donc décomposer la puissance optique émise en la somme d'une puissance optique moyenne $\langle P_{DL} \rangle$ et d'une puissance optique "alternative" d'amplitude ou de valeur crête P_{DL} :

$$P_{DL}(t) \equiv \langle P_{DL} \rangle + P_{DL} \sin(2\pi f_{RF} t) \quad (II-4)$$

II.2.3.1. Télémétrie par modulation de fréquence (pseudo FMCW)

Une première méthode de mesure de distance par temps de vol en régime sinusoïdal permanent repose sur la modulation de la puissance optique par un signal lui-même modulé en fréquence dans un large domaine (une octave). Pour une modulation linéaire en fréquence, la mesure de distance est déterminée par la valeur d'une fréquence de battement issue du décalage entre le signal d'émission et le signal photoélectrique de retour ([Bazin, 1996]). La figure II-b représente le schéma synoptique du dispositif.

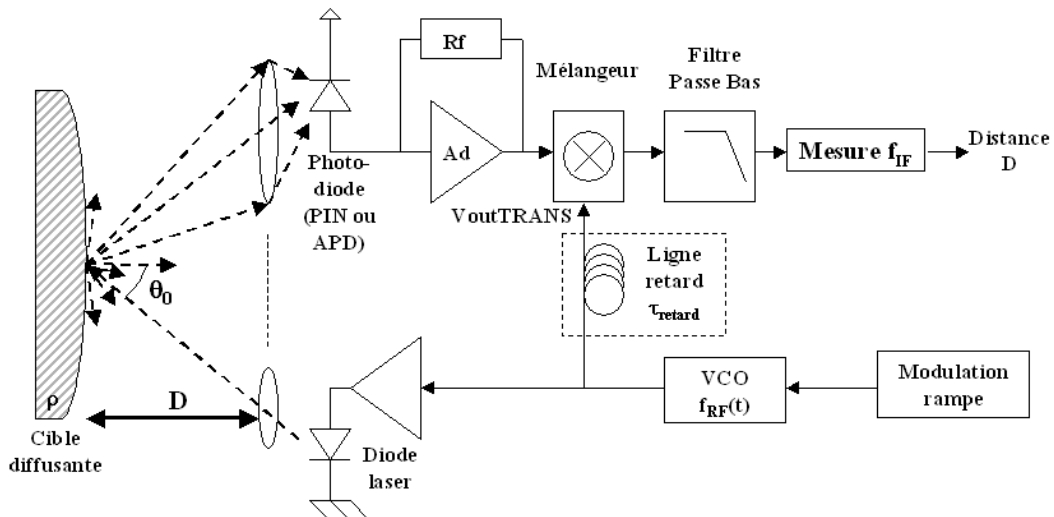


Figure II-b : Schéma synoptique d'un télémètre FMCW. Une ligne à retard facilite la mesure de la fréquence de battement.

Le signal de modulation est généré par un oscillateur haute fréquence contrôlé en tension (VCO). Le signal photoélectrique rétrodiffusé par la cible, est multiplié par le signal de modulation directement issu du VCO par l'intermédiaire d'une ligne à retard ([Dupuy,2001]). On obtient ainsi un signal de battement (de fréquence intermédiaire f_{IF}) disponible à la sortie du mélangeur et du filtre passe-bas. La vitesse de variation a_{VCO} de la fréquence du VCO est très élevée. En modulant linéairement la fréquence du signal généré par le VCO par un signal triangulaire de période T_r , la vitesse de variation a_{VCO} s'écrit :

$$a_{VCO} = \frac{df_{RF}}{dt} = \pm \frac{\Delta f}{T_r/2} \quad (\text{MHz} \cdot \mu\text{s}^{-1}) \quad (II-5)$$

Si Δf est l'excursion en fréquence du VCO, la fréquence intermédiaire est proportionnelle au temps de vol :

$$f_{IF}(t) = \left| \pm 2 \frac{\Delta f}{T_r} \right| \tau_D = \left| \pm 2 \frac{\Delta f}{T_r} \right| \frac{2D}{c} \quad (\text{II-6})$$

Si le temps de vol est petit devant la demi période de la rampe de contrôle, on peut négliger les périodes de transition du changement de signe de la pente de la rampe. Si la distance D est suffisamment grande, l'écart entre la fréquence instantanée du VCO et celle du signal photoélectrique est relativement élevée. On obtient alors un certain nombre de périodes de ce signal de battement durant une durée correspondant à la période T_r de la modulation. La distance minimale D_{0min} mesurable avec une modulation linéaire idéale est ([Bazin, 1996]) :

$$D_{0min} = \frac{c}{4\Delta f} \quad (\text{II-7})$$

Pour une excursion $\Delta f = 100\text{MHz}$ du VCO, on en déduit $D_{0min} = 0,75\text{m}$. Pour accéder à des mesures de distances inférieures à D_{0min} , on améliore le dispositif par la mise en œuvre d'une ligne à retard caractérisée par le temps de retard t_{retard} . L'expression de la fréquence intermédiaire ([Kompa, 1997]) devient, pour une modulation linéaire :

$$f_{IF}(t) = a_{VCO}(t_{retard} \pm \tau_D) = a_{VCO} t_{tot} \quad (\text{II-8})$$

La ligne à retard est placée soit entre le VCO et la diode laser, soit entre le VCO et le mélangeur (figure II-b).

II.2.3.2. Télémétrie laser par mesure de déphasage

La technique par déphasage consiste à moduler la puissance du faisceau laser d'émission avec un signal dont la fréquence f_{RF} est fixe ([Poujouly, 2000]). La figure II-c représente le schéma bloc d'un télémètre par déphasage. La puissance optique émise par la diode laser est modulée par un signal sinus ou carré généré par un oscillateur très stable en fréquence. Le signal optique rétrodiffusé par la cible est capté par la photodiode dont le courant photoélectrique est amplifié par un circuit transimpédance.

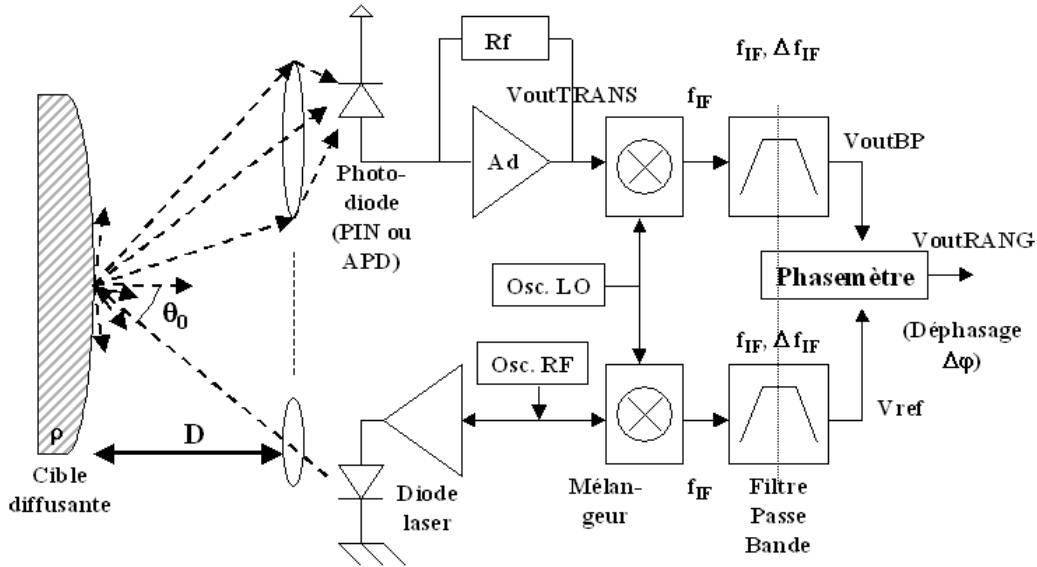


Figure II-c : Schéma de principe du télémètre laser par déphasage avec méthode hétérodyne. Les filtres passe-bande accordés sur la fréquence intermédiaire permettent de réduire la bande passante équivalente de bruit.

La distance D séparant le dispositif de la cible est déduite du glissement de phase $\Delta\phi$ qui existe entre le signal photoélectrique et celui de modulation de la puissance optique :

$$\Delta\phi = 2\pi f_{RF} \tau_D = 2\pi f_{RF} \frac{2D}{c} \quad (\text{II-9})$$

Cependant, la détermination de D est effectuée a priori modulo 2π , ce qui donne une plage d'incertitude égale à un nombre entier de Λ ([Salvadé, 1997]) :

$$\Lambda \equiv D_{\text{Modulo}} = \frac{c}{2f_{RF}} = 0,9\text{m pour } f_{RF} = 166,6\text{MHz} \quad (\text{II-10})$$

Ainsi, plus la fréquence de modulation f_{RF} est importante, plus l'intervalle de mesure est limité. En revanche, la sensibilité de la mesure est accrue à hautes fréquences :

$$\frac{\Delta D}{\Delta\phi} = \frac{c}{2.360.f_{RF}} = 0,25\text{cm/degré pour } f_{RF} = 166,6\text{MHz} \quad (\text{II-11})$$

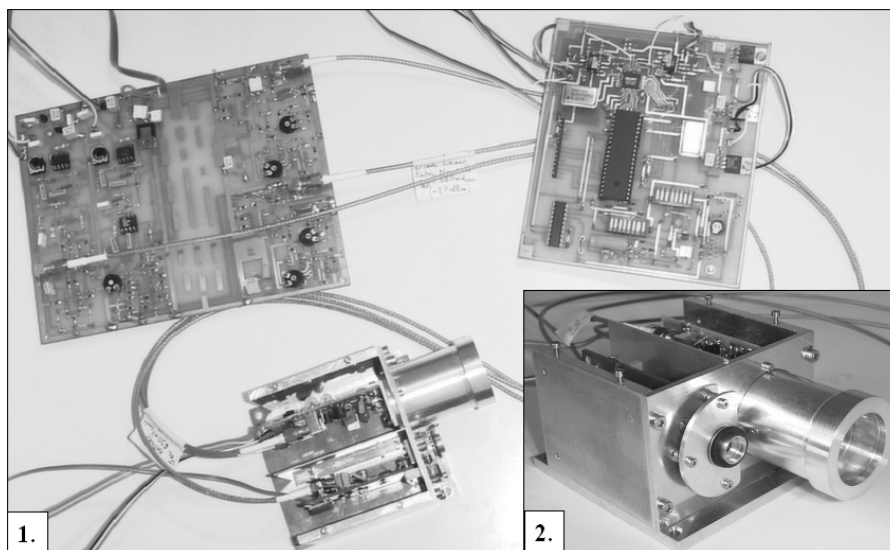
soit, une résolution δD de la mesure en fonction de résolution $\delta\phi$ de la mesure du déphasage :

$$\delta D = \frac{1}{2.360.f_{RF}} \frac{c}{2} \delta\phi = 50\mu\text{m pour } f_{RF} = 166,6\text{MHz et } \delta\phi = 0,02^\circ \quad (\text{II-12})$$

Pour faciliter l'acquisition du déphasage, on effectue la mesure à une fréquence de battement f_{IF} beaucoup plus faible grâce à une technique hétérodyne ([Dziadowiec, 1984]). Le signal utile est alors mélangé avec celui issu d'un oscillateur local f_{LO} de telle sorte que :

$$|f_{RF} - f_{LO}| = f_{IF} \quad \text{avec } f_{IF} \ll f_{RF} \quad (\text{II-13})$$

Un signal auxiliaire est appliqué sur la voie référence du phasemètre. Il est obtenu par le mélange direct entre les signaux issus des deux oscillateurs. Le signal photoélectrique de fréquence intermédiaire f_{IF} est soumis si possible à une amplification à bande étroite de largeur Δf_i afin d'améliorer son rapport signal sur bruit. Pour des raisons de symétrie, un filtre identique est positionné sur la voie référence. On peut montrer que la technique hétérodyne conserve le glissement de phase $\Delta\phi$ entre les signaux de fréquence f_{RF} et ceux de fréquence intermédiaire f_{IF} ([Dziadowiec, 1983], [Poujouly, 2000]). Une photographie du montage décrit ci-dessus est représenté sur la figure II-d.



*Figure II-d : 1. Partie émettrice et réceptrice du télémètre laser par déphasage.
2. Tête optique du télémètre.*

La technique hétérodyne est préférable à celle d'échantillonnage car elle propose un meilleur rapport signal sur bruit et une bande passante équivalente de bruit B_n plus faible. En effet si l'on considère constantes les densités spectrales des bruits d'origine de grenaille et thermique dans la bande passante du signal utile, la bande passante équivalente de bruit s'écrit ([Dziadowiec, 1995]) :

$$B_n = \frac{\pi}{2} \Delta f_i \quad (\text{II-14})$$

Dans la majorité des dispositifs électroniques, le signal minimal détectable est imposé par celui du bruit. Cependant, pour le télémètre par déphasage, une limitation au moins aussi importante, sinon davantage, est le couplage diaphonique électrique qui existe entre les circuits d'émission et ceux de réception. Ce phénomène peut être de très grande importance car "non visible" et avec un niveau relatif bien plus petit que celui du bruit. Ainsi, le bruit "blanc" ne donne qu'une perte de résolution au dispositif, alors que le phénomène de diaphonie ("crosstalk" en anglais) introduit une erreur importante sur la mesure du déphasage

([Bosch, 1997]). Cette erreur varie aussi avec le déphasage réel à déterminer. Une méthode de post correction a été proposée ([Payne, 1992]).

Pour la photodétection classique, l'utilisation d'une APD permet d'obtenir dans les conditions de quasi obscurité un rapport signal sur bruit bien meilleur qu'avec une photodiode PIN (de l'ordre de 10 fois supérieur) (figure II-e).

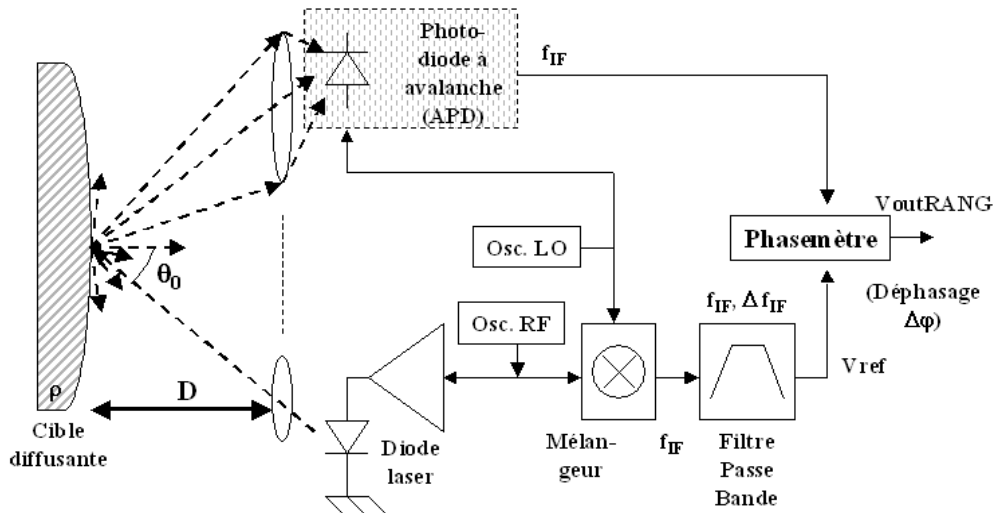


Figure II-e : Schéma de principe du télémètre laser par déphasage avec méthode hétérodyne et mélange optoélectronique au sein de la photodiode à avalanche (APD).

En effet, la bande passante et le rapport signal sur bruit du circuit représenté figure II-c (circuit de type transimpédance associé à la photodiode) sont imposés par la capacité de jonction C_j de la photodiode, par la résistance R_f de contre-réaction et par la réponse en fréquence du gain différentiel en boucle ouverte $A_d(jf)$ de l'amplificateur transimpédance. De plus, pour garantir la stabilité en fréquence et pour obtenir une faible bande passante équivalente de bruit, la marge de phase de l'amplificateur transimpédance doit être supérieure à 45° . Cette condition est difficile à satisfaire avec une photodiode de grande surface active, présentant une grande capacité parasite en entrée de l'amplificateur transimpédance.

Une solution permettant de contourner cette difficulté, améliorer le rapport signal sur bruit et diminuer les effets de la diaphonie est de réaliser un mélange optoélectronique entre l'onde optique retrodiffusée par la cible de fréquence f_{RF} et le signal électrique de fréquence f_{LO} directement au sein d'une photodiode à avalanche (APD) (figure II-e) ([Dupuy, 2002(a)], [Castagnet, 2006]). L'intérêt du mélangeur optoélectronique réside essentiellement dans la suppression de l'amplificateur transimpédance qui génère de forts niveaux de bruit à haute fréquence. Il pourra être remplacé par un simple amplificateur "basse fréquence" dont la tension de bruit ramenée à l'entrée sera très faible. La fréquence du signal utile de sortie sera directement égale à la fréquence intermédiaire f_{IF} .

II.3. CONCLUSION ET PROBLEMATIQUES

II.3.1. Conclusion

De par son principe, la technique par triangulation semble la plus facile à mettre en œuvre. Elle peut être utilisée pour des cibles non coopératives diffusantes. Cependant, si le problème de la zone d'ombre peut être résolu, la dynamique de la plage de mesure reste relativement faible (2cm à 2m environ) pour une résolution relative d'environ 10^{-3} ([Lin, 1983]).

Les mesures de distances issues de l'interférométrie ne semblent pas favorables pour la détection d'obstacles en "temps réel" en raison des limitations associées à la nature et à la rugosité des cibles ([Salvadé, 1997], [Bosch (a), 1995]).

La technique de mesure de distance par temps de vol est la mieux adaptée des trois grandes familles pour la détection d'obstacles. Cette technique demande des composants optoélectroniques et des circuits électroniques plus performants en puissance optique et en bande passante que les techniques issues de la triangulation, mais la plage relative des distances mesurables est plus grande.

La technique en régime impulsionnel permet d'accéder à la mesure de distances sur des cibles non coopératives à priori quelconques dans une plage de 3m à 300m environ, avec une résolution de quelques dizaines de cm en "mono-coup" ([Kaisto, 1983], [Koskinen, 1992]).

De par sa structure, le télémètre laser utilisant une technique de modulation de fréquence possède un certain nombre d'avantages car l'information "distance" est contenue dans une fréquence de battement. Ce second dispositif de mesure par temps de vol présente à priori une meilleure immunité à la diaphonie et aux variations d'amplitude du signal photoélectrique que le télémètre laser par déphasage sur une plage de mesure de 1m à 20m, avec une résolution de l'ordre du millimètre ([Dupuy, 2002(b)]).

La détermination de la distance à partir du déphasage donne de très bons résultats sur des cibles non coopératives, en particulier en terme de résolution, pour des intervalles de mesure allant de 0,5m à 20m. Une résolution proche de la dizaine de μm peut être obtenue à l'aide de cette technique de mesure en modulant la puissance optique du laser à une fréquence de l'ordre du GHz ([Payne, 1992]). C'est cette technique qui a été retenue dans le cadre de cette thèse car elle offre les meilleures performances en terme de précision sur une plage de mesure sans indétermination de l'ordre du mètre.

II.3.2. Problématiques

II.3.2.1. Amplitude du signal photoélectrique reçu

Si la zone éclairée par un faisceau laser d'analyse est considérée plus petite que la surface de la cible non-coopérative et si cette surface cible est supposée diffusante, la lumière rétrodiffusée rayonne selon la loi de Lambert. Ainsi, l'amplitude du courant photoélectrique alternatif de sortie de la photodiode est proportionnelle à l'inverse du carré de la distance D ([Bosch (b), 1995]). L'amplitude du signal de sortie du transimpédance, appelée $V_{outTRANS}$ (figure II-c), peut s'écrire sous la forme :

$$V_{outTRANS} = Z_t i_{PD} = K_1 \frac{\rho \cos(\theta_0)}{D^2} \quad (II-15)$$

avec Z_t la fonction de transfert de l'amplificateur transimpédance, i_{PD} le courant photoélectrique traversant la photodiode (PIN ou APD), ρ_{diff} le coefficient de réflexion diffusant de la cible non-coopérative, θ_0 l'angle d'incidence du faisceau entre la capteur et la cible et K_1 un coefficient de proportionnalité relatif à l'émetteur, au transfert optique et à la réponse de la photodiode. Dans le cadre de l'utilisation d'une photodiode PIN, le terme K_1 s'écrit :

$$K_1 = \frac{1}{\pi} T_\lambda T_R S_\lambda Z_t P_{DL} A_R \quad (II-16)$$

avec T_λ le coefficient de transmission de l'optique d'émission pour une longueur d'onde donnée, T_R le coefficient de réflexion de l'optique de réception, S_λ la réponse spectrale de la photodiode, P_{DL} la puissance crête de l'onde laser transmise et A_R la surface de la pupille de l'optique de photoréception. Le facteur K_1 , indépendant du positionnement et du type de la cible, est considéré comme constant.

II.3.2.2. Mesure de distance

Comme cela a été décrit dans la partie II.2.3.2 relative à la présentation du principe du télémètre laser par déphasage, la mesure de la tension de sortie du télémètre $V_{outRANG}$, proportionnelle à la mesure du déphasage $\Delta\phi$, permet d'obtenir une valeur précise de la distance entre la cible et le télémètre. Néanmoins, la mesure étant limitée par le principe même de mesure de différence de phase (modulo 2π), un choix est à effectuer entre plage utile de mesure et résolution de la mesure. En effet, plus la fréquence de modulation f_{RF} est importante, plus la résolution δD de la mesure est importante, mais plus la plage de mesure Λ est restreinte (équations II-9 à II-12). La problématique réside dans la possibilité d'accroître

l'intervalle de mesure de distance à partir du même dispositif, sans nécessiter de mesures supplémentaires, tout en gardant une résolution élevée.

II.3.2.3. Classification de surfaces

Le coefficient de réflexion diffusant est caractéristique d'une surface testée à une longueur d'onde considérée. Si l'on se place à une distance D et à un angle d'incidence θ_0 donnés, il est possible de classer des cibles de différentes rugosités grâce au signal $V_{outTRANS}$ (II-15). La figure II-f.1 montre un cas simple de discrimination de quatre types de surfaces S_1 à S_4 (respectivement une surface plastifiée, du papier glacé, un mur peint et une surface poreuse) par l'intermédiaire du signal $V_{outTRANS}$ à une distance de 1m et pour un angle d'incidence nul.

Les différentes droites représentent en échelle log-log l'amplitude du signal de sortie de l'amplificateur transimpédance $V_{outTRANS}$ pour les quatre types de surfaces dans le cas idéal, sur l'intervalle de distance [0.5m ; 2m] pour un angle d'incidence nul. Les coefficients de réflexion diffusants apparents ρ de chaque surface, déterminés expérimentalement à l'aide du télémètre, sont respectivement égaux à 0.9, 0.6, 0.35 et 0.15. L'amplitude du signal photoélectrique $V_{outTRANS}$ a été normalisée, de telle sorte qu'une amplitude de 1V corresponde au test de la surface plastique (surface la plus réfléchissante des quatre surfaces dans la direction normale) pour la distance la plus courte ($D_{min} = 0.5m$). Lorsque le télémètre se situe à une distance fixe de la cible ($D=1m$) et que l'angle d'incidence est constant ($\theta_0=0$), chaque tension de sortie V_{outS_i} observée ($i \in [1;4]$) est relative à un type de surface S_i donné (figure II-f.1). Cette bijection évite toute confusion possible entre deux types de surfaces.

Imaginons maintenant que la distance entre la cible et le télémètre soit inconnue, toujours à incidence normale (figure II-f.2). Par exemple, plaçons-nous dans le cas où la tension $V_{outTRANS}$ observée est égale à 100mV. Cette amplitude peut correspondre à une surface plastifiée S_1 située à 1.55m, à une feuille de papier glacée S_2 située à 1.25m, à un mur peint S_3 situé à 0.95m ou à une surface poreuse S_4 à 0.6m. Cette fois-ci, il n'y a plus du tout bijection entre la tension de sortie $V_{outTRANS}$ et le type de surface S_i .

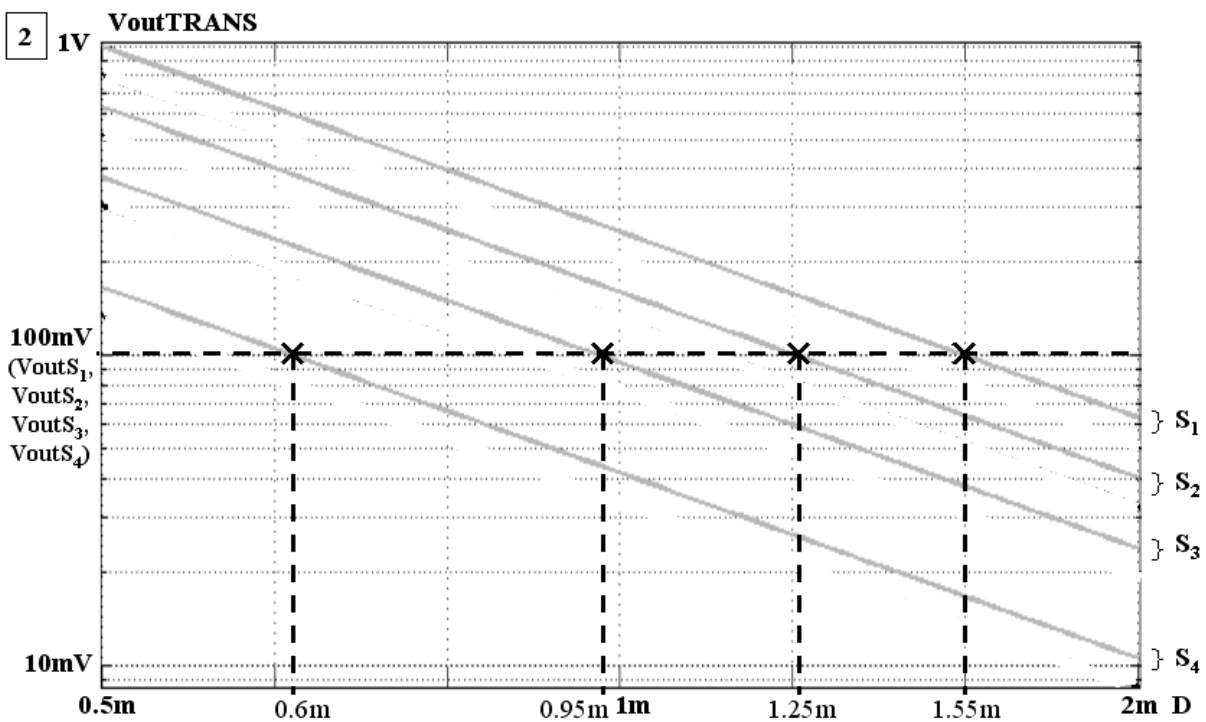
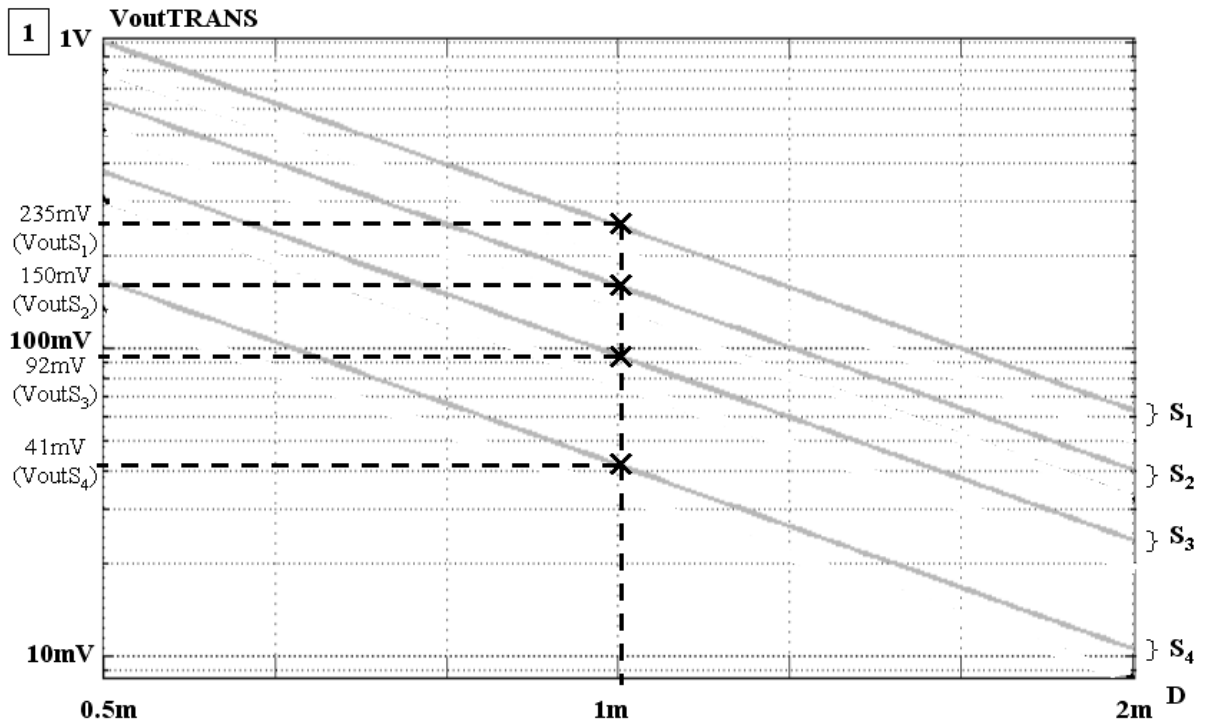


Figure II-f: Amplitude du signal de sortie de l'amplificateur transimpédance $V_{outTRANS}$ en fonction de la distance pour les 4 types de surfaces S_1 à S_4 à incidence normale.

1. Pour une valeur de distance fixée ($D=1m$), les 4 surfaces peuvent être facilement discriminées par l'intermédiaire de l'amplitude du signal $V_{outTRANS}$.
2. Lorsque la distance entre le télémètre et la cible n'est pas fixée, une même amplitude $V_{outTRANS}$ peut correspondre aux 4 types de surfaces testées, placées à des distances différentes.

Ainsi, aucune distinction entre ces quatre surfaces n'est possible en n'utilisant que l'information contenue dans l'amplitude du signal photoélectrique V_{out}^{TRANS} , à moins de limiter l'intervalle de mesure à un intervalle restreint ($[0.93m;1.05m]$ sur la figure II-g.1) et de fixer l'angle d'incidence à une valeur donnée (ici, à $\theta_0=0$). Dans ce cas restreint, chaque amplitude appartenant à l'intervalle $V_{out}^{S_i}$ ($i \in [1;4]$) est bijective par rapport au type de surface S_i testée (figure II-g.1).

Néanmoins, plus les surfaces testées sont proches en terme de valeur de coefficient de réflexion diffusant apparent ρ , plus l'intervalle de mesure est limité. De plus, cet intervalle est d'autant plus restreint si l'on tient compte d'une possible variation de l'angle d'incidence du faisceau laser par rapport à la cible, augmentant la probabilité de confusion entre deux surfaces. La figure II-g.2 montre un exemple de classification des quatre types de surfaces S_i pour un angle d'incidence variant de $-\pi/5$ à $+\pi/5$ (de -36° à 36°) dans le cas d'une surface cible lambertienne. Pour que ces quatre types de surfaces puissent être discriminés uniquement à l'aide du signal V_{out}^{TRANS} , il est nécessaire de limiter l'intervalle à une longueur de 5cm (par exemple, $[0.97m;1.02m]$). Plus l'angle d'incidence est grand, plus l'intervalle sur lequel la discrimination est possible est restreint.

La seconde problématique réside dans la possibilité de classer différents types de cibles à partir du signal V_{out}^{TRANS} , mais pour un angle d'incidence variable et sur un intervalle de mesure de l'ordre du mètre.

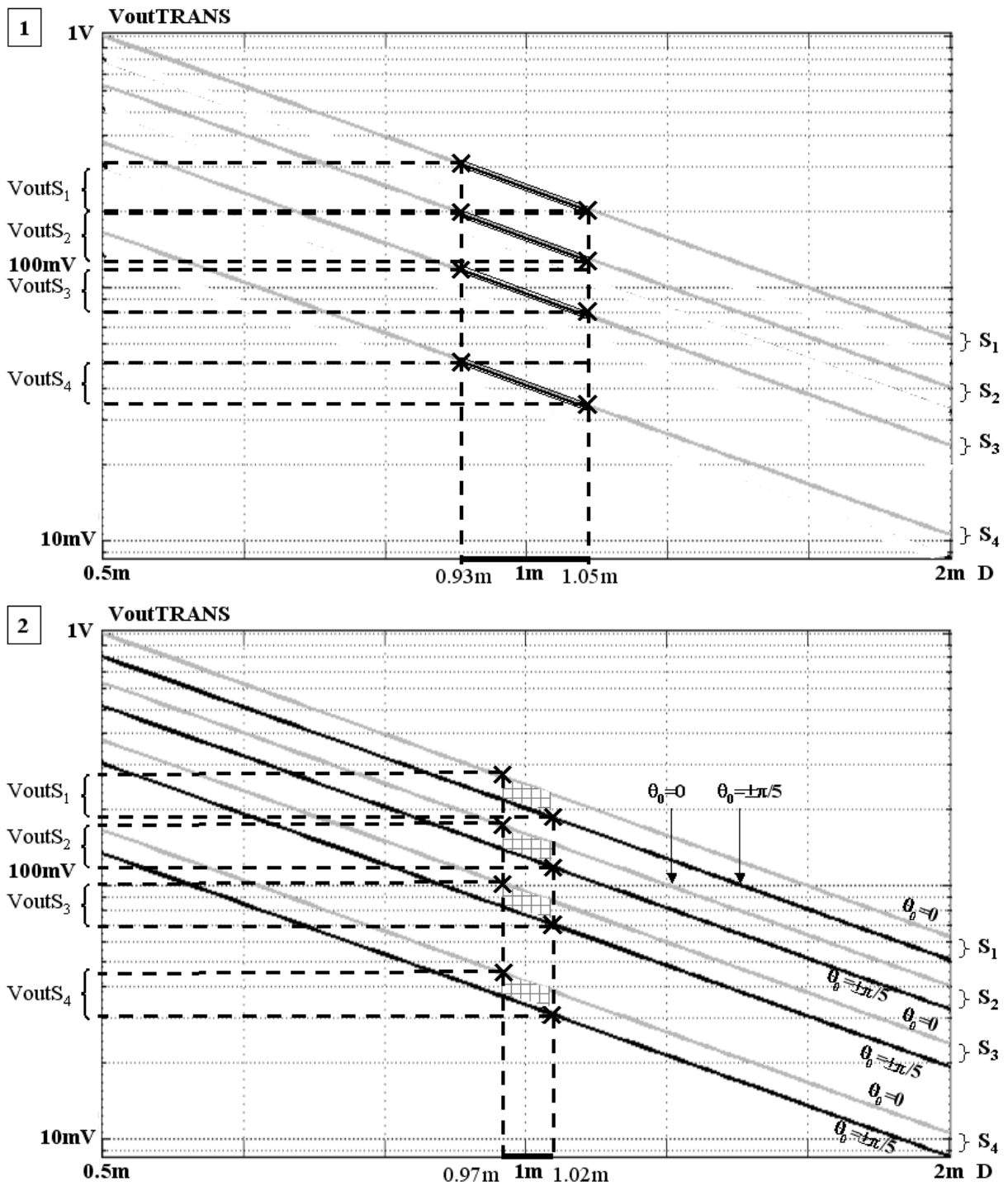


Figure II-g : 1. Restriction de l'intervalle de distance à l'intervalle [0.93m;1.05m], de manière à ce que chaque amplitude appartenant à l'intervalle $VoutS_i$ ($i \in [1;4]$) soit bijective par rapport au type de surface S_i testée pour un angle d'incidence nul.

2. Restriction de l'intervalle de distance à l'intervalle [0.97m;1.02m], de manière à ce que chaque amplitude appartenant à l'intervalle $VoutS_i$ ($i \in [1;4]$) soit bijective par rapport au type de surface S_i testée pour un angle d'incidence variant de $-\pi/5$ à $+\pi/5$. Les droites grises et noires représentent l'amplitude du signal $VoutTRANS$ pour un angle d'incidence respectivement égal à 0 et à $\pm\pi/5$.

II.3.2.4. Bilan

Ainsi, cette thèse repose sur deux problématiques associées à la télémétrie laser par déphasage :

- dans le cadre de la mesure de distance, la problématique réside dans la possibilité d'accroître l'intervalle de mesure de distance à partir du même dispositif, sans faire appel à plusieurs fréquences de modulation, tout en gardant une résolution élevée.

- dans le cas de la détection de surfaces, l'objectif est d'exploiter l'amplitude du signal photoélectrique issu du télémètre laser afin de discriminer plusieurs types de cibles ayant un coefficient de réflexion diffusant apparent différent. Ces surfaces doivent être classées sur un intervalle de mesure de l'ordre du mètre et pour un angle d'incidence variable, en corrélant l'amplitude du signal photoélectrique à la "rugosité" ou aux "aspérités" de la cible testée.

L'augmentation de l'intervalle de mesure et la détection de surfaces doivent être effectuées à partir de signaux déjà disponibles au sein du système télémétrique, sans faire appel à la mise en œuvre d'une autre mesure spécifique. De plus, la solution développée doit pouvoir être implantable de manière à réaliser ces deux opérations directement au sein de l'application télémétrique. Pour que le système puisse être embarqué, une basse consommation et une taille minimale seront des avantages supplémentaires.

Une solution aux deux problématiques explicitées ci-dessus et répondant aux différentes restrictions soulevées est proposée dans la partie suivante.

II.4. REFERENCES

- [Ailisto, 2002] : AILISTO H., HEIKINEN V., MITIKKA R., MYLLALA R., KOSTAMOVARA J., MANTYNIEMI A., KOSKINEN M., Scannerless imaging pulsed-laser range finding, *J. Opt. A: Pure Appl. Opt.* 4, 337-346 (2002).
- [Amann, 2001] : AMANN M.C., BOSCH T., LESCURE M., MYLLYLA R., ROUX M., Laser ranging: a critical review of usual techniques for distance measurement, *Opt. Eng.*, 40 (1), 1-10 (2001).
- [Baglio, 2002] : BAGLIO S., MUSCATO G., SAVALLI N., Tactile measuring systems for the recognition of unknown surfaces, *IEEE Transactions on Instrumentation and Measurement*, 51 (3), 522 – 531 (2002).
- [Bazin, 1996] : BAZIN G., JOURNET B., A new range finder based on FMCW-like method, *IEEE Instrumentation and Measurement Technology Conference, Brussels (Belgium)*, June 4-6, 1996.
- [Bosch, 1994] : BOSCH T., Modélisation de fonctions optoélectroniques appliquées à la télémétrie laser par déphasage, *Thèse du 22/10/1992, INSA, Toulouse*.
- [Bosch (a), 1995] : BOSCH T., LESCURE M., Selected Papers on Laser Distance Measurements, *SPIE Milestone Series, Vol. MS 115, SPIE Optical Engineering Press, Bellingham, Washington, USA*, 1995.
- [Bosch (b), 1995] : BOSCH T., LESCURE M., Experimental determination of the useful reflexion coefficient of non-cooperative targets for a time-of-flight laser rangefinder, *Optical Review*, Vol.2, No.4, 1995, 289-291.
- [Bosch, 1997] : BOSCH T., LESCURE M., Error Analysis of 1-10 m Laser Phase-Shift Rangefinder, *IEEE Transaction on Instrumentation and Measurement*, vol 46, n°6, pp1224-1228, Dec. 1997.
- [Carullo, 1996] : CARULLO A., FERRARIS F., GRAZIANI S., GRIMALDI U., PARVIS M., "Ultrasonic distance sensor improvement using a two-level neural-network", *Instrumentation and Measurement, IEEE Transactions on*, Volume 45, Issue 2, April 1996, Page(s):677 – 682.
- [Castagnet, 2006] : CASTAGNET D., TAP-BETEILLE H., LESCURE M., APD-based heterodyne optical head of a phase-shift laser rangefinder, *Opt. Eng.*, 40 (4), 43003_1 – 43003_7 (2006).
- [CEI 825] : Norme de la Commission Electrotechnique Internationale, *Sécurité des Appareils à Laser, Classification des Matériels, Prescription et Guide de l'Utilisateur*.
- [Cinquin, 1995] : CINQUIN P., BAINVILLE E., BARBE C., BITTAR E., BOUCHARD V. et al., Computer assisted medical interventions, *Engineering in Medicine and Biology Magazine, IEEE*, 14 (3), 254 – 263 (1995).
- [De Almeida, 2000] : DE ALMEIDA A.R., FREIRE E.O., RONNOW C.A., VIANNA J.E.S., ROSI R.M., Neural network-based geometric references recognition applied to ultrasound echo signals, *Circuits and Systems, Proceedings of the 43rd IEEE Midwest Symposium on*, Vol. 3, 1344 - 1347 (2000).
- [Dietmayer, 2001] : DIETMAYER K., SPARBERT J., STRELLER D., Model Based Classification and Object Tracking in Traffic Scenes from Range-Images, *IEEE Symposium Intelligent Vehicles 2001*.

- [Donati, 2004] : DONATI S., *Electro-optical instrumentation, sensing and measuring with lasers*, Prentice Hall, 2004.
- [Dorsch, 1994] : DORSCH R.G., HAUSLER G., HERMANN J.M., Laser Triangulation : fundamental uncertainty in distance measurement, *Applied Optics*, vol 33 (7), pp1306-1314.
- [Dupuy,2001] : DUPUY D., LESCURE M., COUSINEAU M., Télémètre laser FMCW comportant une ligne à retard et une détection non linéaire, *Colloque Instrumentation, Interdisciplinarité et Innovation, C2I'2001*, Paris, janvier 2001.
- [Dupuy, 2002(a)] : DUPUY D., LESCURE M., TAP-BETEILLE H., Analysis of an avalanche photodiode used as an optoelectronic mixer for a frequency modulated continuous wave laser range finder, *Journal of Optics A : Pure and Applied Optics*, Vol. 4, pp. S332-336 (2002).
- [Dupuy, 2002(a)] : DUPUY D., LESCURE M., Improvement of the FMCW Laser Range-Finder by an APD Working as an Optoelectronic Mixer, *IEEE Transactions on Instrumentation and Measurement*, Vol. 51, n°5, Oct 2002.
- [Dziadowiec, 1983] : DZIADOWIEC A., LESCURE M., BOUCHER J., Comparaison de différents dispositifs de mesure de déphasage en haute fréquence, *L'onde électrique*, 1983, vol. 63, n°6-7.
- [Dziadowiec, 1984] : DZIADOWIEC A., LESCURE M., BOUCHER J., A heterodyne low level signal phase meter operating over 1 MHz to 300 MHz, *IEEE Transactions on Instrumentation and Measurement*, Vol IM 33, n°1, 03/1984.
- [Dziadowiec, 1995] : DZIADOWIEC A., LESCURE M., Fonctions à Amplificateurs Opérationnels, *Editions Eyrolles*, 1995.
- [Fiorillo, 2002] : FIORILLO A.S., D'ANGELO G., Echo signals processing with neural network in bat-like sonars based on PVDF, *Ultrasonics Symposium, 2002. Proceedings. 2002 IEEE*, Volume 1, 8-11 Oct. 2002, Page(s):781 – 784, vol.1.
- [Fukui, 1997] : FUKUI H., TAKAGI J., MURATA Y., TAKEUSHI M., An image processing method to detect road surface condition using optical spatial frequency, *Intelligent Transportation System, IEEE Conference on*, 1005 – 1009 (1997).
- [Giubbolini, 1999] : GIUBBOLINI L., A Microwave Imaging Radar in the near Field for Anti-Collision (MIRANDA), *IEEE Trans. on Microwaves Theory and Techniques*, Vol. 47 (1999).
- [Häusler, 1997] : HAUSLER G., KREIPEL S. ; LAMPALZER R., SCHIELZETH, A., SPELLENBERG B., New Range Sensors at the Physical Limit of Measuring Uncertainty, *Proc. Odimap'97, EOS Optoelectronic Distance/Displacement Measurements and Applications. Nantes*, July 1997.
- [Hueber, 2000] : HUEBER G., OSTERMANN T., BAUERNFEIND T., RACSHHOFER R., HAGELAUER R., New approach of ultrasonic distance measurement technique in robot applications, *Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on*, Volume 3, 21-25 Aug. 2000, Page(s):2066 – 2069, vol.3.
- [Joodaki, 2001] : JOODAKI M., KOMPA G., AHMADI V., FARSHI M.K.M., Using neural networks for high resolution distance measurements in pulsed laser radar, *Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE*, Volume 2, 21-23 May 2001, Page(s):1242 – 1246, vol.2.

- [Kaisto, 1983] : KAISTO I., KOSTAMOVAARA J., MANNINEN M., MYLLYLÄ R., Optical range finder for 1.5-10-m distances, *Applied Optics*, Vol. 22, Issue 20, pp. 3258.
- [Koskinen, 1992] : KOSKINEN M., KOSTAMOVAARA J., MYLLYLÄ R., Comparison of continuous-wave and pulsed time-of-flight laser range-finding techniques, *Optics, Illumination, and Image Sensing for Machine Vision VI, Proceedings of SPIE*, Volume 1614, March 1992, pp. 296-305.
- [Lang, 1997] : LANG T., GENON-CATALOT, D., DANDREA, P., SHANEN-DUPOINT, I., BENECH, P., Integrated optic displacement sensor with four quadrature phase shifted output signals, *Proc. Odimap'97, EOS Optoelectronic Distance/Displacement Measurements and Applications. Nantes, July 1997*.
- [Liessem, 2002] : LIESSEM A., GRIMPE F., OESTERLEIN L., State-of the art quality control during production of linepipes, *Proceedings of 4th International Pipeline Conference, Sept. 29- Oct. 3, 2002, Calgary, Alberta, Canada*.
- [Lin, 1983] : LIN J.C., CHI Z.C., Accuracy Analysis of a Laser Camera Based 3D Measurement System, *Sensory Controls, Proc. SPIE*, Vol. 449, pp.158-170 (1983).
- [Määttä, 1993] : MAATTA K., KOSTAMOVAARA J., MYLLYLÄ R., Profiling of hot surfaces by pulsed time-of-flight laser range finder techniques, *Applied Optics*, Vol. 32, No 27, pp. 5334 - 5347, September 1993.
- [Neelamegan, 2003] : NEELAMEGAN P., RAJENDRAN A., "Classification of surface roughness and distance measurement using artificial neural network", *Instrumentation Science and Technology*, 31 (4), 417-423 (2003).
- [Nerguizian, 2004] : NERGUIZIAN V., RAFAF M., PACKIRISAMY M., STIHARU I., Intelligent sensors for ultra violet detection, *Industrial Technology, 2004. IEEE ICIT '04. 2004 IEEE International Conference on*, Volume 2, 8-10 Dec. 2004, Page(s):991 – 995.
- [Norris, 1999] : NORRIS JR V.J.; EVANS R.S., CURRIE D.G., Performance comparison of visual, infrared, and ultraviolet sensors for landing aircraft in fog, *Proc. of the 18th Digital Avionics Systems Conference*, vol.1, 24-29 Oct. 1999, pp:5.D.2-1 - 5.D.2-19.
- [Okada, 2005] : OKADA K., OGURA T., HANEDA A., INABA M, Autonomous 3D walking system for a humanoid robot based on visual step recognition and 3D foot step planner, *Robotics and Automation, Proceedings of the 2005 IEEE International Conference on*, 623 – 628 (2005).
- [Omura, 1999] : OMURA Y., FUNABIKI S., TANAKA T., A monocular vision-based position sensor using neural networks for automated vehicle following, *Power Electronics and Drive Systems, 1999. PEDS '99. Proceedings of the IEEE 1999 International Conference on*, Volume 1, 27-29 July 1999, Page(s):388 – 393, vol.1.
- [Ozdemir, 2001] : OZDEMIR S.K, SHINOHARA S., ITO S., TAKAMIYA S., YOSHIDA H., Compact optical instrument for surface classification using self-mixing interference in a laser diode, *Opt. Eng.*, 40 (1), 38-43 (2001).
- [Pauchard, 2000] : PAUCHARD A.R., MANIC D., FLANAGAN A., BESSE P.A., POPOVIC R.S., A method for spark rejection in ultraviolet flame detectors, *Industrial Electronics, IEEE Transactions on*, Volume 47, Issue 1, Feb. 2000 Page(s):168 – 174.
- [Payne, 1992] : PAYNE J.M., PARKER D., BRADLEY R.F., Rangefinder with fast multiple range capability, *Review of Scientific Instruments*, Vol. 63(6), pp. 3311-3316 (june 1992).

- [Poujouly, 2000] : POUJOULY S., JOURNET B., Laser range finding by phase-shift measurement: moving towards smart systems, *Proceedings of the SPIE Machine Vision and Three-Dimensional Imaging Systems for Inspection and Metrology*, 4189, 152-160 (2000).
- [Rouveure, 2003] : ROUVEURE R., FAURE P., MONOD M.-O., A multi-layer feed-forward perceptron for microwave signals processing, *Geoscience and Remote Sensing Symposium, 2003, IGARSS '03. Proceedings. 2003 IEEE International*, Volume 6, 21-25 July 2003, Page(s):3519 – 3521, vol.6.
- [Salvadé, 1997] : SALVADE Y., ZIMMERMANN E, DANDLIKER R., Limitations of interferometry due to frequency fluctuations of laser diodes, *Proc. Odimap'97, EOS Optoelectronic Distance/Displacement Measurements and Applications. Nantes, July 1997.*
- [Sarabandi, 1997] : SARABANDI K., LI E.S., NASHASHIBI A., Modeling and measurements of scattering from road surfaces at millimeter-wave frequencies, *Antennas and Propagation, IEEE Transactions on*, 45 (11), 1679 – 1688 (1997).
- [Tiedeke, 1990] : TIEDEKE J., SCHABEL P., RILLE E., Vehicle distance sensor using a segmented IR laser beam, *Vehicular Technology Conference, 1990 IEEE 40th, 6-9 May 1990*, Page(s):107 – 112.
- [Vossiek, 1994] : VOSSIEK M., MAGORI V., ERMERT H., An Ultrasonic Multielement Sensor System for Position Invariant Object Identification, *IEEE Ultrasonics Symposium* (1994).
- [Williams, 1986] : WILLIAMS C.C., WICKRAMASINGHE H;K., Optical ranging by wavelength multiplexed interferometry, *Journal of applied Physics*, Vol 60, n°6, Sept 1986, pp1900-1903.
- [Williams, 1995] : WILLIAMS J.D., Advanced technologies for perimeter intrusion detection sensors, *Security and Detection, European Convention on*, 133 – 137 (1995).
- [Yamada, 1993] : YAMADA O., TAKAHASHI H., KIMURA M., NAITOH H., SANO R., CO₂ laser radar 3D vision sensor for a mobile robot, *Advanced robotics*, 1993, vol. 7, n°3, pp. 251-260.
- [Yoda, 1995] : YODA S., OKABE H., TAKAGI J., YAMASHITA T., Road surface recognition sensor using an optical spatial filter, *Proceedings on Intelligent Vehicles 95 Symposium*, 253-257 (1995).

Chapitre III

Utilisation de Réseaux de Neurones dans le domaine de la télémétrie laser par déphasage

Ce chapitre est centré sur l'application de réseaux de neurones dans deux domaines relatifs à la télémétrie laser par déphasage : la mesure de distance et la classification de surfaces.

III.1. INTRODUCTION

III.1.1. Mesure de distance

Rappelons brièvement la problématique. Pour une cible donnée, avec un angle d'incidence donnée, il n'est pas possible, avec le télémètre laser par déphasage seul, d'avoir une très grande résolution sur la mesure de la distance sur un grand intervalle de distance. L'objectif est de garder le même système de mesure permettant d'obtenir une grande résolution et d'y ajouter un système annexe dont le but est de lever l'indétermination modulo 2π introduite par la mesure du déphasage et ainsi accroître l'intervalle de mesure.

Les deux principales solutions qui s'offrent à nous pour augmenter l'intervalle de mesure du télémètre laser par déphasage sont :

- la modulation de la diode laser à deux fréquences très différentes, l'une faible de manière à évaluer grossièrement la valeur de la distance et l'autre élevée de manière à avoir une résolution importante ([Journet, 2002]).

- un traitement du signal adapté à partir d'un des signaux issus du système télémétrique, de manière à évaluer un ordre de grandeur de la distance sans indétermination sur sa valeur et sans faire appel à une mesure supplémentaire.

Notre objectif est d'accroître l'intervalle de mesure de distance à partir du même dispositif tout en gardant une résolution élevée, sans complexifier la structure globale. Nous proposons donc d'effectuer un traitement adéquat d'un ou de plusieurs signaux issus du télémètre afin

d'évaluer la distance avec une précision modeste, de façon à lever l'indétermination modulo 2π introduite par la mesure de la phase. La précision nécessaire sur l'évaluation de la valeur de la distance par le réseau de neurones est déduite de deux valeurs successives de distance correspondant à un même déphasage, soit la valeur de l'intervalle maximum de mesure sans indécision Λ . Par exemple, pour une fréquence de modulation de 166MHz, l'intervalle Λ est égal à 0.9m. Ainsi, la différence de phase observée en sortie du télémètre laser par déphasage sera identique pour toute distance modulo Λ . Pour distinguer deux valeurs de distance, il suffit d'évaluer la distance à $\Lambda/2$ près, soit 0.45m pour notre exemple.

Pour obtenir une approximation de la valeur de la distance, il faut pouvoir traiter un signal contenant l'information "distance" provenant de la partie réception du télémètre, avant qu'il n'ait été multiplié avec le signal de référence. Le signal le plus simple à traiter est celui qui est envoyé au phasemètre après hétérodynage et filtrage, appelé V_{outBP} (figure II-c). L'amplitude de ce signal, proportionnelle à celle de la sortie de l'amplificateur transimpédance $V_{outTRANS}$, suit la relation :

$$V_{outBP} = K'_1 \frac{\rho \cos(\theta_0)}{D^2} \quad (III-1)$$

où K'_1 est un coefficient de proportionnalité relatif à la liaison optoélectronique et aux fonctions de transfert du mélangeur et du filtre passe-bande. Rappelons que K'_1 est un terme constant indépendant de la cible et de sa position dans l'espace. La surface étant supposée invariante et l'angle d'incidence constant, le signal V_{outBP} ne varie qu'en fonction de l'inverse du carré de la distance.

L'objectif est donc d'évaluer la valeur de la distance à partir de l'amplitude du signal photoélectrique à la fréquence f_{IF} . La majeure partie des systèmes de traitement du signal font appel à des systèmes numériques. La méthode la plus simple pour extraire la valeur de la distance du signal alternatif issu du télémètre consiste alors à réaliser les étapes suivantes :

- échantillonner le signal à l'aide d'un convertisseur analogique-numérique,
- retirer la valeur moyenne ou la tension de décalage,
- extraire la valeur maximale,
- effectuer l'inversion et la racine carrée.

Une autre solution consiste à extraire l'information "distance" du signal V_{outBP} à l'aide d'un système simple, constitué d'une entrée (reliée à la sortie du filtre passe-bande) et d'une sortie (dont la variation en fonction de la distance est simple à interpréter). Les valeurs expérimentales de l'amplitude du signal V_{outBP} mesurées à une distance D sont facilement

accessibles et en grand nombre, sur un intervalle limité par les défauts de défocalisation de l'optique de réception à courte distance et le niveau de bruit électronique à partir de quelques mètres. Le système adéquat semble donc être un approximateur de fonction réalisant une fonction g de telle sorte que la donnée issue de $g(V_{outBP})$ permette de déterminer une valeur approximative de la distance D facilement accessible et exploitable. Seulement, la relation (III-1) issue du bilan de liaison est idéale et ne tient pas compte des défauts d'alignement optiques, ni des problèmes de défocalisation, de variation de température, de vieillissement, de diaphonie, etc. Comme cette relation varie en fonction de chacun de ces paramètres expérimentaux difficilement accessibles, établir un approximateur de fonctions quelconque réalisant une fonction g mènerait à une donnée $g(V_{outBP})$ difficilement exploitable.

La solution provient d'un système d'approximation de fonctions capable d'être entraîné dans différentes conditions expérimentales et d'effectuer la fonction souhaitée de manière à ce que la donnée $g(V_{outBP})$ soit directement corrélée à l'information "distance". Nous sommes donc dans une situation où l'utilisation d'un réseau de neurones est toute indiquée de manière à effectuer l'approximation de la fonction g ([Gatet, 2006]).

III.1.2. Classification de surfaces

Rappelons brièvement la problématique. La classification de surfaces à l'aide de la seule amplitude du signal $V_{outTRANS}$ (ou V_{outBP}) s'avère très peu intéressante du fait de la limitation de l'intervalle de mesure, même pour un angle d'incidence fixe et pour des surfaces dont les coefficients de réflexion diffusant sont très différents. L'objectif est donc de pouvoir discriminer différentes cibles sur un intervalle de mesure plus important et pour un angle d'incidence variable à partir de signaux issus du télémètre laser sans en modifier la structure, afin de conserver une résolution élevée.

L'idée est d'exploiter un deuxième signal issu du système de mesure. Le premier signal (V_{outBP}) permettant d'obtenir une information relative au type de surface testé, ce deuxième signal aurait pour utilité d'évaluer la distance entre le télémètre et la cible, de manière à accroître le domaine de mesure sur lequel la discrimination est possible. La donnée la plus accessible relative à la distance est le signal de sortie du télémètre laser par déphasage $V_{outRANG}$ (figure II-c). Ce signal est directement proportionnel à la distance sur un intervalle de distance limité par l'indécision modulo 2π de la mesure de phase. La précision de la mesure de la distance n'étant pas primordiale, la fréquence de modulation f_{RF} de la diode

laser sera étalonnée de manière à ce que la plage de mesure sur laquelle sera effectuée la détection de surfaces soit incluse dans l'intervalle de mesure sans indécision Λ .

A l'aide de ces deux données, un système discriminant est capable de différencier plusieurs types de surfaces pour un angle d'incidence θ_0 variable. La plage de variation de l'angle d'incidence est liée à la proximité des surfaces en terme de coefficient de réflexion diffusant apparent. En effet, plus les valeurs de ρ sont éloignées les unes des autres, plus la variation de l'angle d'incidence peut être importante. La valeur de l'angle d'incidence maximum tolérable sera discutée plus loin. Appelons K_2 le coefficient de proportionnalité entre l'amplitude du signal de sortie $V_{outRANG}$ du télémètre laser par déphasage et la distance D .

$$V_{outRANG} = K_2 D \quad (III-2)$$

Plusieurs systèmes de traitement des signaux sont capables de réaliser de la discrimination de surfaces à partir de deux jeux de données. Le système numérique le plus simple consiste à effectuer les mêmes opérations de numérisation que pour le système d'accroissement de l'intervalle de mesure de distance, puis à corrélérer les informations issues des deux signaux pour déduire le type de surfaces à partir de jeux de valeurs mémorisés. Ce système n'est valable que si la relation (III-1) est vérifiée.

Là encore, il est nécessaire de faire appel à un système capable d'être entraîné à classer les différentes surfaces en fonction d'un nombre important de données expérimentales prélevées directement sur le système télémétrique, notamment du point de vue de la corrélation entre le type de surface et l'amplitude des signaux de sortie V_{outBP} et $V_{outRANG}$. Nous sommes donc de nouveau dans un cas adéquat d'utilisation d'un réseau de neurones pour effectuer la discrimination des différentes surfaces à partir de ces deux signaux ([Gatet, 2007]).

Les deux parties suivantes expliquent la manière dont les deux réseaux de neurones ont été mis en œuvre pour achever leur tâche. Le but final est de faire en sorte de ne concevoir qu'un seul réseau de neurones, capable de réaliser l'une ou l'autre tâche en fonction de l'application dans laquelle il sera placé, de manière à simplifier une intégration future du système complet.

III.2. MESURE DE DISTANCE

III.2.1. Principe

L'objectif est d'utiliser l'amplitude du signal photoélectrique alternatif V_{outBP} et les facultés d'approximation du réseau de neurones à partir de données expérimentales de manière à évaluer avec suffisamment de précision la valeur de la distance D pour lever l'indétermination modulo 2π introduite par la mesure du déphasage. En corrélant cette information à celle issue de la mesure de la phase, le système global est capable de donner une valeur de la distance avec la même précision tout en ayant augmenté l'intervalle de mesure. Cet intervalle, limité à $[0.5m ; 1.4m]$ par la mesure directe du déphasage, est étendu à $[0.5m ; 3.5m]$, intervalle de mesure limite du télémètre utilisé. Les valeurs 0.5m et 3.5m sont respectivement liées aux défauts de défocalisation de l'optique de réception et au mauvais rapport signal sur bruit. Ainsi, l'intervalle de mesure a été augmenté d'un facteur 3.3 avec la même résolution sans avoir modifié la structure du télémètre ni fait appel à un système de mesure supplémentaire ([Gatet, 2006]).

La fonction d'approximation est choisie de telle sorte que le signal de sortie V_{outNN} du réseau de neurones soit linéaire avec la distance :

$$V_{outNN} = K_0 D \quad (III-3)$$

Si l'on appelle g la fonction réalisée par le réseau et V_{inNN} le signal d'entrée continu du réseau proportionnel à l'amplitude du signal photoélectrique alternatif V_{outBP} , on obtient la relation suivante :

$$V_{outNN} = g(V_{inNN}) \quad (III-4)$$

D'un point de vue structurel, le réseau possède une entrée et une sortie dont la variation doit être proportionnelle à la distance D . Le nombre de neurones placés sur la couche cachée sera explicité plus loin.

Pour une cible relative à un mur peint, les courbes théoriques et expérimentales relatives à la tension d'entrée V_{inNN} du réseau sont représentées sur la figure III-a en échelle log-log en fonction de la distance, sur l'intervalle de mesure $[0.5m ; 3.5m]$, pour des conditions expérimentales de laboratoire.

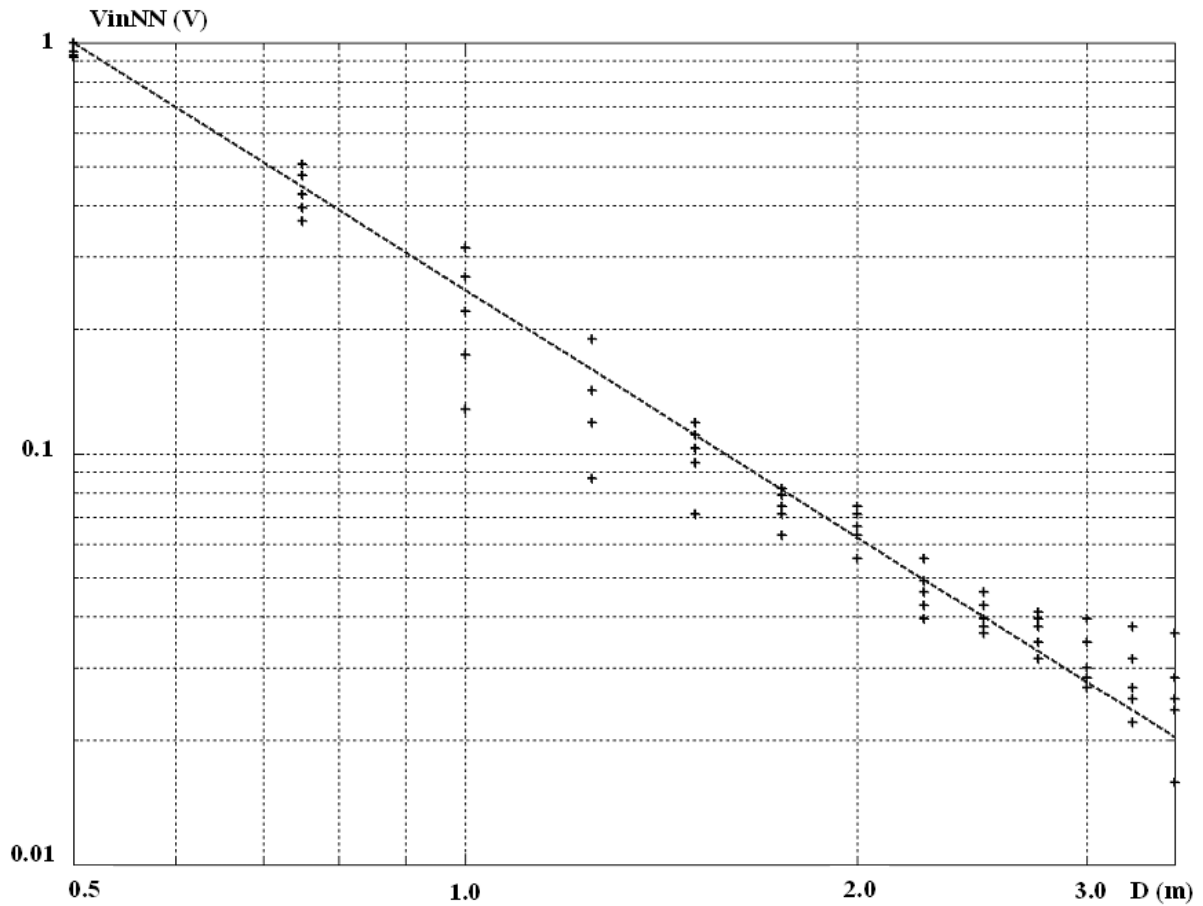


Figure III-a : Variations de l'amplitude du signal photoélectrique en fonction de la distance. La caractéristique idéale proportionnelle à l'inverse du carré de la distance est tracée en pointillés ; l'ensemble des points expérimentaux est représenté par des croix. Pour chaque valeur de distance, 5 mesures ont été réalisées dans des conditions expérimentales différentes.

La tension de sortie V_{outNN} du réseau de neurones doit être telle que :

$$V_{outNN} = g(V_{inNN}) = g\left(\frac{K'' \cos(\theta_0)}{D^2}\right) = K_0 D \quad (III-5)$$

La fonction que le réseau de neurones doit réaliser de manière à ce que sa sortie soit proportionnelle à la distance D est donc :

$$V_{outNN} = g(V_{inNN}) = \frac{K_{01}}{\sqrt{V_{inNN}}} \quad (III-6)$$

Le système complet et la fonction g que doit approcher le réseau pour que sa sortie soit linéaire par rapport à la distance sont représentées sur la figure III-b.

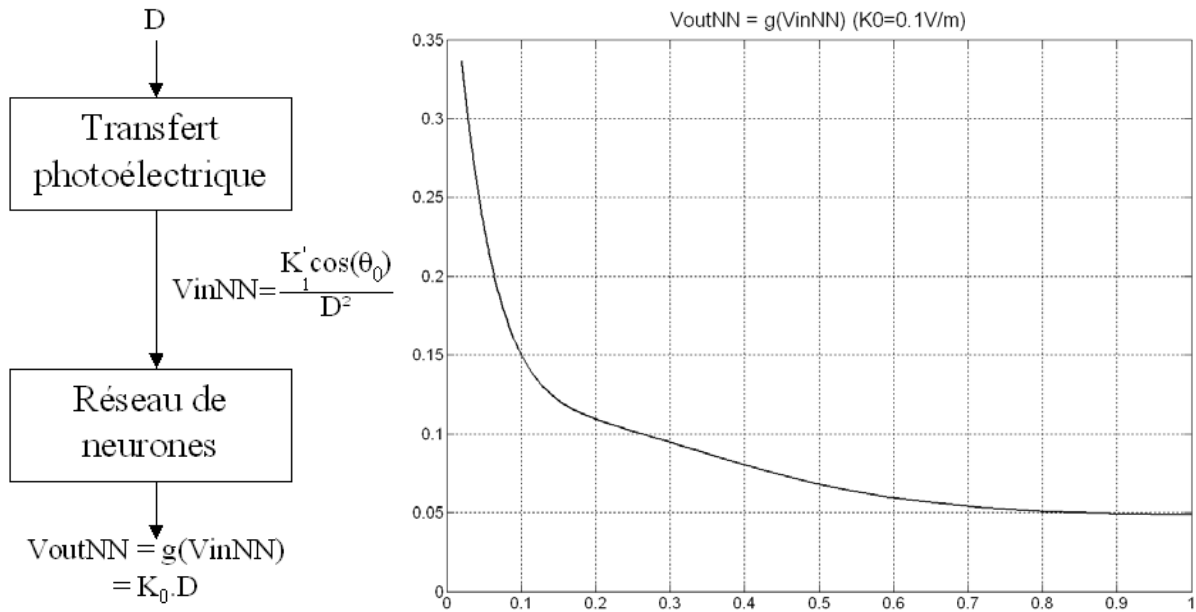


Figure III-b : Fonction g que doit réaliser le réseau de neurones de manière à ce que sa sortie soit proportionnelle à la distance (L'angle d'incidence θ_0 est considéré constant et K_0 est choisi égal à $0.1V.m^{-1}$).

Pour simplifier le traitement du signal que le réseau de neurones doit effectuer, le signal photoélectrique V_{outBP} de fréquence f_{IF} est converti en signal continu à l'aide d'un simple convertisseur AC-DC. Ce signal continu constituera le signal d'entrée du réseau de neurones. Le système complet est représenté figure III-c.

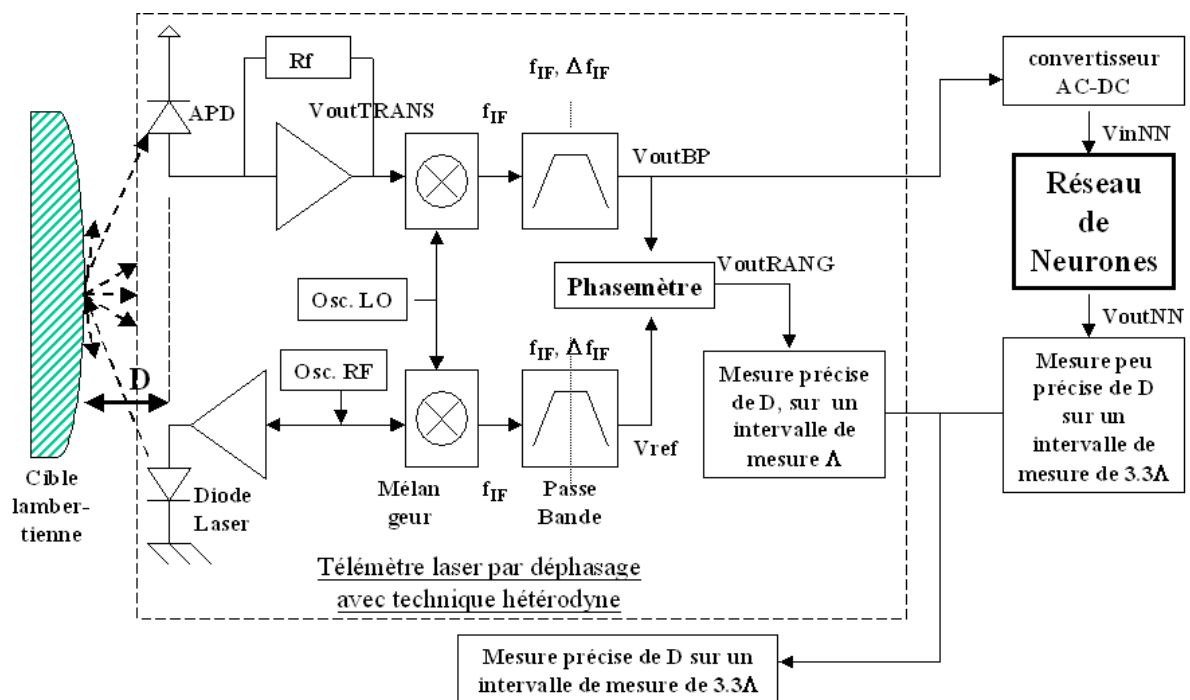


Figure III-c : Système complet de levée d'indétermination sur la mesure de distance issue d'un télémètre laser par déphasage à l'aide d'un réseau de neurones.

III.2.2. Simulations comportementales sous Matlab

Nous avons établi précédemment que le réseau de neurones non bouclé à apprentissage supervisé, utilisé pour effectuer la levée de l'indétermination sur la mesure de distance, est constitué d'une entrée et d'un neurone de sortie. Reste à déterminer le nombre de couches cachées et de neurones par couches.

La sortie du réseau de type perceptron multi-couches peut être considérée comme la somme de fonctions "tangente hyperbolique" pondérées par les poids w_{2j} . Un seul neurone sur la couche cachée revient à chercher à approcher la fonction (III-6) par l'intermédiaire de la relation $V_{outNN} = b_2 + w_{2j} \cdot \tanh(w_{1j} V_{inNN} + b_{1j})$ pour une tension V_{inNN} variant de 20mV à 1V. Au vu de la caractéristique de la fonction tangente hyperbolique, il est impossible de faire en sorte qu'elle se superpose avec la courbe tracée figure III-b avec un degré de précision suffisant. Le constat est identique pour un réseau à deux neurones cachés. C'est la raison pour laquelle le nombre de neurones cachés a été initialisé à 3 (figure III-d).

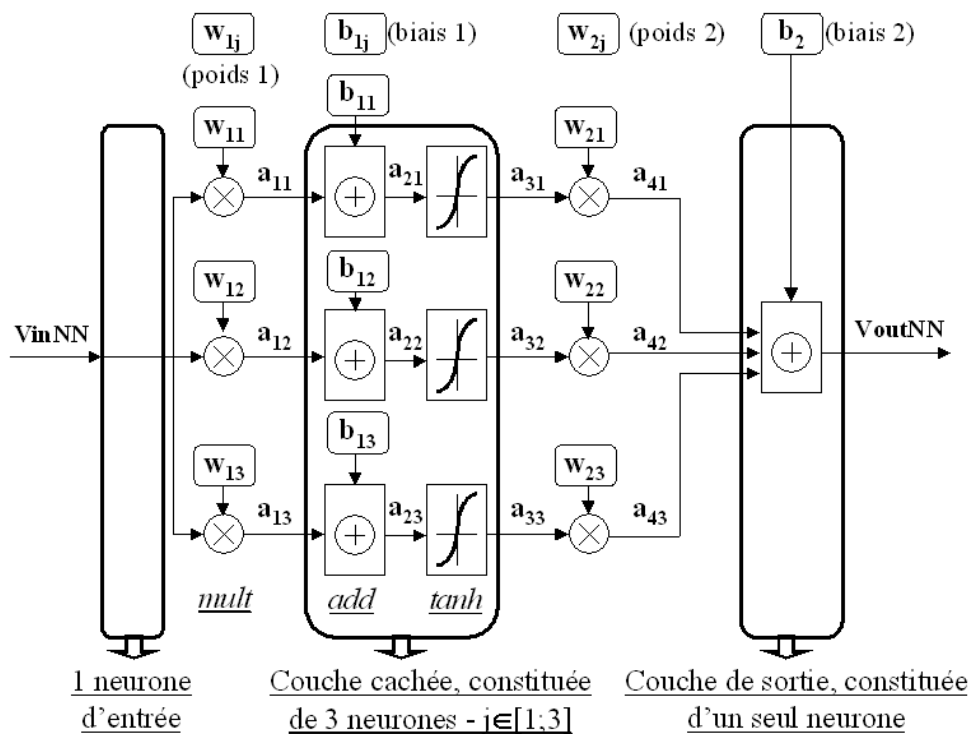


Figure III-d : Structure du réseau de neurones réalisant la fonction g .

Une fois le nombre de neurones cachés fixé, la procédure explicitée figure I-e est utilisée. Un jeu de valeurs de distance est déterminé dans l'intervalle [0.5m ; 3.5m] de manière à constituer l'ensemble d'apprentissage. Le choix de ces valeurs se porte sur deux critères : le nombre de valeurs parmi l'ensemble des valeurs mesurables accessibles et leur répartition sur l'intervalle de distance. Il est évident qu'un nombre insuffisant ou une mauvaise répartition des

valeurs d'apprentissage (par exemple un choix de trois valeurs de distance seulement ou une répartition de l'ensemble des valeurs de distance entre 2.5m et 3.5m) conduirait à une mauvaise approximation de la fonction pour des valeurs autres que celles appartenant à l'ensemble d'apprentissage. Un raisonnement rapide conduirait à prendre l'ensemble des valeurs accessibles et d'entraîner le réseau avec ces valeurs. Le risque est de constater une "non-convergence" des paramètres du réseau lors de la phase d'apprentissage. De plus, le temps d'entraînement associé à un réseau de neurones est d'autant plus important que le nombre de valeurs d'apprentissage est grand.

La tension d'entrée du réseau de neurones suivant une loi inversement proportionnelle au carré de la distance (figure III-e), les valeurs de distance de l'ensemble d'apprentissage ont été choisies de manière à ce qu'elles soient uniformément réparties en échelle logarithmique. En effet, l'amplitude du signal de sortie V_{outBP} subit des variations d'amplitude plus importantes pour des distances faibles. Il faut donc faire en sorte que le nombre de valeurs d'apprentissage soit plus important pour les faibles distances.

Le nombre de valeurs de distance a été choisi égal à 9 (figure III-e), de manière à ce que les temps d'apprentissage ne soient pas excessifs et qu'un nombre important de valeurs de distance ait été conservé pour le test du réseau une fois la phase d'apprentissage terminée.

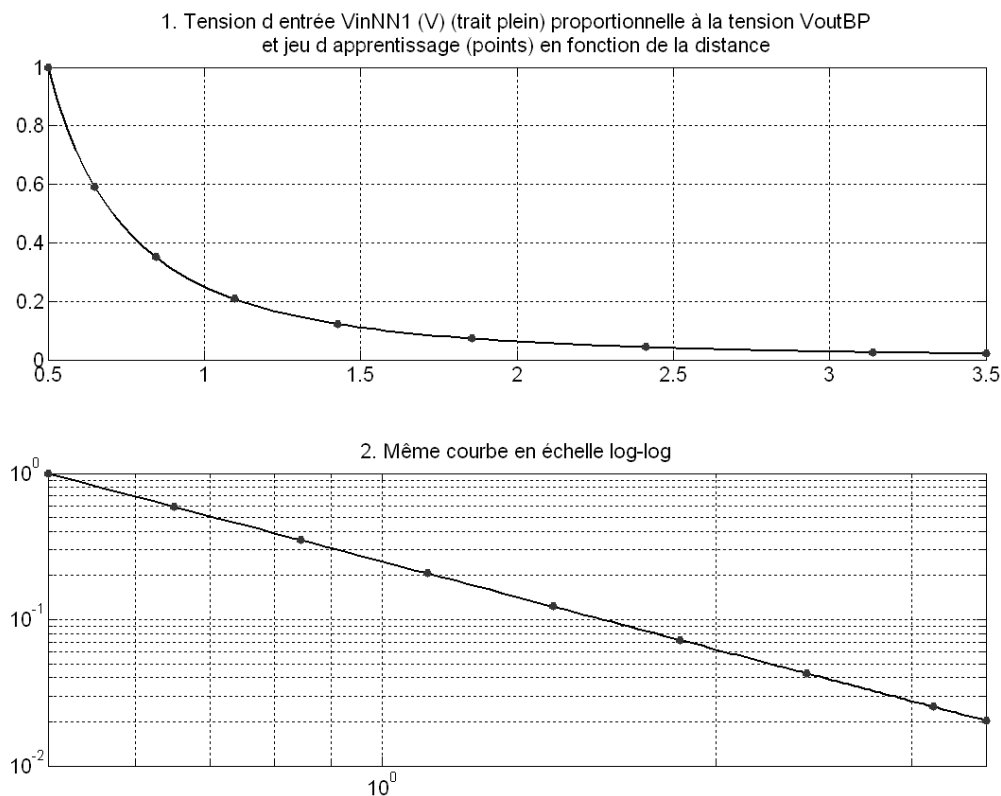


Figure III-e : Courbes représentant la répartition des valeurs des amplitudes d'entrée du réseau de neurones, fonction des mesures de distance choisies dans l'intervalle [0.5m;3.5m].

Remarque : en choisissant 9 valeurs de distance appartenant à l'ensemble d'apprentissage, le nombre de paramètres du réseau (10) devient supérieur au nombre de données. Dans ce cas de figure, il existe une infinité de solutions à ce problème. De plus, d'autres approximateurs plus simples que les réseaux de neurones et facilement intégrables, tels que les approximateurs polynomiaux, permettent de résoudre ce type de problème (figure III-f).

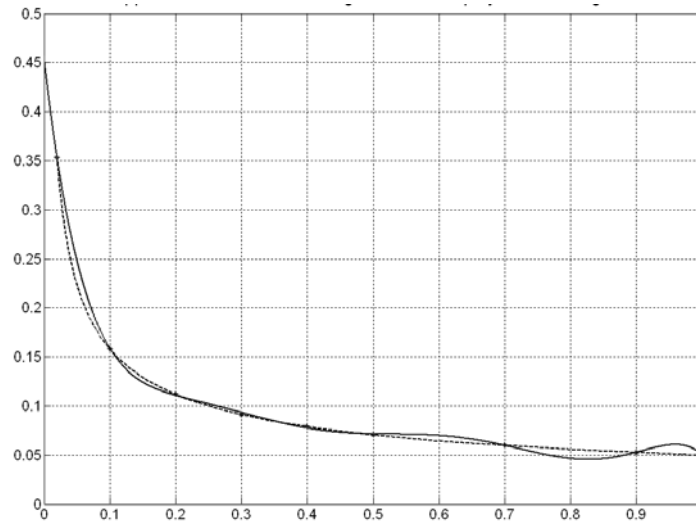


Figure III-f : Approximation de la fonction g idéale (pointillés) à partir des 9 valeurs de distance (points) à l'aide d'un polynôme de degré 7 (trait plein). L'erreur commise est inférieure à 10mV.

Néanmoins, le but de la thèse est de trouver une méthode commune aux deux problématiques, celle sur la levée d'indétermination sur la mesure de distance et celle sur la classification de surfaces. Concernant l'application relative à la classification de surfaces, le nombre de données expérimentales est égal à 96 (paragraphe II-3.2).

De plus, cette thèse est une première étape avant la réalisation d'un système implanté bouclé capable de prendre en compte les variations des conditions expérimentales par mise à jour des paramètres à l'aide d'un système numérique. Le réseau de neurones a été choisi car en plus de répondre aux problématiques relatives aux deux applications, il permet un apprentissage adaptatif en fonction des conditions expérimentales.

L'idée de cette validation comportementale peut se résumer au fait de vouloir simplement vérifier que le réseau est capable d'approcher une fonction avec une précision suffisante pour lever l'indétermination due à la mesure du déphasage.

Un apprentissage non adaptatif aurait pu être réalisé pour effectuer cette validation comportementale. Néanmoins, dans la perspective d'une mise à jour des paramètres directement au sein de l'application où le réseau de neurones sera intégré en fonction des conditions expérimentales, un apprentissage adaptatif a été choisi.

Une fois le nombre de valeurs de distance et leur répartition déterminés, le calcul des poids et biais se fait selon la procédure indiquée figure III-g.

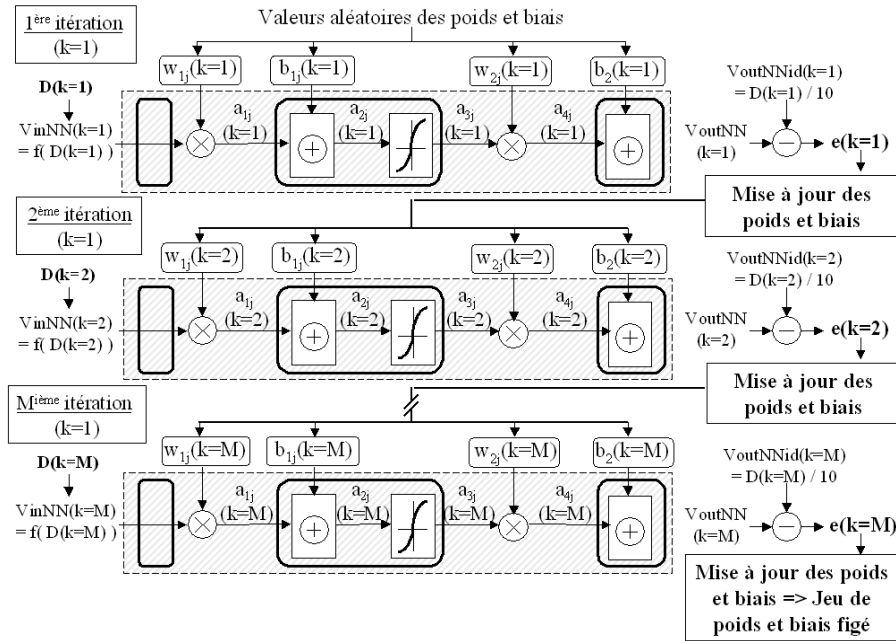


Figure III-g : Méthode utilisée pour calculer les poids et biais à partir des valeurs de distance issues du jeu d'apprentissage.

A chaque itération k , une valeur de mesure de distance $D(k)$ est choisie aléatoirement dans le jeu d'apprentissage. A cette mesure de distance est associée la mesure de la tension d'entrée $V_{inNN}(k)$ du réseau, correspondant à l'amplitude V_{outBP} . Cette tension $V_{inNN}(k)$ est alors multipliée par un vecteur de 3 poids $w_{1j}(k)$ correspondant au nombre de neurones sur la couche cachée. A chaque neurone correspond un indice j , avec $j \in [1;3]$. Le signal de sortie $a_{1j}(k)$ de chaque multiplieur est additionné à un biais $b_{1j}(k)$ puis distordu par l'intermédiaire d'une fonction non-linéaire sigmoïdale (figure III-d). La sortie de chaque fonction sigmoïdale $a_{2j}(k)$ est ensuite multipliée par un deuxième poids $w_{2j}(k)$. Les sorties $a_{4j}(k)$ des trois multiplieurs sont alors sommées entre elles, puis à un deuxième biais $b_2(k)$ pour former la sortie $V_{outNN}(k)$ du réseau de neurones (figure III-d). En résumé, à chaque itération k , chaque valeur de tension $V_{inNN}(k)$ subit les opérations suivantes :

$$a_{1j}(k) = K_{a1} w_{1j}(k) V_{inNN}(k) \quad (III-7)$$

$$a_{2j}(k) = a_{1j}(k) + K_{a2} b_{1j}(k) \quad (III-8)$$

$$a_{3j}(k) = K_{a3} \tanh(K_{th} a_{2j}(k)) \quad (III-9)$$

$$a_{4j}(k) = K_{a4} w_{2j}(k) a_{3j}(k) \quad (III-10)$$

$$V_{outNN}(k) = K_{sum} \sum_{j=1}^3 a_{4j}(k) + K_{b2} b_2(k); \quad (III-11)$$

avec K_{a1} , K_{a2} , K_{a3} , K_{th} , K_{a4} , K_{sum} et K_{b2} des constantes associées à chaque opération. Ces valeurs seront amenées à être paramétrées plus tard, lors de la simulation puis du test du réseau de neurones intégré. Pour l'instant, considérons toutes ces constantes égales à 1. A chaque itération k , la sortie $V_{outNN}(k)$ du réseau s'écrit donc (figure III-d):

$$V_{outNN}(k) = b_2(k) + \sum_{j=1}^3 w_{2j}(k) \tanh(w_{1j}(k) V_{inNN}(k) + b_{1j}(k)) \quad (III-12)$$

Ainsi, à la première itération $k=1$, une valeur de mesure de distance $D(k=1)$ est choisie aléatoirement dans l'ensemble d'apprentissage. La tension $V_{inNN}(k=1)$ correspondante est injectée dans le réseau de neurones. Les différents poids et biais mis en jeu lors de la première itération sont choisis aléatoirement avec une distribution uniformément répartie autour de 0. La tension de sortie $V_{outNN}(k=1)$ est calculée d'après l'équation (III-12). Cette tension est comparée à la tension $V_{outNNid}(k=1)$ correspondant à la valeur idéale de la tension de sortie du réseau, proportionnelle à la distance (équation (III-5)). Le facteur de proportionnalité K_0 est choisi arbitrairement égal à 0.1Vm^{-1} . La comparaison des deux tensions $V_{outNN}(k=1)$ et $V_{outNNid}(k=1)$ conduit à une erreur $e(k=1)$:

$$e(k=1) = V_{outNNid}(k=1) - V_{outNN}(k=1) = 0.1D(k=1) - V_{outNN}(k=1) \quad (III-13)$$

Cette erreur permet de mettre à jour les poids et biais initiaux par l'intermédiaire d'une méthode appelée rétropropagation de l'erreur ([Rumelhart, 1986], [Werbos, 1974]), de façon à minimiser la différence entre V_{outNN} et $V_{outNNid}$. Il existe plusieurs types d'algorithmes utilisés pour effectuer cette mise à jour. Deux algorithmes de mise à jour ont été testés pour cette application : l'algorithme du gradient ([Haykin, 1994]) et l'algorithme de Levenberg-Marquardt ([Levenberg, 1944], [Marquardt, 1963]). L'algorithme de Levenberg-Marquardt, algorithme du second ordre, a été choisi car il permettait d'obtenir une convergence plus rapide. Les autres méthodes du second ordre, telles que les filtres de Kalman étendu, n'ont pas testées dans le cadre de cette thèse

Une fois la mise à jour effectuée, la deuxième itération commence ($k=2$), avec une nouvelle valeur de mesure de distance $D(k=2)$ choisie aléatoirement dans l'ensemble d'entraînement, à laquelle correspond une nouvelle entrée $V_{inNN}(k=2)$ et une nouvelle sortie idéale $V_{outNNid}(k=2)$ (figure III-g – 2^{ème} itération). Une nouvelle erreur $e(k=2)$ est calculée par comparaison des sorties idéale et réelle ($V_{outNN}(k=2)$) du réseau. Cette nouvelle erreur permet de mettre à jour les poids et biais calculés à la fin de la première itération pour la prochaine itération.

Les itérations se succèdent jusqu'à ce que le jeu de poids et biais permette d'obtenir une erreur en sortie du réseau inférieure à e_{\max} en valeur absolue, et ce pour n'importe quelle valeur de l'ensemble d'apprentissage. Pour déterminer la valeur de e_{\max} , reprenons l'exemple donné dans le paragraphe (III.1.1). Le domaine de mesure Λ est égal à 0.9m. Pour distinguer deux valeurs de distance modulo Λ , il suffit d'avoir une évaluation de la distance à $\pm\Lambda/2$ près, soit $\pm 0.45\text{m}$. Or, d'après l'équation III-5, $V_{\text{outNNid}} = K_0 D$ avec $K_0 = 0.1\text{Vm}^{-1}$. Donc la valeur maximale de l'erreur tolérée de manière à évaluer la distance avec une précision suffisante pour lever l'indétermination modulo 2π sur la mesure de phase est égale à $K_0\Lambda/2$, soit 45mV.

En résumé, si l'erreur sur la tension de sortie du réseau de neurones est inférieure à 45mV, l'erreur faite sur l'évaluation de la distance est inférieure à 0.45m. Afin de se ménager une plage plus confortable, l'erreur maximale e_{\max} est prise égale à $\pm 10\text{mV}$ (figure III-h-7. en pointillés). Ainsi, la tension de sortie du réseau de neurones permettra une évaluation de la mesure de la distance à $\pm 0.1\text{m}$ sur l'intervalle [0.5m;3.5m], ce qui est largement suffisant pour lever l'indétermination due à la mesure de phase. Le nombre d'itérations n_{ptse} , choisi afin de constater que les n_{ptse} dernières valeurs de l'erreur $e(k)$ sont bien inférieures à la valeur maximale de l'erreur e_{\max} , est choisi égal à 100, de manière à ce que la probabilité que chacune des 9 valeurs de distance appartenant à l'ensemble d'apprentissage ait été tirée au moins une fois soit quasiment égale à 1.

Plusieurs phases d'apprentissage ont été réalisées, avec des initialisations différentes des paramètres, afin de valider la capacité du réseau à approcher la fonction (III-6). Comme le nombre d'exemples est inférieur au nombre de paramètres, des solutions différentes ont été relevées à chaque phase. Le but de cette phase comportementale n'est pas de déterminer la solution optimale, mais juste de montrer qu'il existe au moins une. La figure III-h montre un exemple de convergence de l'algorithme. Seules les dernières itérations ont été tracées. Les figures III-h.1. à 4. représentent la valeur des poids et biais lors des $n_{\text{ptse}}=100$ dernières itérations. Chacun des paramètres a atteint une valeur constante. Cela signifie que le gradient de l'erreur est devenu faible devant la valeur des paramètres, et ce pour n'importe quelles valeurs de distance $D(k)$ choisies aléatoirement dans l'ensemble d'apprentissage. La figure III-h.5. montre la variation de l'entrée $V_{\text{inNN}}(k)$ du réseau, relative aux différentes valeurs de distance $D(k)$, lors des 50 dernières itérations. La figure III-h.6. représente, en trait plein, la variation de la sortie $V_{\text{outNN}}(k)$ du réseau à l'itération k , et en pointillés, la variation idéale de la sortie $V_{\text{outNNid}}(k)$ du réseau lors des 50 dernières itérations. Les deux courbes sont quasiment superposées. La figure III-h.7. montre l'erreur $e(k)$ entre les deux dernières courbes

lors des 50 dernières itérations. On constate que cette erreur est inférieure à $e_{\max} = 10\text{mV}$ en valeur absolue. Le critère de convergence ayant été atteint, les itérations se sont arrêtées.

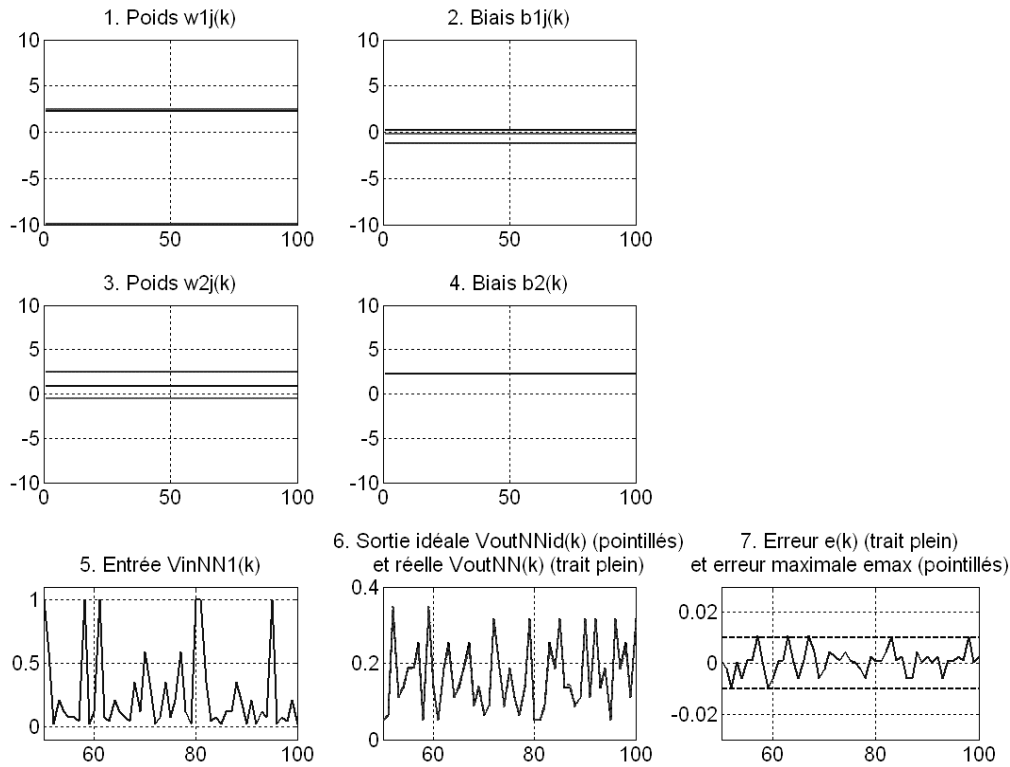
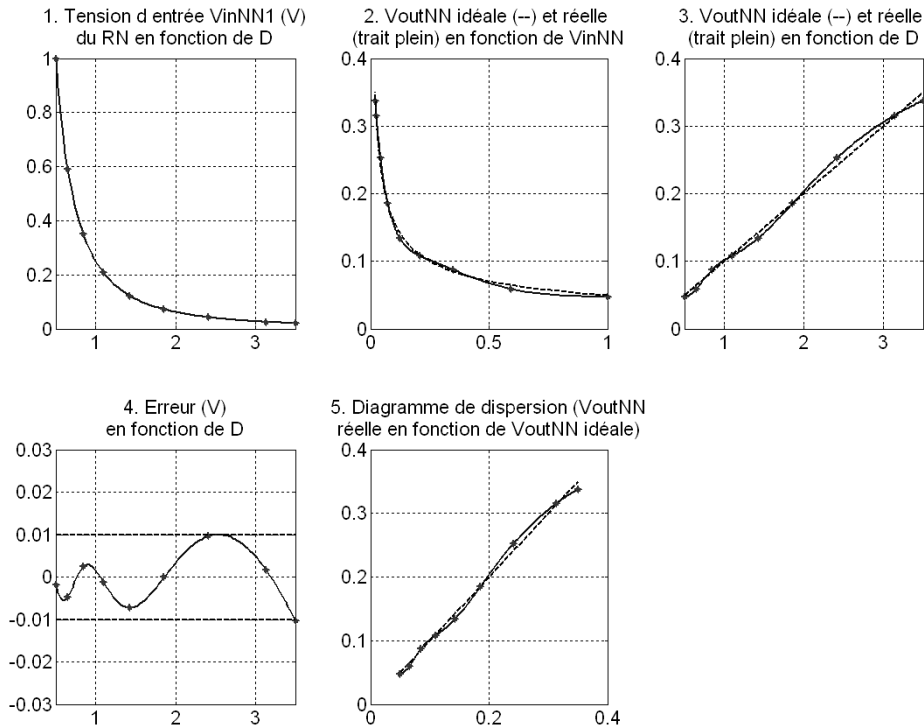


Figure III-h : De 1. à 4.: Variation des poids et biais lors des $n_{\text{prise}}=100$ dernières itérations. 5.: Valeurs $V_{\text{inNN}}(k)$ présentées en entrée du réseau de neurones lors des 50 dernières itérations, correspondant aux valeurs aléatoires de distance choisies dans l'ensemble d'apprentissage. 6.: Sorties idéale $V_{\text{outNNid}}(k)$ (pointillés) et réelle $V_{\text{outNN}}(k)$ lors des 50 dernières itérations. Les deux courbes sont quasiment superposées. 7. Erreur $e(k)$ (III-17) (trait plein) et erreur maximale e_{\max} (pointillés).

Ainsi, le réseau paramétré avec les valeurs finales des paramètres trouvées à la dernière itération est capable d'approximer la relation (III-6) pour les valeurs de distance choisies dans l'ensemble d'apprentissage. Pour valider la bonne capacité d'approximation du réseau, ce jeu de paramètres est testé pour 100 nouvelles valeurs de mesure de distance appartenant au jeu de validation. Ces 100 valeurs sont uniformément réparties sur l'intervalle $[0.5\text{m};3.5\text{m}]$. La figure III-i.1. montre les 100 valeurs de V_{inNN} correspondantes, en fonction de la distance, avec les 9 valeurs relatives au jeu d'apprentissage représentées par des points. Les tensions de sortie V_{outNN} et $V_{\text{outNNid}}=0.1\text{D}$ du réseau de neurones sont respectivement représentées en trait plein et en pointillés sur les figures III-i.2 et III-i.3, en fonction de V_{inNN} et en fonction de D . L'erreur commise par rapport à la tension idéale linéaire est représentée figure III-i.4. Si les points d'apprentissage ne sont pas parfaitement appris, c'est que les itérations s'arrêtent à partir du moment où l'erreur commise entre sortie idéale et sortie du réseau est inférieure à une valeur $e_{\max} = 10\text{mV}$. En réduisant la valeur de e_{\max} , la fonction réalisée par le réseau se

rapprochera de la fonction idéale. Le diagramme de dispersion, représentant la tension de sortie du réseau en fonction de la tension attendue, est tracée sur la figure III-i.



*Figure III-i : 1.: Ensemble des valeurs d'entrée V_{inNN} du réseau relatives aux 100 valeurs de distance appartenant au jeu de validation. Les 9 valeurs de l'ensemble d'apprentissage sont représentées par des points.
 2.: Tensions de sortie idéale ($V_{outNNid}$) et réelle (V_{outNN}) du RN en fonction de V_{inNN} .
 3.: Tensions de sortie idéale ($V_{outNNid}$) et réelle (V_{outNN}) du RN en fonction de D.
 4.: Erreur (trait plein) ; l'erreur maximale $\pm e_{max}$ est représentée en pointillés.
 5. Diagramme de dispersion.*

En résumé, le jeu de paramètres final issu de la phase d'apprentissage permet d'approcher l'équation (III-6) avec une précision inférieure à 10mV pour l'ensemble des valeurs appartenant au jeu de validation. Comme les valeurs de distance sont en grand nombre et réparties uniformément sur l'intervalle [0.5m ; 3.5m], on peut considérer que le réseau vérifie l'équation (III-6) sur tout cet intervalle. On vérifie ainsi la faculté de généralisation du réseau, capable d'approcher une fonction monotone sur tout un intervalle alors qu'il n'a été entraîné que pour un nombre fini de valeurs (en l'occurrence 9 valeurs de distance). La structure représentée figure III-d, permet donc d'approcher l'équation (III-6) avec la précision suffisante de manière à ce que sa tension de sortie soit linéaire avec la distance. Ainsi, par l'intermédiaire de cette structure, il est possible de déterminer la valeur de la distance avec suffisamment de précision pour lever l'indétermination modulo 2π sur la mesure de déphasage. En corrélant les résultats donnés par la sortie du phasemètre et celle du réseau de neurones, la valeur de distance peut donc être donnée avec une résolution télémétrique pour un intervalle de distance 3.3 fois supérieur à celui restreint par la mesure du déphasage seule.

III.3. CLASSIFICATION DE SURFACES

III.3.1. Principe

L'objectif est d'extraire à l'aide d'un réseau de neurones l'information relative au type de surface, contenue dans le signal photoélectrique d'amplitude V_{outBP} , dans le but de classer différentes cibles sur un certain intervalle de distance et pour un angle d'incidence variable. Pour cette application, la précision de la mesure relative à la distance n'étant pas primordiale, la fréquence de modulation f_{RF} de la diode laser est réduite à 50MHz. Ainsi l'intervalle de mesure sans indécision Δ et la précision δD sont respectivement égaux à 3m et 166 μ m.

Tout comme dans le système de levée d'indétermination sur la mesure de distance décrit précédemment, le signal V_{outBP} de fréquence f_{IF} est converti en signal continu. Celui-ci constitue le signal d'entrée V_{inNN_1} du réseau de neurones. L'ensemble des tensions d'entrée V_{inNN_1} , représentées en échelle logarithmique en fonction de la distance sur la figure III-j, est scindé en quatre zones. Chacune de ces zones contient l'ensemble des tensions d'entrées relatives à une surface donnée S_i (S_1 : surface plastifiée, S_2 : papier glacé, S_3 : mur peint et S_4 : surface poreuse), dans des conditions de test données (angle d'incidence, température, défocalisation, etc.). Ces zones sont délimitées par les amplitudes du signal V_{outBP} pour des angles d'incidence égaux à 0 et à $\pm\theta_{0max}$. Des frontières S_i-S_j ont été tracées en pointillés à équidistance de chacune des zones relatives aux surfaces S_i et S_j .

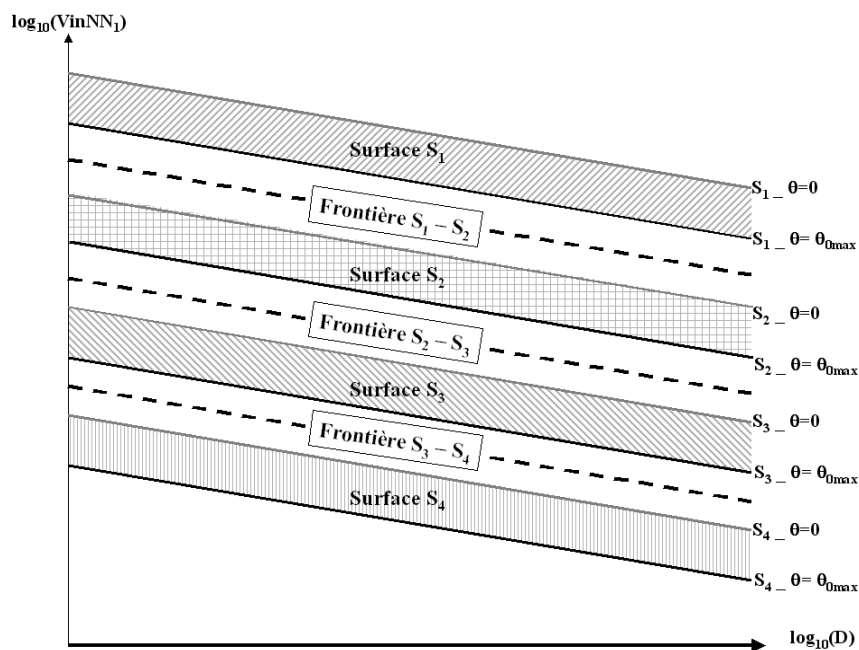


Figure III-j : Découpage de l'espace des tensions d'entrée en quatre zones relatives à chacune des surfaces. Chaque zone est limitée par les valeurs des amplitudes des signaux pour des angles d'incidence respectivement égaux à 0 et $\pm\theta_{0max}$. Chaque frontière S_i-S_j est équidistante des zones relatives aux surfaces S_i et S_j .

Un exemple de courbes expérimentales relatives à la variation de l'entrée V_{inNN_1} en fonction de la distance pour les quatre surfaces testées est représenté sur la figure III-k, pour un angle d'incidence égal à $\pi/6$ (30°), une température ambiante de 25°C et avec une légère défocalisation. Les frontières S_i-S_j pour ces conditions de simulations spécifiques ont aussi été représentées. On constate que chaque courbe est bien incluse dans la zone relative à la surface testée.

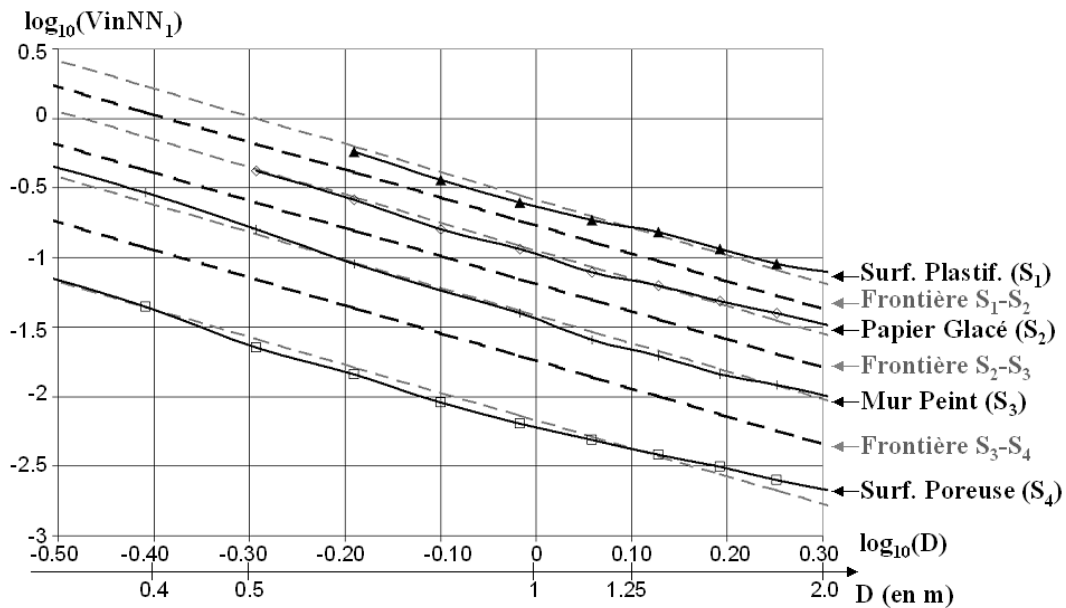


Figure III-k : Tension d'entrée V_{inNN_1} du réseau de neurones en fonction de la distance en échelle log-log pour un angle d'incidence égal à $\pi/6$. Les courbes en trait plein sont relatives aux mesures expérimentales réalisées avec le télémètre laser par déphasage. Les droites grises pointillées correspondent aux courbes idéales (suivant la relation III-1) relatives à chacune des surfaces. Les droites noires pointillées représentent les frontières S_i-S_j entre les zones relatives à chacune des surfaces.

Le rôle du réseau de neurones est de déterminer le type de surface testée par l'intermédiaire de sa ou ses tensions de sortie. Pour accroître les plages de distance et d'angle d'incidence sur lesquelles sera effectuée la classification, on exploite le signal de sortie $V_{outRANG}$ du phasemètre, directement proportionnel à la distance. Le signal de sortie du phasemètre $V_{outRANG}$, déjà continu, constitue le second signal d'entrée V_{inNN_2} du réseau de neurones.

Une troisième tension d'entrée V_{inNN_3} , contenant une information relative à l'angle d'incidence θ_0 , pourrait être ajoutée de manière à accroître les plages de classification du système. Néanmoins, les mesures précises d'angles d'incidence ne sont pas faciles à réaliser dans la plupart des systèmes de détection de surfaces, et, en accord avec le fait que l'intensité lumineuse d'une source lambertienne varie en $\cos(\theta_0)$, les simulations comportementales montrent qu'un réseau de neurones à deux entrées donnent des résultats quasi similaires à ceux d'un réseau à trois entrées, mais pour un temps de convergence nettement plus faible. De

plus, comme l'objectif est d'implanter le réseau de neurones dans la tête optique du télémètre laser afin de rendre le système transportable, il est nécessaire de minimiser au maximum la taille du réseau.

L'angle d'incidence n'étant pas mesuré, la capacité de discrimination du système est limitée à un certain intervalle d'angles d'incidence θ_0 , de manière à ce que l'amplitude du signal V_{outBP} ne corresponde pas, pour une distance D donnée, à deux types de surface S_i et S_j différents, à des angles d'incidence θ_{0i} et θ_{0j} différents. En d'autres termes, au regard de la figure III-j, on souhaite choisir des intervalles d'angle d'incidence θ_0 de manière à ce que les zones relatives à chacune des surface soient disjointes. Pour cette raison, et dans le but de détecter quatre types de surfaces différents, l'intervalle associé à l'angle d'incidence θ_0 est limité à $[-\theta_{0max} ; \theta_{0max}]$, avec $\theta_{0max}=\pi/5$ et l'intervalle de distance est restreint à $[0.5m ; 2m]$ ([Gatet, 2007]). Comme l'intervalle de mesure sans indécision Λ est égal à 3m, le second signal d'entrée du réseau de neurones V_{inNN_2} relatif au signal de sortie du phasemètre est bien linéaire par rapport à la distance sur l'intégralité de l'intervalle choisi.

Le système complet de détection de surface est représenté figure III-l.

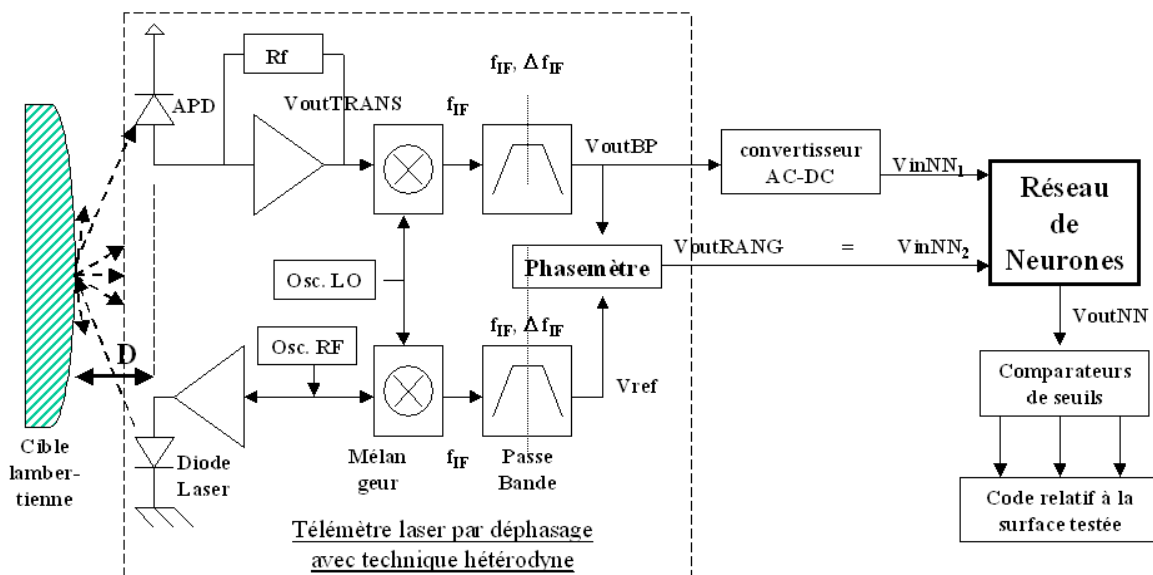


Figure III-l : Représentation du système de détection de surface mettant en œuvre un télémètre laser par déphasage et un réseau de neurones.

Lorsque un problème de classification est un problème à plusieurs classes, deux approches sont possibles :

- soit concevoir un réseau possédant autant de sorties que de classes à discriminer,
- soit décomposer le problème en plusieurs sous-problèmes de classification à deux classes.

La principale difficulté de la première approche est relative à la discrimination des classes adjacentes. La deuxième approche est souvent beaucoup plus sûre, car elle permet de réaliser

des réseaux plus simples, plus petits, et de choisir des paramètres d'entrée différents en fonction des classes à discriminer ([Dreyfus, 2004]). Néanmoins, l'objectif ici est de simplifier au maximum le système neuronal choisi, afin d'être intégré par la suite au sein de l'application télémétrique. C'est la raison pour laquelle la conception d'un seul réseau a été retenue.

Comme l'application développée a pour but de classer quatre types de surfaces, la première approche conduit à choisir quatre neurones sur la couche de sortie, chacun relatif à un type de surface testé. Le rôle de chaque neurone de sortie est de ne présenter que deux tensions "binaires" (0V ou 1V), selon que la surface est éclairée ou non.

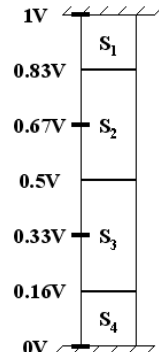
Par souci de simplification du réseau et de généralisation de la structure avec celle développée pour lever l'indétermination sur la mesure de distance, il a été choisi de ne placer qu'un seul neurone de sortie. La discrimination entre les différents types de surfaces sera effectuée en sortie du réseau, par l'intermédiaire de détecteurs de seuil, en fonction de la valeur de la tension de sortie V_{outNN} (figure III-1).

Pour cela, en considérant toujours la valeur de la tension de sortie normalisée, on distribue sur l'intervalle [0V ; 1V] quatre valeurs de tension idéales relatives à chacune des surfaces et trois valeurs de tensions correspondant aux frontières entre chacune des surfaces. Les quatre valeurs de tension idéales sont uniformément réparties sur l'intervalle [0V ; 1V], bornes incluses. Les tensions 0V et 1V peuvent être choisies comme tensions idéales pour des raisons électroniques, puisque la tension de sortie ne pourra ni dépasser la valeur de la tension d'alimentation (normalisée à 1V) ni celle de la masse (0V). Ainsi, les quatre valeurs idéales de tension de sortie relatives à la surface plastifiée, au papier glacé, au mur peint et à la surface poreuse sont respectivement 1V, 0.67V, 0.33V et 0V.

Les trois valeurs de tension relatives aux frontières entre chacune des surfaces sont choisies à équidistance de chacune des tensions de référence : 0.83V, 0.5V et 0.17V (tableau III- α). Les intervalles de sortie relatifs à chacune des surfaces sont donc [0V ; 0.17V], [0.17V ; 0.5V], [0.5V ; 0.83V] et [0.83V ; 1V], avec pour tension de sortie idéale respective 0V, 0.33V, 0.67V et 1V. Les tensions de commutation des comparateurs de seuil correspondent aux tensions relatives à chacune des frontières entre les différents types de surface. Ainsi, l'intervalle dans lequel se situe la valeur de la tension V_{outNN} est corrélé à un code relatif aux sorties de chacun des comparateurs (tableau III- α).

Tableau III-α : Les tensions de sortie idéales du réseau de neurones et les intervalles de sortie associés à chacun des types de surfaces sont représentées dans les deux premières colonnes. Les codes associés aux sorties respectives de chacun des détecteurs de seuil et relatifs au type de surface testée sont représentés dans les trois dernières colonnes.

Surface	Type	Tension de sortie idéale	Intervalle de sortie associé	Sortie du premier détecteur de seuil V_{out_1} (0.83V)	Sortie du second détecteur de seuil V_{out_2} (0.5V)	Sortie du troisième détecteur de seuil V_{out_3} (0.16V)
S ₁	Surface plastique	1V]0.83;1['1'	'1'	'1'
S ₂	Papier glacé	0.67V]0.5;0.83['0'	'1'	'1'
S ₃	Mur peint	0.33V]0.16;0.5['0'	'0'	'1'
S ₄	Surface poreuse	0V]0;0.16['0'	'0'	'0'



Pour résumer, chaque surface testée est corrélée à un intervalle de tensions de sortie V_{outNN} inclus dans l'intervalle [0V ; 1V], "centré" sur une tension de sortie idéale, et à un code relatif aux sorties des comparateurs de seuil placés en sortie du réseau. Si par exemple, la surface testée est un mur peint, la tension idéale de sortie attendue est 0.33V. Donc si un mur peint est éclairé par le télémètre à une distance comprise entre 0.5m et 2m et pour un angle d'incidence compris entre $-\pi/5$ et $\pi/5$, la tension attendue en sortie du réseau doit se rapprocher de la tension 0.33V, avec une tolérance de $\pm 0.16V$, de manière à ce que cette tension appartienne à l'intervalle [0.17V ; 0.5V]. Si c'est le cas, les sorties respectives des trois comparateurs de seuils indiquent 0V, 0V et 1V, soit le code 001. Le raisonnement est identique pour les trois autres types de surface (plastique, papier glacé et surface poreuse), donc les codes correspondants sont respectivement 111, 011 et 000.

Utiliser seulement un neurone de sortie suivi de détecteurs de seuil présente l'intérêt majeur de la reconfigurabilité du système. En effet, si le nombre de surfaces à discriminer est différent de quatre, il suffit de subdiviser de manière différente l'intervalle des tensions de sortie [0V ; 1V] et d'utiliser le nombre adéquat de comparateurs de seuil. Par exemple, pour cinq types de surfaces à discriminer, les intervalles de sortie seraient [0V ; 0.125V], [0.125V ; 0.375V], [0.375V ; 0.625V], [0.625V ; 0.875V] et [0.875V ; 1V]. Le nombre de comparateurs de seuil sera égal à quatre et les codes relatifs à chacune des surfaces seront respectivement 0000, 0001, 0011, 0111 et 1111. Donc il n'est pas nécessaire de changer la structure du réseau pour discriminer un nombre différent de surfaces.

Au-delà de l'aspect reconfigurabilité, ce réseau a une structure qui possède le même nombre de sorties que le réseau utilisé pour la levée d'indétermination sur la mesure de distance décrit précédemment. Les simulations comportementales montrent qu'une seule

couche cachée constituée de trois neurones cachés suffit pour effectuer la discrimination des différents types de surfaces considérées, pour une distance et un angle d'incidence variables (paragraphe III.3.2). Ainsi, avec deux entrées, trois neurones cachés et un neurone de sortie, la structure neuronale capable d'effectuer la classification de surfaces au sein du télémètre laser (figure III-m) est très proche de celle détaillée dans l'application précédente (figure III-d). Il suffit de prendre l'entrée V_{inNN_2} égale à 0 pour pouvoir utiliser la structure de la figure III-m dans l'application de levée de l'indétermination sur la mesure de distance. Dans l'optique d'une intégration du réseau, cette structure a été choisie pour valider son aptitude à effectuer la classification des surfaces.

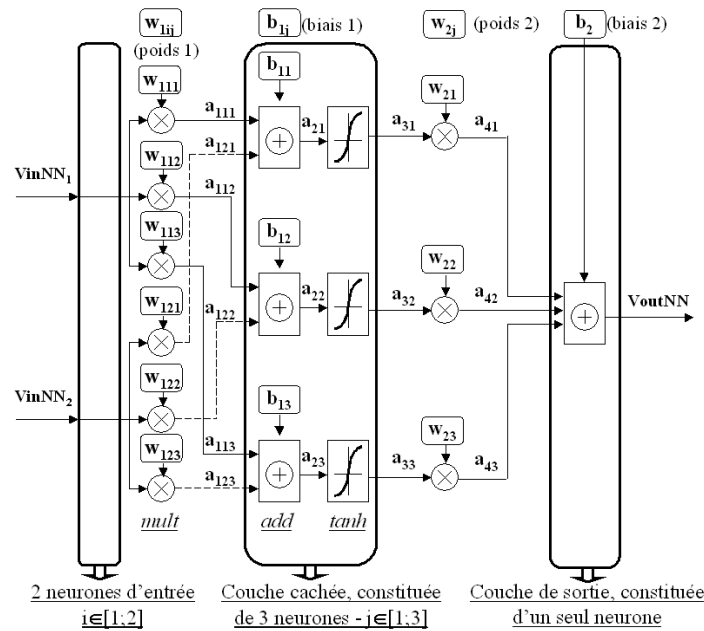


Figure III-m : Structure du réseau de neurones pour la discrimination de surfaces.

III.3.2. Simulations comportementales sous Matlab

Comme le réseau de neurones choisi pour cette application est non bouclé, les poids et biais doivent être calibrés durant une phase d'apprentissage de manière à ce que le réseau soit apte à discriminer chaque surface pour toute distance D et tout angle d'incidence θ_0 respectivement compris dans les intervalles $[0.5m ; 2m]$ et $[-\pi/5 ; \pi/5]$. Pour cela, il faut déterminer l'ensemble des différentes valeurs de distance $D(k)$, d'angle d'incidence $\theta_0(k)$ et de coefficients de réflexion diffusants $\rho(k)$ qui vont faire partie du jeu d'entraînement du réseau et permettre de déterminer les différentes valeurs d'entrées $V_{inNN_1}(k)$ et $V_{inNN_2}(k)$ à l'itération k .

Comme quatre types de surfaces sont testées, le nombre de coefficients de réflexion diffusants $n_{S_{app}}$ est égal à 4. Les valeurs des coefficients de réflexion diffusants apparent des différentes surfaces testées ont été déterminés expérimentalement à l'aide du télémètre laser (Figure III-k). Ils sont respectivement égaux à 0.9, 0.6, 0.35 et 0.15. Le nombre d'angles d'incidence $n_{\theta_{app}}$ a été choisi égal à 2 de manière à minimiser les temps de convergence durant la phase d'apprentissage. Les deux valeurs d'angle choisies sont 0 et $|\theta_{0max}|=\pi/5$. Enfin, de la même façon que pour l'entraînement du réseau dédié à la levée d'indétermination sur la mesure de distance, le choix des différentes valeurs de distance a été optimisé de manière à avoir une répartition uniforme sur une échelle logarithmique. Pour déterminer le "pas" entre chaque valeur de distance, on utilise la méthode suivante (Figure III-n). On détermine dans un premier temps quelles sont les deux types de surface les plus difficiles à discriminer l'une par rapport à l'autre, en choisissant celles dont les coefficients de réflexion diffusant sont les plus proches. Soient ρ_j et ρ_{j+1} les coefficients de réflexion diffusants de ces deux surfaces S_j et S_{j+1} . Plaçons-nous à la distance minimale $D_0=D_{min}=0.5m$ pour un angle d'incidence θ_0 nul. Relevons la valeur de $VinNN_1$ correspondant à la surface S_{j+1} . Soit D_1 la valeur de distance la plus proche de D_0 pour laquelle on obtient la même tension $VinNN_1$ pour la surface S_j avec un angle $\theta_0=\theta_{0max}$ (figure III-n). On a donc l'égalité suivante :

$$VinNN_1(S_{j+1}, D_0, \theta_0=0) = VinNN_1(S_j, D_1, \theta_0=\theta_{max}) \quad (III-14)$$

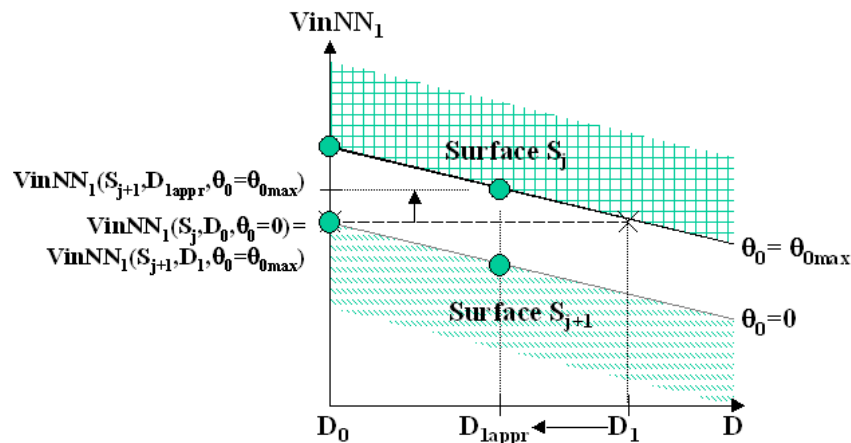


Figure III-n :Méthode de répartition des valeurs de distance appartenant au jeu d'entraînement, représentée en échelle log-log. Les différentes valeurs d'entraînement choisies pour calibrer le réseau sont mises en valeur par des points.

Si l'on choisit D_0 et D_1 comme valeurs d'entraînement, il existerait deux jeux de valeurs donnant la même valeur de tension $VinNN_1$. Durant la phase d'entraînement, le rôle du réseau de neurones serait donc de discriminer les deux types de surfaces, sachant que sa première entrée serait la même pour les deux types de surface (équation III-14) et que sa deuxième

entrée V_{inNN_2} prendrait des valeurs peu différentes selon la surface testée (la tension V_{inNN_2} est proportionnelle à la distance, et D_0 et D_1 sont proches l'une de l'autre).

Pour faciliter l'entraînement du réseau, la distance D_{1app} qui suit D_0 dans le jeu d'apprentissage a été choisie inférieure à D_1 (figure III-p). Ainsi, chaque jeu de valeurs d'entrée correspond à une tension d'entrée V_{inNN_1} relative à un seul type de surface. Dans la pratique, $D_{1app} = 0.9D_1$. Les autres valeurs de distance appartenant au jeu d'entraînement sont déterminées de la même manière. La figure III-o représente l'ensemble des valeurs de tension d'entrée V_{inNN_1} pour les quatre types de surface, pour une distance variant de 0.5m à 2m et pour cinq angles d'incidence θ_0 répartis uniformément entre 0 et θ_{0max} . Les différentes valeurs d'entraînement choisies pour calibrer le réseau sont mises en valeur par des points.

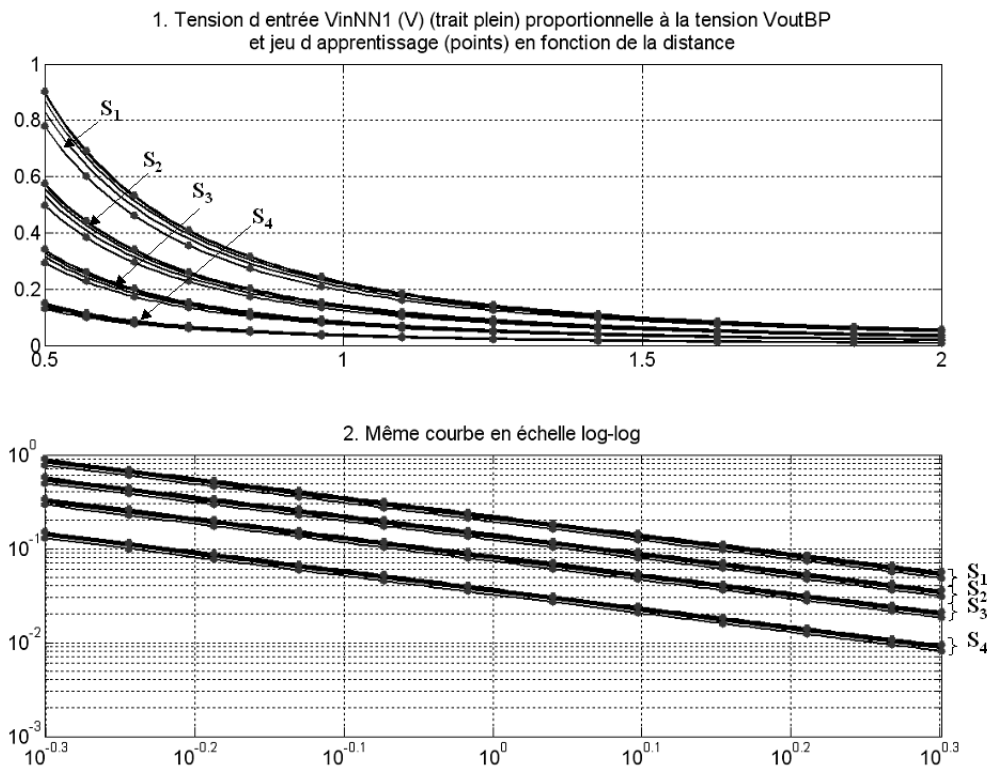


Figure III-o : Courbes représentant l'amplitude d'entrée V_{inNN_1} du réseau de neurones associée aux 4 types de surfaces en fonction de la distance dans l'intervalle [0.5m;2m], pour cinq valeurs d'angles d'incidence comprises entre 0 et $\pi/5$. Les points représentent les valeurs d'entraînement.

Une fois ces trois jeux d'apprentissage fixés, on utilise la méthode décrite figure III-q, proche de celle utilisée sur la figure III-g. A chaque itération, trois valeurs aléatoires sont choisies : une valeur de coefficient de réflexion diffusant $\rho(k)$ parmi les $n_{Sapp}=4$ types de surfaces à discriminer, une valeur d'angle d'incidence $\theta_0(k)$ parmi $\{0 ; |\theta_{0max}| = \pi/5\}$ et une valeur de distance $D(k)$ parmi les $n_{Dapp}=12$ distances appartenant à l'ensemble d'apprentissage. A ces trois mesures sont associées les tensions d'entrée $V_{inNN_1}(k)$ et $V_{inNN_2}(k)$ du réseau, proportionnelles aux signaux V_{outBP} et $V_{outRANG}$ (III-1 et III-2).

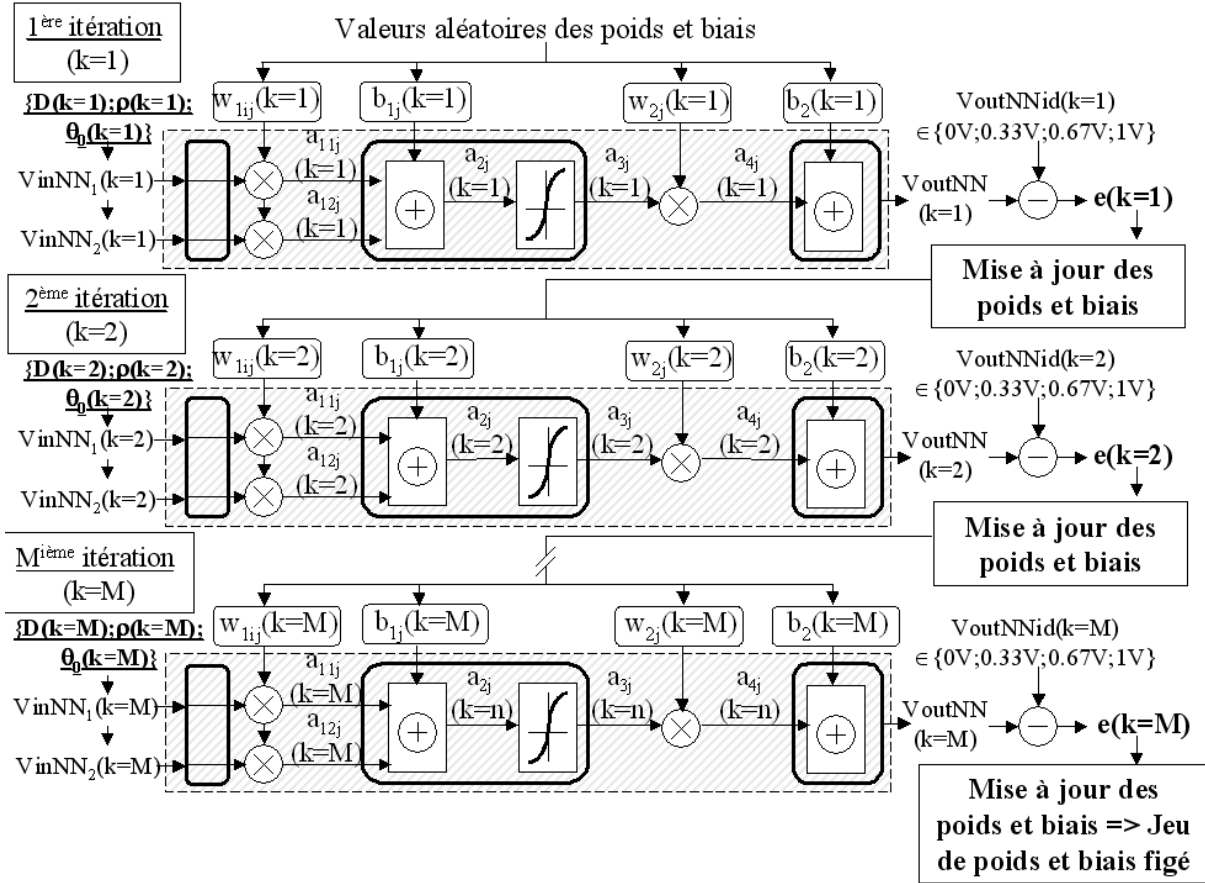


Figure III-p : Méthode utilisée pour calculer les poids et biais à partir des valeurs de distance et d'angle d'incidence issues de leur jeu d'apprentissage respectifs.

Ces tensions $V_{inNN_1}(k)$ et $V_{inNN_2}(k)$ sont alors respectivement multipliées par trois poids $w_{11j}(k)$ et $w_{12j}(k)$, avec l'indice j correspondant au $j^{\text{ème}}$ neurones sur la couche cachée ($j \in [1;3]$) (figure III-m). Les sorties $a_{11j}(k)$ et $a_{12j}(k)$ des multiplieurs sont sommées entre elles, puis additionnées à un biais $b_{1j}(k)$. A chaque sortie $a_{2j}(k)$ d'additionneur est appliqué une non-linéarité de type sigmoïdale. Chaque sortie de fonction sigmoïdale $a_{3j}(k)$ est multipliée par un deuxième poids $w_{2j}(k)$. Les sorties $a_{4j}(k)$ des multiplieurs sont alors sommées entre elles, puis à un deuxième biais $b_2(k)$ pour former la sortie $V_{outNN}(k)$ du réseau de neurones (figure III-m). En résumé, à chaque itération k , chaque valeur de tension $V_{inNN_i}(k)$ ($i \in [1;2]$) subit les opérations suivantes :

$$a_{1ij}(k) = K_{a1i} w_{1ij}(k) V_{inNN_i}(k) \quad (III-15)$$

$$a_{2j}(k) = a_{11j}(k) + a_{12j}(k) + K_{a2} b_{1j}(k) \quad (III-16)$$

$$a_{3j}(k) = K_{a3} \tanh(K_{th} a_{2j}(k)) \quad (III-17)$$

$$a_{4j}(k) = K_{a4} w_{2j}(k) a_{3j}(k) \quad (III-18)$$

$$V_{outNN}(k) = K_{sum} \sum_{j=1}^3 a_{4j}(k) + K_{b2} b_2(k) \quad (III-19)$$

avec K_{a1i} , K_{a2} , K_{a3} , K_{th} , K_{a4} , K_{sum} et K_{b2} des constantes associées à chaque opération. Ces valeurs seront amenées à être paramétrées plus tard, lors de la simulation puis du test du réseau de neurones intégré. Pour l'instant, considérons toutes ces constantes égales à 1. A chaque itération k , la sortie $V_{outNN}(k)$ du réseau s'écrit donc (figure III-m):

$$V_{outNN}(k) = b_2(k) + \sum_{j=1}^3 w_{2j}(k) \tanh(w_{11j}(k)V_{inNN_1}(k) + w_{12j}(k)V_{inNN_2}(k) + b_{1j}(k)) \quad (III-20)$$

Ainsi, à la première itération $k=1$ (figure III-p – 1^{ère} itération), une valeur de coefficient de réflexion diffusant $\rho(k=1)$, une mesure d'angle d'incidence $\theta_0(k=1)$ et une mesure de distance $D(k=1)$ sont prises au hasard dans leurs ensembles d'apprentissage respectifs. Les tensions $V_{inNN_1}(k=1)$ et $V_{inNN_2}(k=1)$ correspondantes sont injectées dans le réseau de neurones. Les différents poids et biais mis en jeu lors de la première itération sont choisis aléatoirement avec une distribution uniformément répartie autour de 0. La tension de sortie $V_{outNN}(k=1)$ est calculée d'après l'équation (III-20). Cette tension est comparée à la tension $V_{outNNid}(k=1)$ correspondant à la valeur idéale souhaitée de la tension de sortie du réseau (0V, 0.33V, 0.67V ou 1V). Cette valeur idéale est relative au choix de la surface observée, par l'intermédiaire du choix du coefficient de réflexion diffusant $\rho(k=1)$ en début d'itération (tableau III- α). La comparaison des deux tensions $V_{outNN}(k=1)$ et $V_{outNNid}(k=1)$ conduit à une erreur $e(k=1)$:

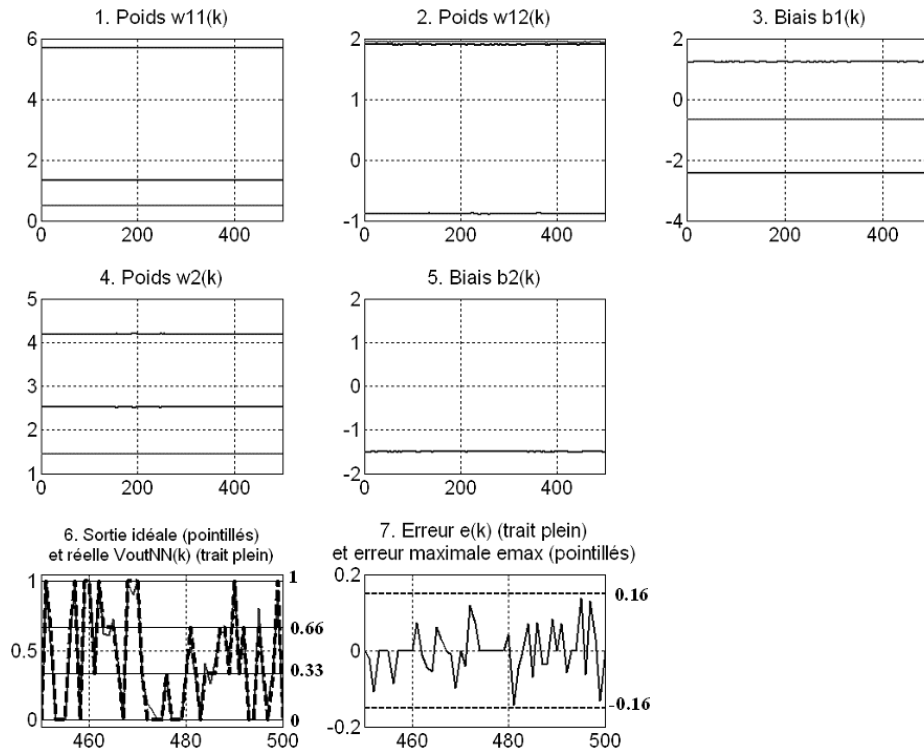
$$e(k=1) = V_{outNNid}(k=1) - V_{outNN}(k=1) = 0.1D(k=1) - V_{outNN}(k=1) \quad (III-21)$$

Cette erreur permet de mettre à jour les poids et biais initiaux par l'intermédiaire de la méthode de rétropropagation de l'erreur, associée à l'algorithme de Levenberg-Marquardt, de façon à minimiser la différence entre V_{outNN} et $V_{outNNid}$. Une fois que la mise à jour a été effectuée, la deuxième itération commence ($k=2$), avec une nouvelle valeur de mesure de distance $D(k=2)$ choisie aléatoirement dans l'ensemble d'entraînement, auxquelles correspondent deux nouvelles valeurs d'entrées $V_{inNN_1}(k=2)$ et $V_{inNN_2}(k=2)$, et une nouvelle sortie idéale $V_{outNNid}(k=2)$ (figure III-p – 2^{ème} itération). Une nouvelle erreur $e(k=2)$ est calculée par comparaison des sorties idéale et réelle ($V_{outNN}(k=2)$) du réseau. Cette nouvelle erreur permet de mettre à jour les poids et biais calculés à la fin de la première itération pour la prochaine itération.

Les itérations se succèdent jusqu'à ce qu'un critère d'arrêt soit atteint. Le critère de convergence choisi est celui de la valeur maximale de l'erreur e_{max} . Sa valeur doit être inférieure à 0.166V en valeur absolue pour l'ensemble des combinaisons possibles entre les trois jeux d'entraînement. En effet, si l'erreur atteint la valeur maximale 0.166V, la tension de sortie est à la limite de l'intervalle qui lui est alloué. La valeur de l'erreur e_{max} a été choisie

égale à 0.16V, de manière à accroître l'écart entre les tensions de sortie relatives à chacune des surfaces. Ainsi, les itérations continuent tant que, pour un nombre d'itérations n_{ptse} , l'erreur n'est pas inférieure à e_{max} . Ce nombre d'itérations n_{ptse} est fonction du nombre de combinaisons possibles entre les trois jeux d'apprentissage. Afin d'être sûr que le réseau a bien été entraîné pour chacune des combinaisons, le nombre d'itérations n_{ptse} est choisi 5 fois supérieur au nombre de combinaisons possibles entre chacun des éléments des trois jeux d'entraînement. Ainsi, les itérations ne s'arrêtent que si l'erreur constatée en sortie est bien inférieure à 0.16V lors des $n_{ptse}=5 \times 4 \times 2 \times 12=480$ dernières itérations.

La figure III-q montre un exemple de convergence de l'algorithme. Seules les dernières itérations ont été représentées. Les figures III-q.1. à 5. représentent la valeur des poids et biais lors des 500 dernières itérations. Chacun des poids et biais a atteint une valeur constante. Cela signifie que le gradient de l'erreur est devenu très faible devant la valeur des poids et biais, et ce quels que soient le coefficient de réflexion diffusant $\rho(k)$, l'angle d'incidence $\theta_0(k)$ ou la distance $D(k)$ choisis aléatoirement dans leur ensemble d'apprentissage respectif. La figure III-q.6. représente, en trait plein, la variation de la sortie $V_{outNN}(k)$ du réseau en fonction de son entrée $V_{inNN}(k)$ et des différents poids et biais à l'itération k , et en pointillés, la variation idéale de la sortie $V_{outNNid}(k)$ du réseau, relative à la sortie idéale pour le type de surface testé, lors des 50 dernières itérations. Les deux courbes sont quasiment superposées. La figure III-q.7. montre l'erreur $e(k)$ existant entre les deux courbes de la figure III-q.6. lors des 50 dernières itérations. On constate que cette erreur est inférieure en valeur absolue à $e_{max}=0.16V$ (droites pointillées). Le critère de convergence ayant été atteint, les itérations se sont arrêtées. Ainsi, pour les $n_{Dappi}=12$ distances appartenant à l'intervalle [0.5m ; 2m] et pour les angles d'incidence égaux à 0 et θ_{0max} , le système est capable de discriminer les quatre types de surface choisis. Les valeurs des poids et biais trouvées lors de la dernière itération sont sauvegardés.



*Figure III-q : De 1. à 5.: Variation des poids et biais lors des $n_{pise}=500$ dernières itérations.
6.: Sorties idéale ($V_{outNNid}(k)$ à l'itération k en pointillés) et réelle ($V_{outNN}(k)$ à l'itération k en trait plein) lors des 50 dernières itérations. Les deux courbes sont quasiment superposées.
7. Erreur $e(k)$ (III-17) (trait plein) et erreur maximale $\pm e_{max}$ (pointillés).*

Maintenant, vérifions que le système est capable, à partir de ces poids et biais finaux, de détecter les quatre surfaces pour toute valeur de distance dans l'intervalle $[0.5m ; 2m]$ et pour tout angle d'incidence appartenant à l'intervalle $[-\pi/5 ; \pi/5]$. Pour cela, on choisit $n_{Dval}=60$ valeurs de distance et $n_{\theta val}=10$ valeurs d'angle d'incidence uniformément réparties dans leurs intervalles respectifs, de manière à tester la faculté de généralisation du réseau de neurones. Un exemple de résultat, relatif au test des 4 types de surface, est représenté figure III-r. Les tensions d'entrées de test V_{inNN_1} et V_{inNN_2} sont tracées en fonction de la distance D et pour les 10 angles d'incidence figure III-r.1. et III-r.2. Les courbes des figures III-r.2 sont toutes superposées car la sortie du phasemètre ne dépend pas de l'angle d'incidence. La tension de sortie V_{outNN} du réseau de neurones correspondante, calculée à partir des poids et biais finaux et de l'équation (III-20), est tracée figure III-r.3. Chaque courbe en trait plein est relative à un angle d'incidence donné. La courbe en pointillés montre la tension de sortie idéale $V_{outNNid}=1V$. La valeur absolue de l'erreur entre les deux courbes est représentée figures III-r.4. La courbe en pointillés est relative à l'erreur maximale $e_{max}=0.16V$ ne devant pas être dépassée pour que la détection de la surface soit correcte. On peut donc conclure des 4 figures III-r.4. que les 4 types de surfaces sont bien détectés pour toute valeur de distance sur l'intervalle $[0.5m;2m]$ et tout angle d'incidence appartenant à l'intervalle $[-\pi/5 ; \pi/5]$. Les

figures III-r-5. représentent les erreurs commises par le réseau. Pour chaque couple de distance et d'angle d'incidence pour lequel l'erreur dépasse la valeur e_{max} , la valeur e_{bin} est égale à 1. Si cette erreur est inférieure à e_{max} , la valeur e_{bin} est égale à 0. Comme la valeur e_{bin} est égale à 0 sur tout l'intervalle de distance pour n'importe quelle valeur d'angle d'incidence et pour les 4 types de surface, la conclusion tirée des suites de la figure III-r.4. est confirmée. Ainsi, la faculté de généralisation du réseau de neurones est validée. Le réseau est capable d'effectuer sa tâche de discrimination pour les quatre types de surfaces, pour n'importe quelle valeur de distance et d'angle d'incidence appartenant à leur intervalle respectif [0.5m ; 2m] et $[-\pi/5, \pi/5]$, et ce après n'avoir été entraîné qu'à partir de 12 valeurs de distance et 2 valeurs d'angle d'incidence. Les valeurs des poids et biais établies selon la procédure décrite sur la figure III-p sont définitivement figées.

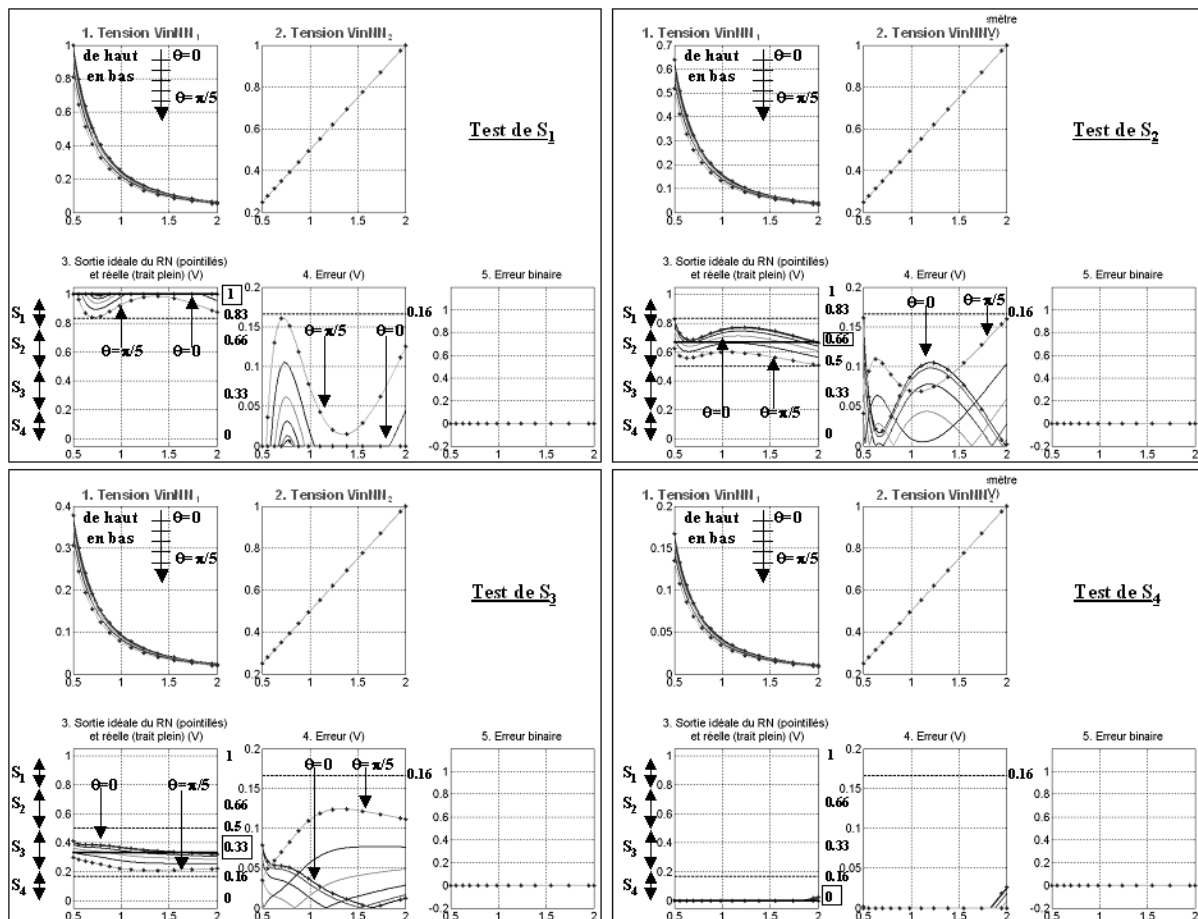


Figure III-r : 1. et 2.: Ensemble des valeurs d'entrée V_{inNN_1} et V_{inNN_2} du réseau relatives aux 60 valeurs de distance et aux 10 valeurs d'angle d'incidence appartenant à leur jeu de validation respectif. Les valeurs relatives à l'ensemble d'apprentissage sont représentées par des points.
 3.: Tensions de sortie idéales relatives aux surfaces testées ($V_{outNNid}$ - trait gras) et réelles (V_{outNN} - pointillés) du réseau de neurones.
 4.: Erreurs en valeur absolue (trait plein) et erreur maximale e_{max} .
 5.: Erreurs binaires.

III.4. CONCLUSION

Ainsi, d'après les précédentes simulations comportementales, l'utilisation d'un réseau de neurones au sein d'une application télémétrique permet d'accroître l'intervalle de mesure de distance sans perte de résolution et de classer différents types de surfaces sur des plages de mesure de distance et d'angle d'incidence variables. De plus, les structures neuronales choisies pour valider ces deux applications sont très proches l'une de l'autre, au nombre d'entrées près. La proximité de ces deux architectures rend possible l'implantation d'une seule et même structure, capable de répondre des deux problématiques soulevées. Une intégration électronique permettrait d'effectuer les tâches de levée d'indétermination sur la mesure de distance et de classification de surfaces directement au sein de l'application télémétrique.

C'est la raison pour laquelle il a été décidé de concevoir les cellules de base du perceptron multi-couches représenté figure III-m. Un faible encombrement et une faible consommation ont été privilégiés, de manière à valider la possibilité d'une intégration dans la tête optique du télémètre. Deux types d'intégration ont été réalisées puis comparées, l'une entièrement analogique et l'autre entièrement numérique. Le réseau de neurones analogique a été réalisé à partir de trois neurones élémentaires intégrés dans un ASIC. L'intégration des diverses cellules analogiques est explicitée au chapitre IV. Une comparaison avec l'intégration numérique est détaillée dans le chapitre V.

III.5. REFERENCES

- [Bosch, 1997] : BOSCH T., LESCURE M., Crosstalk analysis of 1 m to 10 m laser phase-shift range finder, *IEEE Transactions on Instrumentation and Measurement*, Volume 46, Issue 6, Dec. 1997, Page(s):1224 – 1228.
- [Dreyfus, 2004] : DREYFUS G., MARTINEZ J.M., SAMUELIDES M., GORDON M.B., BADRAN F., THIRIA S., HERAULT L., *Réseaux de neurones, méthodologie et applications, sous la direction de Gérard Dreyfus*, 2^{ème} édition, Eyrolles, avril 2004.
- [Gatet, 2006] : GATET L., TAP-BETEILLE H., LESCURE M., Raising of indecision of a Phase-Shift Laser Rangefinder thanks to feedforward MLP Neural Network, *Proceedings of the 7th WSEAS International Conference on NEURAL NETWORKS (NN'06)*, Cavtat, Croatia, June 12-14, 2006, Pages: 67-72
- [Gatet, 2007] : GATET L., TAP-BETEILLE H., LESCURE M., Surface Detection using Phase-Shift Laser Rangefinder and Neural Network", *Optical Engineering, Optical Sensors (44)*, Vol. 46, No. 7, July 2007, to be published.
- [Haykin, 1994] : HAYKIN S., Neural Network : A Comprehensive Foundation, *Prentice Hall*, 1994.
- [Journet, 2002] : JOURNET B., Time of flight range finding systems, *Seventh International Symposium on Laser Metrology Applied to Science, Industry, and Everyday Life, Proceedings of the SPIE*, Volume 4900, pp. 466-477 (2002).
- [Levenberg, 1944] : LEVENBERG K., A method for the solution of certain non-linear problems in least squares, *Quarterly Journal of Applied Mathematics*, 2, p.164-168.
- [Marquardt, 1963] : MARQUARDT D.W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society of Industrial and Applied Mathematics*, 11, p. 431-441.
- [Rumelhart, 1986] : RUMELHART D.E., HINTON G.E., WILLIAMS R.J., Learning internal representations by error backpropagation, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, p.318-362, MIT Press.
- [Werbos, 1974] : WERBOS P.J., *Beyond regression : new tools for prediction and analysis in the behavioural sciences*, Ph. D. thesis, Harvard University.

Chapitre IV

Conception du réseau de neurones analogique

Ce chapitre est centré essentiellement sur la conception des cellules élémentaires d'un réseau de neurones visant à être implanté dans les deux applications télémétriques.

IV.1. INTRODUCTION

La technologie choisie pour implanter le réseau de neurones est la technologie CMOS 0.35 μ m. La faible valeur de tension d'alimentation autorisée par cette technologie (3.3V) impose un travail important de conception au niveau de chaque cellule de manière à optimiser la dynamique des signaux. Cette contrainte basse tension fait que l'ensemble des cellules a été conçu de manière à fonctionner de préférence en transfert de courant, afin de limiter au maximum la distorsion non souhaitée des signaux. En effet, pour obtenir une bande passante relativement élevée, il est nécessaire a priori que les transistors fonctionnent en régime de forte inversion. De ce fait, les tensions de polarisation grille-source V_{GS} doivent être relativement importantes par rapport à la tension d'alimentation.

Afin d'améliorer la dynamique, il est intéressant de disposer de signaux différentiels en entrée de chacune de ces cellules afin que le dispositif fonctionne en symétrique. De plus, une implantation d'un réseau de neurones analogique s'avère très consommatrice en nombre d'entrées/sorties alors que la surface de silicium est moins contraignante. Ainsi, pour minimiser le nombre de broches, des paires différentielles à entrées asymétriques et à sorties symétriques ont été intégrées à chaque entrée et chaque signal différentiel de sortie est ramené à un signal asymétrique.

IV.2. CELLULES ANALOGIQUES EN TECHNOLOGIE CMOS BASSE TENSION

La conception de circuits intégrés analogiques précis et à bande passante élevée en technologies CMOS submicroniques basse tension nécessite la mise en œuvre de nouvelles architectures de circuits. Ainsi, une architecture optimisée pour des technologies à tension d'alimentation 5V ne pourra être utilisée telle quelle pour des technologies 3.3V et inférieures. En effet, la tension de seuil V_{th} des transistors MOS ne présente pas une décroissance proportionnelle aux tensions maximales admissibles d'alimentation. Une tension V_{GS} devient donc une fraction importante du budget en tension d'un circuit analogique. C'est la raison pour laquelle on analyse dans cette partie le budget minimal nécessaire en sortie du transistor MOS de manière à ce que celui-ci demeure en zone source de courant.

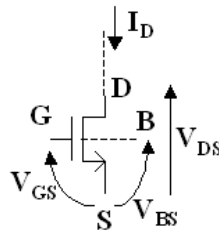


Figure IV-a : Schéma d'un transistor NMOS.

IV.2.1. Modèle niveau 1

Le courant drain I_D d'un transistor NMOS modélisé par le niveau 1 en régime forte inversion ($V_{GS} > V_{th}$), en zone source de courant ($V_{DS} > V_{GS} - V_{th}$), a pour expression ([Gray, 2001]):

$$I_D = \frac{\mu C_{ox}}{2} \left(\frac{W}{L}\right) (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \quad (IV-1)$$

avec μ la mobilité des porteurs (des électrons pour un transistor NMOS), C_{ox} la capacité de l'oxyde de grille par unité de surface, W/L le rapport de la largeur sur la longueur du canal du transistor et λ le facteur de modulation de la longueur du canal. Pour une technologie CMOS AMS 0.35 μ m, la tension seuil V_{th} a pour valeur typique 0.5V pour un transistor NMOS (0.6V pour un transistor PMOS). Le produit μC_{ox} est environ égal à 180 μ A/V² pour un transistor NMOS (60 μ A/V² pour un transistor PMOS). Par commodité d'écriture, on introduit le terme β :

$$\beta = \frac{\mu C_{ox}}{2} \left(\frac{W}{L}\right) \quad (IV-2)$$

Comme l'essentiel des transistors utilisés dans ce chapitre est de type NMOS, pour éviter d'alourdir la lecture, les notations utilisées au sein des formules sont relatives à celles des transistors NMOS

Introduisons la notation ΔV_G relative à la différence $V_{GS}-V_{th}$. Ainsi, on admettra qu'un transistor MOS reste en zone source de courant si $V_{DS}>\Delta V_G$. Il fonctionne en zone ohmique si $V_{DS}<\Delta V_G$.

IV.2.2. Schéma équivalent petit signal

La figure IV-b présente le schéma équivalent petit signal (modèle haute fréquence ([Geiger, 1990])) d'un transistor NMOS, en considérant que le potentiel du substrat (*bulk*) est différent de celui de la source d'un point de vue dynamique. Dans notre étude, le substrat des transistors NMOS et PMOS seront respectivement reliés au potentiel le plus faible (la masse) et au potentiel le plus élevé (V_{CC}).

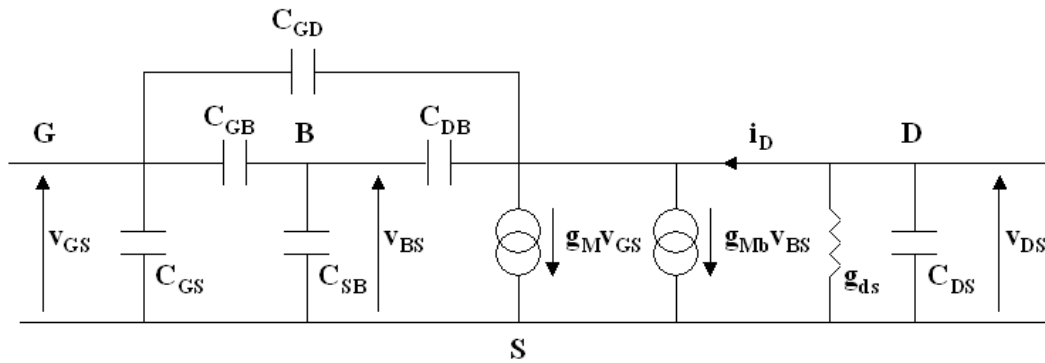


Figure IV-b : Modèle petit signal haute fréquence du transistor MOS en zone source de courant.

La transconductance g_M est liée à la modulation petit signal du courant drain I_D par la variation de la tension V_{GS} en prenant les tensions V_{DS} et V_{BS} constantes. En régime de forte inversion, l'expression analytique de la transconductance est :

$$g_M = \left(\frac{\partial I_D}{\partial V_{GS}} \right)_{V_{DS}=\text{cste et } V_{BS}=\text{cste}} = \sqrt{4\beta I_D} \quad (\text{IV-3})$$

La transconductance g_{Mb} est liée à la modulation petit signal du courant drain I_D par la tension V_{BS} en prenant les tensions V_{GS} et V_{DS} constantes.

La conductance de sortie g_{ds} est liée à la modulation λ de la longueur du canal par la tension V_{DS} pour des tensions V_{GS} et V_{BS} constantes. L'expression analytique de la transconductance est :

$$g_{ds} = \left(\frac{\partial I_D}{\partial V_{DS}} \right)_{V_{GS}=\text{cste et } V_{BS}=\text{cste}} \cong \lambda I_D \quad (\text{IV-4})$$

Le paramètre λ sera considéré au premier ordre comme constant. Sa valeur est extraite graphiquement par extrapolation de la caractéristique statique $I_D(V_{DS})$ à V_{GS} et V_{BS} constantes en zone source de courant. A titre d'exemple, la figure IV-c montre l'évolution de la caractéristique statique $I_D(V_{DS})$ d'une technologie $0.35\mu\text{m}$, à V_{GS} et V_{BS} constantes (respectivement 1V et 0V), pour différentes longueurs de canal, avec un rapport W/L constant de 10. On note qu'une diminution de la longueur de canal entraîne une pente de plus en plus importante, donc une tension $V_A=1/\lambda$ de plus en plus faible à l'intersection de la tangente à la caractéristique avec l'axe des abscisses. Ainsi, une diminution de la longueur de canal entraîne une augmentation de la conductance de sortie g_{ds} . Des exemples de valeurs de conductance de sortie g_{ds} associées ont été regroupées dans le tableau IV- α .

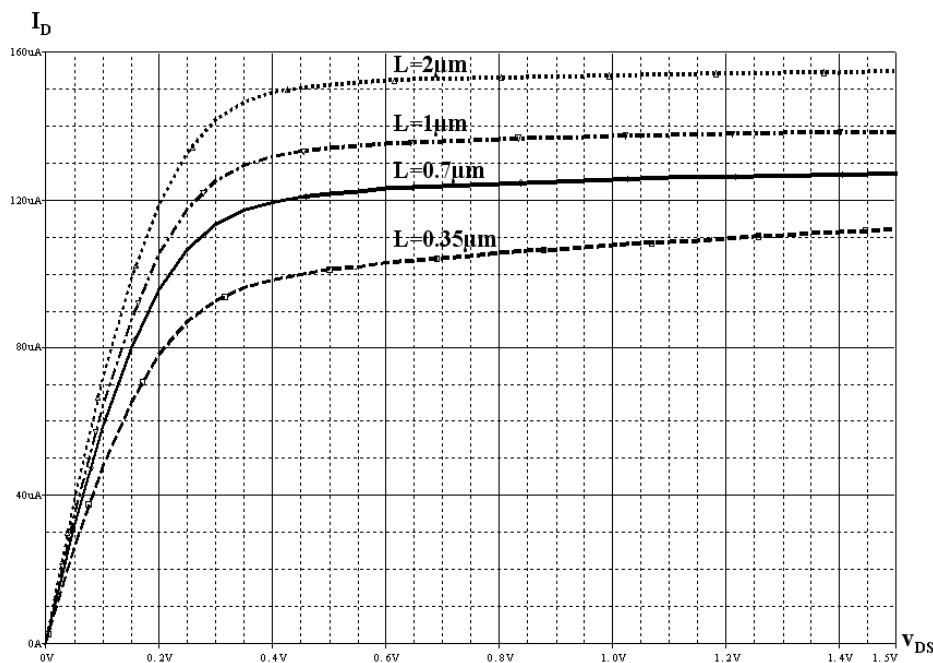


Figure IV-c : Evolution de la caractéristique $I_D(V_{DS})$ d'un transistor MOS en technologie $0.35\mu\text{m}$ pour différentes longueurs de canal, avec un rapport W/L constant de 10 et pour un V_{GS} constant égal à 1V.

Tableau IV- α : Variation de la conductance de sortie g_{ds} pour différentes longueurs de canal, avec un rapport W/L constant de 10 et pour $V_{GS} = 1\text{V}$ et $V_{DS} = 1\text{V}$.

L (μm)	0.35	0.7	1	2
g_{ds} (μS)	10.2	4.4	1.7	1.3

Enfin, au premier ordre, les capacités grille-substrat (C_{GB}), grille-drain (C_{GD}) et grille-source (C_{GS}) sont fonctions des dimensions géométriques du transistor ([Geiger, 1990], [Antognetti, 1993]). Ainsi, à même W/L , plus le transistor est long, plus la valeur des différentes capacités associées est importante, ce qui entraîne une limitation de la bande passante.

IV.2.3. Transfert par miroir de courant

IV.2.3.1. Introduction

Le miroir de courant est l'une des cellules de base des circuits intégrés analogiques. Ils imposent très souvent la polarisation (ou point de fonctionnement au "repos") des circuits. Le budget de tension étant souvent faible, il est préférable d'utiliser au maximum la plage de tension admissible en sortie. Les principaux critères de bon fonctionnement d'un miroir de courant sont les suivants :

- un facteur de copie en courant précis et constant,
- une plage de tension admissible en sortie la plus grande possible,
- une faible résistance dynamique d'entrée,
- une faible conductance dynamique de sortie,
- une réponse en fréquence élevée du courant de court-circuit en sortie.

Les deux sous-parties suivantes présentent les deux structures de miroir de courant les plus communes : le miroir de courant simple et le miroir cascodé.

IV.2.3.2. Miroir de courant simple

Le miroir de courant le plus simple est représenté figure IV-d. Le transistor M_{N01} est câblé en "transdiode". Comme sa tension drain-source est égale à sa tension grille-source, la condition $V_{DS} > \Delta V_G$ est bien respectée. Ce transistor opère toujours en zone source de courant (dite zone de saturation). L'équation d'état (IV-1) est donc toujours valable.

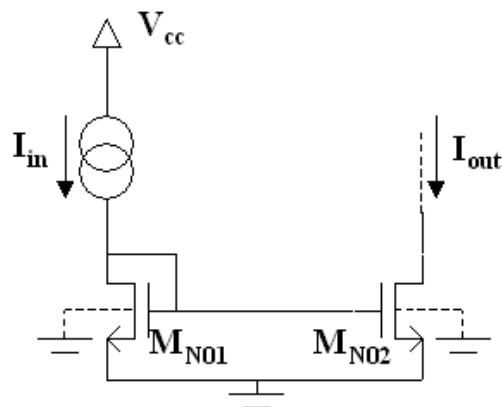


Figure IV-d : Miroir de courant simple.

Afin d'assurer une recopie fidèle en courant, le transistor M_{N02} doit aussi opérer en zone source de courant. D'après l'équation (IV-1), on peut déduire l'expression du transfert en courant du miroir au premier ordre :

$$\frac{I_{out}}{I_{in}} = \frac{W_2/L_2}{W_1/L_1} \frac{1+\lambda_2 V_{DS2}}{1+\lambda_1 V_{DS1}} \quad (IV-5)$$

Pour obtenir un courant de sortie I_{out} égal au courant d'entrée I_{in} , les tailles des transistors M_{N01} et M_{N02} sont prises identiques ($W_2/L_2 = W_1/L_1$). Pour réduire l'influence de la tension V_{DS} des deux transistors apparaissant dans le second terme, une solution simple consiste à choisir des transistors de canal long, de manière à minimiser la valeur de paramètre λ .

Ce type de miroir présente la plus grande plage de tension admissible en sortie. Cette plage n'est limitée que par la polarisation du transistor M_{N02} en zone source de courant. La conductance de sortie du miroir de courant est donnée par l'équation (IV-4) :

$$g_{out} = g_{ds2} = \lambda_2 I_{out} \quad (IV-6)$$

Nous avons fait l'hypothèse que le substrat des transistors est connecté à la masse au point de vue régime dynamique ($v_{BS}=0V$). La figure IV-e présente le schéma équivalent petit signal du miroir de courant simple. La conductance G_L et la capacité C_L de sortie ont été représentées. En régime petit signal, le transistor M_{N01} câblé en transdiode présente une forte valeur de conductance dynamique ($g_{M1}+g_{ds1}$) grâce à la contre réaction parallèle entre drain et grille.

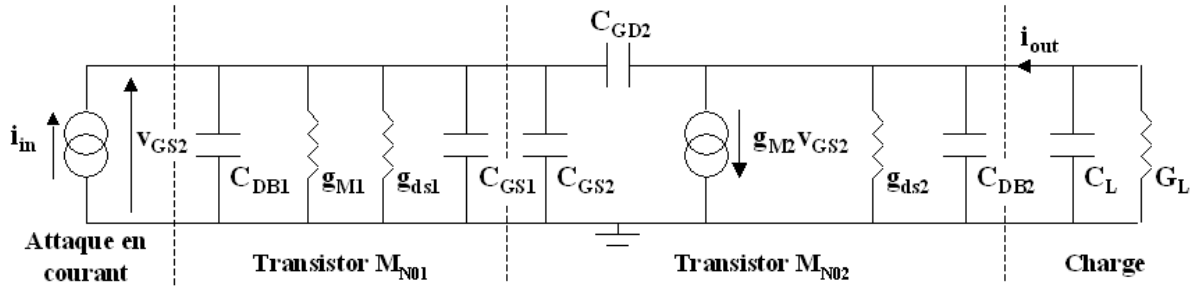


Figure IV-e : Schéma équivalent petit signal du miroir de courant simple.

Le transfert en courant peut s'écrire sous la forme :

$$\frac{i_{out}(p)}{i_{in}} = \frac{W_2/L_2}{W_1/L_1} \frac{1}{1+\tau_1 p} \frac{1}{1+\tau_2 p} \quad (IV-7)$$

avec τ_1 et τ_2 les constantes de temps associées respectivement aux nœuds d'entrée et de sortie, en utilisant l'approximation de Miller pour M_{N02} :

$$\tau_1 = \frac{C_{DB1} + C_{GS1} + C_{GS2} + C_{GD2} [1 + g_{M2}/(g_{ds2} + G_L)]}{g_{ds1} + g_{M1}} \quad (IV-8)$$

$$\tau_2 = \frac{C_{DB2} + C_{GD2} + C_L}{g_{ds2} + G_L} \quad (IV-9)$$

Remarque 1 : Les relations (IV-8) et (IV-9) seront admises avec l'hypothèse de la hiérarchie des constantes de temps $\tau_1 \gg \tau_2$.

Remarque 2 : On note que souvent, $g_{M1} \gg g_{ds1}$ car le rapport g_{M1}/g_{ds1} , correspondant au gain en tension intrinsèque d'une source commune, est de l'ordre de 30 ([Josse, 2003]).

Si la charge du transistor M_{N02} est une transdiode M_{P03} ($G_L \cong 1/g_{m3}$), la fréquence de coupure associée à l'entrée s'écrit :

$$f_1 = \frac{1}{2\pi\tau_1} \cong \frac{g_{M1}}{2\pi(C_{DB1} + C_{GS1} + C_{GS2} + C_{GD2}(1 + g_{M2}/g_{M3}))} \quad (IV-10)$$

Cela nous amène à la limite intrinsèque du miroir de courant simple, qui correspond à une fréquence de coupure produit gain – bande passante au premier ordre du transistor.

Comme nous l'avons vu précédemment, un transistor de sortie ayant une grande longueur de canal permet d'augmenter la précision de la recopie du courant par une réduction de la conductance de sortie du miroir. En revanche, à même W/L, plus ce transistor est long, plus la taille des différentes capacités associées est importante, ce qui entraîne une limitation de la bande passante.

IV.2.3.3. Miroir de courant cascodé

Le miroir de courant cascodé est une alternative largement utilisée (figure IV-f). Pour améliorer la bande passante du miroir de courant simple, tout en conservant une bonne recopie de courant et une grande résistance de sortie, un second transistor M_{N04} monté en grille commune reçoit les variations de la tension de sortie tandis que le transistor M_{N02} se charge de la bonne recopie du courant.

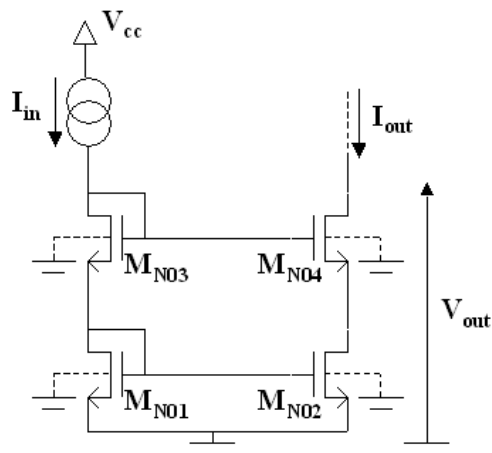


Figure IV-f: Miroir de courant cascodé.

L'expression du transfert en courant est identique à celle du miroir de courant simple (équation (IV-5)). La précision de la recopie est nettement améliorée, car la condition $V_{DS1}=V_{GS1}=V_{DS2}$ est satisfaite. Pour déterminer la conductance de sortie, on utilise le schéma équivalent représenté sur la figure IV-g.

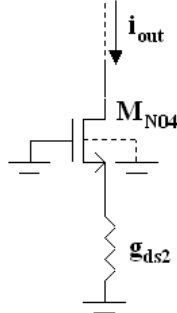


Figure IV-g : Schéma équivalent pour la détermination de la conductance de sortie du miroir cascode.

La contre-réaction série sur la grille commune (M_{N04}) donne l'expression de la conductance de sortie g_{out} :

$$g_{out} = \frac{g_{ds4}}{1+(g_{M4}/g_{ds2})} \quad (IV-11)$$

Prenons comme valeur de référence le gain en tension intrinsèque d'un circuit source commune ayant une conductance de charge extérieure nulle : $|g_M/g_{ds}| \cong 30$ ([Josse, 2003]). La structure cascode dont les longueurs de canal sont voisines de la longueur minimale permet d'obtenir une conductance de sortie environ 30 fois plus faible qu'un miroir de courant simple. De plus, la diminution des tailles des transistors entraîne des tailles de capacités associées plus faibles, un effet Miller réduit, et donc une augmentation de la bande passante du transfert en courant.

En revanche, le principal inconvénient de cette structure est la limite de la plage de tension admissible en sortie. En effet, la polarisation en sortie du miroir doit respecter la condition de saturation $V_{DS} > \Delta V_G$ des transistors. Ainsi, pour un miroir de courant cascodé (figure IV-f), la tension de sortie V_{out} du miroir doit être supérieure à $V_{out_mini} \cong 2\Delta V_G + V_{th}$.

Différentes méthodes existent de manière à accroître l'excursion en tension de la sortie du miroir de courant cascode. Leur principe commun est de faire en sorte que le transistor M_{N02} soit placé à la limite de la zone source de courant ([Sooch, 1985], [Babanezhad, 1987]). Pour cela, on fixe la tension de la grille du transistor de sortie M_{N04} à la polarisation $2\Delta V_G + V_{th}$. Cette méthode permet donc de diminuer d'un V_{th} la tension grille du transistor de sortie M_{N04} . La tension de sortie V_{out} du miroir ne doit alors être supérieure qu'à $V_{out_mini} \cong 2\Delta V_G$.

IV.2.3.4. Discussion

Le choix de la structure de miroir à utiliser dépend essentiellement des critères de précision de recopie, de budget de tension admissible en sortie et de bande passante. Les structures cascodées améliorées à large excursion permettent de satisfaire tous ces critères. En terme de précision de recopie, pour un miroir de courant simple (figure IV-d) avec une transdiode pour charge extérieure, l'erreur de recopie en régime petit signal est de l'ordre de g_{ds}/g_M , soit environ 1/30, alors qu'elle peut être évaluée à $(g_{ds}/g_M)^2$, soit environ 1/900, pour un miroir de courant cascodé (figure IV-f).

Néanmoins, les structures cascodées sont plus complexes à mettre en oeuvre. De plus, elles nécessitent plusieurs courants de polarisation supplémentaires. En terme de simplicité, de précision de recopie et de budget de tension admissible en sortie, le miroir de courant simple utilisant des transistors relativement longs (de l'ordre de deux fois la longueur minimale) présente des performances suffisantes, avec une erreur de recopie de l'ordre de 3%.

C'est la raison pour laquelle cette structure a été choisie, malgré l'inconvénient lié à la limitation de la bande passante. Un soin tout particulier a été apporté lors du choix des tailles des transistors de sortie de manière à optimiser le budget de tension admissible en sortie le plus important possible tout en ne bridant pas trop la bande passante globale du circuit.

IV.2.4. Circuit différentiel à sources couplées

IV.2.4.1. Introduction

Le circuit différentiel remplit l'une des fonctions les plus importantes de l'électronique analogique. Il permet de réaliser, commodément, l'amplification à liaisons directes des signaux continus ou très lentement variables en fonction du temps. Il est présent dans de nombreuses fonctions telles que les amplificateurs opérationnels, les comparateurs, les amplificateurs vidéo ou à large bande passante, les régulateurs de tension, les circuits comprenant des contrôles automatiques de gains, etc. Le circuit différentiel est aussi très utilisé dans de nombreuses applications de télécommunications analogiques sous la forme de cellules de Gilbert ([Gray, 2001], [Gatet, 2006(a)], [Gatet, 2007(a)]). Cette cellule permet de réaliser les fonctions de multiplication, de modulation d'amplitude, ou encore de comparaison de phases. Le circuit différentiel est l'un des éléments de base des circuits neuronaux que l'on retrouve par exemple dans la fonction multiplication par un poids qui apparaît entre chaque neurone [Gatet, 2006(b)].

Le circuit différentiel permet de réaliser des circuits à liaisons directes. La séparation entre les régimes statiques et dynamiques n'apparaît pas, de ce fait, aussi nettement que dans les circuits à couplages capacitifs. Il est alors indispensable de distinguer les conditions de fonctionnement linéaire et non-linéaire de ce circuit, non seulement en fonction de la valeur de l'amplitude du signal, mais aussi en fonction de la composante moyenne qui influence ses points de fonctionnement.

Les deux tensions v_1 et v_2 appliquées sur les grilles des transistors M_{N11} et M_{N12} (figure IV-h) imposent de définir deux types de paramètres :

♦ ceux du mode différentiel auquel correspond une réponse de l'amplificateur dépendant de la seule tension différentielle d'entrée v_d :

$$v_d = v_1 - v_2 \quad (IV-12)$$

♦ ceux du mode commun auquel correspond une réponse de l'amplificateur dépendant de la seule tension de mode commun v_M :

$$v_M = \frac{v_1 + v_2}{2} \quad (IV-13)$$

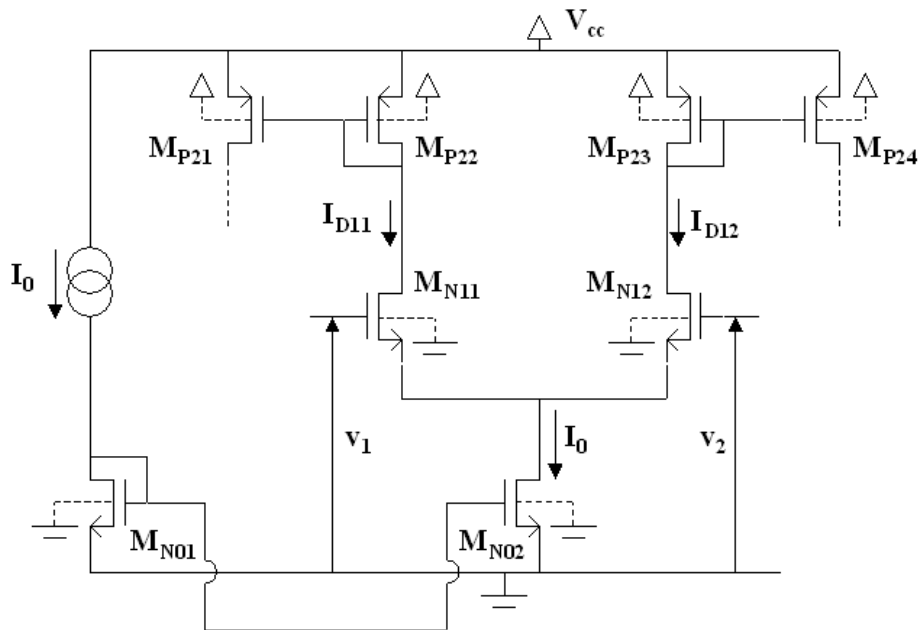


Figure IV-h : Circuit différentiel à sources couplées.

Le signal de sortie d'un circuit différentiel ne doit généralement dépendre que de la tension différentielle v_d qui constitue de ce fait le signal utile. L'influence du mode commun, représenté par v_M , doit être aussi faible que possible. On la caractérise par le rapport de réjection de la tension de mode commun (RRMC), relatif au rapport entre le gain en tension différentiel et le gain en tension de mode commun.

Le but de cette étude est d'optimiser le domaine de linéarité des paires différentielles impliquées dans chaque multiplieur (cellule de Gilbert repliée) et de contrôler la zone de "compression" de la fonction tangente hyperbolique. Pour chacune des cellules, nous nous intéresserons à l'équation de transfert liant le courant différentiel de sortie aux différents paramètres d'entrée.

IV.2.4.2. Transfert du circuit différentiel

Equation de transfert globale :

Les transistors M_{N11} et M_{N12} de la paire différentielle sont supposés identiques. Posons β_{diff} le terme correspondant à :

$$\beta_{diff} = \frac{\mu C_{ox}}{2} \left(\frac{W}{L}\right)_{diff} \quad (IV-14)$$

Supposons que l'on travaille en régime quasi-statique et que l'ensemble des transistors de la figure IV-h fonctionnent en régime de forte inversion en zone source de courant ($V_{DS} > \Delta V_G$). Par l'intermédiaire de l'équation (IV-1), en supposant une symétrie rigoureuse, la tension différentielle de sortie v_d du circuit différentiel à sources couplées s'écrit :

$$v_d = V_{GS1} - V_{GS2} = \sqrt{\frac{I_{D11}}{\beta_{diff}}} - \sqrt{\frac{I_{D12}}{\beta_{diff}}} \quad (IV-15)$$

Les deux courants drain I_{D11} et I_{D12} des transistors M_{N11} et M_{N12} sont aussi liés par la présence de la source de courant I_0 :

$$I_{D11} + I_{D12} = I_0 \quad (IV-16)$$

On en déduit les variations des courants drain I_{D11} et I_{D12} ([Geiger, 1990]) :

$$I_{D11} = \frac{I_0}{2} + \frac{I_0}{2} \sqrt{\frac{2\beta_{diff} v_d^2}{I_0} - \frac{\beta_{diff}^2 v_d^4}{I_0^2}} \quad (IV-17)$$

$$I_{D12} = \frac{I_0}{2} - \frac{I_0}{2} \sqrt{\frac{2\beta_{diff} v_d^2}{I_0} - \frac{\beta_{diff}^2 v_d^4}{I_0^2}} \quad (IV-18)$$

soit la variation du courant différentiel de sortie ΔI_{D1} suivante :

$$\Delta I_{D1} = I_{D11} - I_{D12} = I_0 \sqrt{\frac{2\beta_{diff} v_d^2}{I_0} - \frac{\beta_{diff}^2 v_d^4}{I_0^2}} = v_d \sqrt{2\beta_{diff} I_0} \sqrt{1 - \frac{\beta_{diff} v_d^2}{2I_0}} \quad (IV-19)$$

Le terme contenu sous la racine carrée devant être positif, l'équation (IV-19) n'est valable que pour $v_d < \sqrt{\frac{2I_0}{\beta_{diff}}}$.

Equation de transfert petit signal :

Avec l'hypothèse d'une équation d'état $I_D(V_{GS})$ parabolique du transistor MOS en régime de forte inversion, on peut admettre que le courant différentiel de sortie ΔI_{D1} est linéaire par rapport à la tension différentielle v_d si $\beta_{diff} v_d^2 / 2I_0 \ll 1$. Ainsi, pour $v_d \ll \sqrt{\frac{2I_0}{\beta_{diff}}}$, nous obtenons, en régime linéaire ("petit signal") :

$$\Delta i_{D1} = i_{D11} - i_{D12} = \sqrt{4\beta_{diff} \frac{I_0}{2}} v_d \equiv G_M v_d \quad (IV-20)$$

avec G_M la transconductance de la fonction, qui correspond dans ce cas à la transconductance du transistor M_{N11} ou M_{N12} ($G_M = g_{M11} = g_{M12}$ (IV-3)).

Equation de transfert retenue :

L'équation (IV-19) peut être relativement lourde à utiliser. Pour éviter d'ajouter des contraintes algorithmiques au niveau de l'extraction de la fonction de transfert de chaque cellule, l'expression du courant différentiel de sortie ΔI_{D1} sera utilisée par la suite sous la forme d'une fonction tangente hyperbolique, valable pour toute valeur de v_d :

$$\Delta I_{D1} = I_0 \tanh\left(\frac{G_M v_d}{I_0}\right) \quad (IV-21)$$

Ainsi, à petit signal, on retrouve l'équation (IV-20), tandis qu'à fort signal, le courant différentiel de sortie est bien limité à $\pm I_0$. L'expression (IV-21) est issue en toute rigueur d'une loi exponentielle $I_D(V_{GS})$ que l'on retrouve en régime de faible inversion du MOS et dans la caractéristique de transfert $I_C(V_{BE})$ du transistor bipolaire ([Gray, 2001]). La figure IV-i représente la comparaison du transfert ΔI_{D1} "exact" d'une paire différentielle en fonction de v_d et le tracé de la fonction relative à l'équation (IV-21). On remarque que les deux courbes sont quasiment superposées.

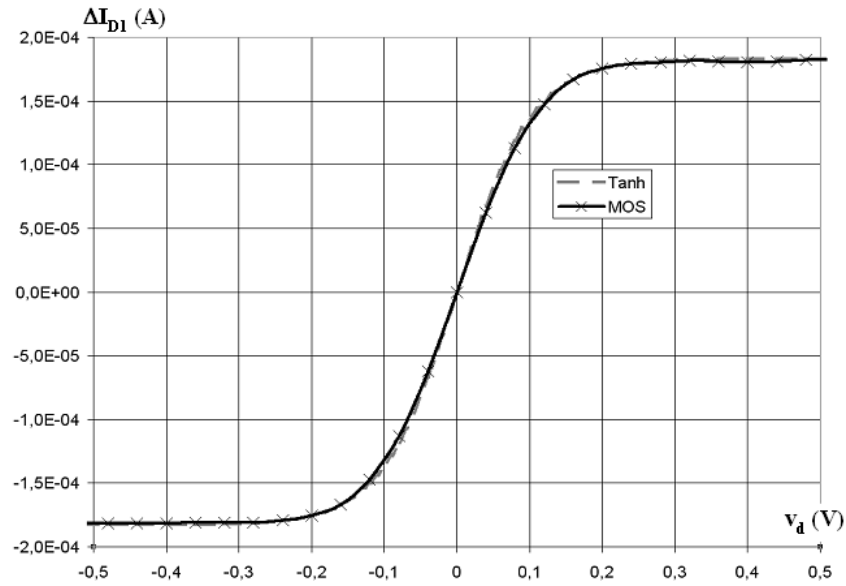


Figure IV-i : Comparaison entre la caractéristique de transfert "exacte" de la paire différentielle NMOS ($I_0=180\mu\text{A}$, $W/L=35\mu\text{m}/1\mu\text{m}$) et la fonction relative à l'équation (IV-21) avec $G_M = 1.5\text{mS}$. Les deux courbes sont quasiment superposées.

IV.2.4.3. Domaine de linéarité du mode différentiel

En règle générale, le critère retenu est l'éloignement en terme d'écart relatif de la caractéristique de transfert par rapport à une caractéristique de transfert linéaire. La caractéristique de transfert linéaire la plus communément choisie correspond à celle dont le coefficient directeur est égal à la pente à l'origine de la caractéristique correspondant à la transconductance $G_M = g_{M11}$ (droite 1 - figure IV-j).

Une deuxième méthode consiste à tracer une droite entre le point relatif à $v_d = 0\text{V}$ et le point pour lequel le courant différentiel de sortie a atteint un certain pourcentage (90%, 95% ou 99%) du courant I_0 (droite 2 - figure IV-j).

Une autre méthode consiste à choisir une caractéristique de transfert linéaire relativement proche en terme de pourcentage de la caractéristique de transfert, avec une erreur relative choisie par l'utilisateur (droite 3 - figure IV-j).

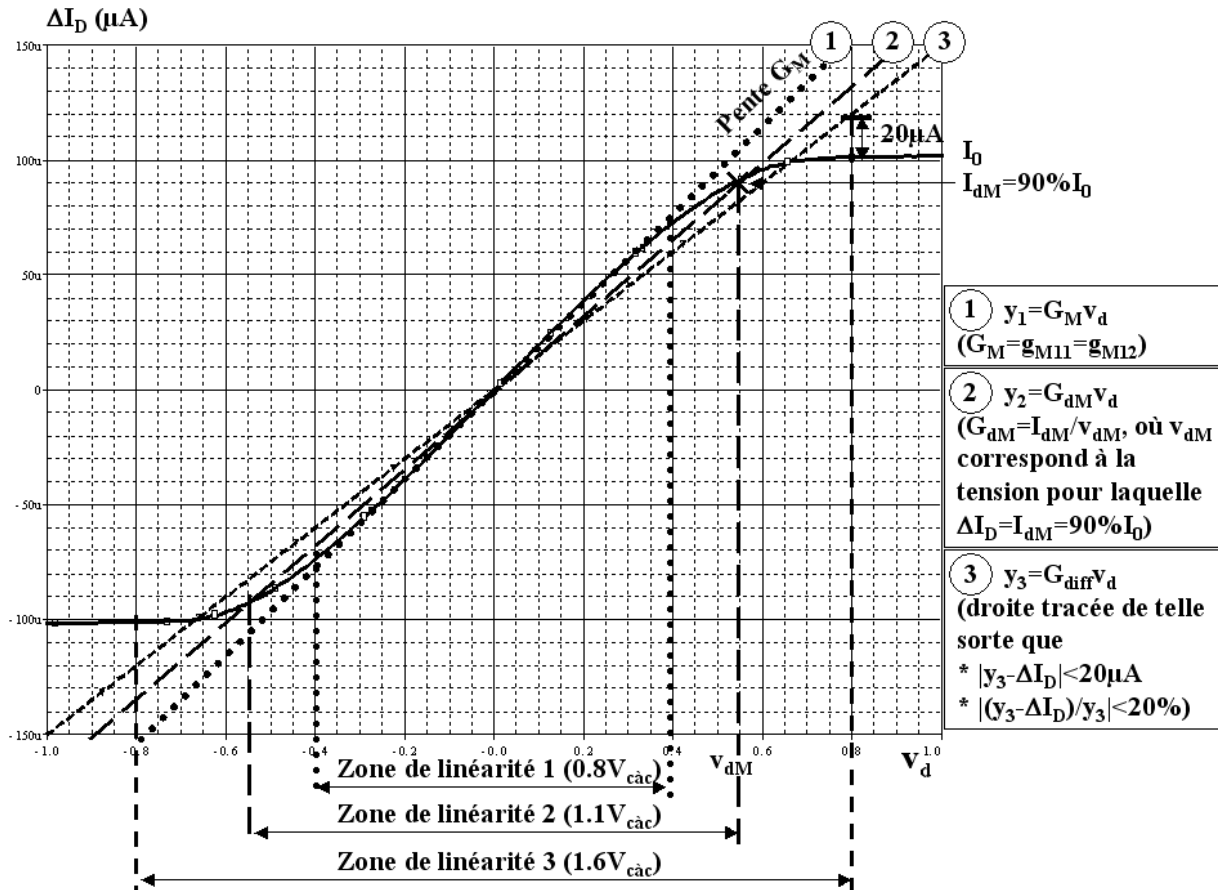


Figure IV-j : Domaine de linéarité du transfert d'une paire différentielle NMOS, avec $W/L = 2\mu\text{m}/1\mu\text{m}$ et $I_0 = 100\mu\text{A}$.

En toute rigueur, la condition de linéarité correspond à $v_d \ll \sqrt{\frac{2I_0}{\beta_{\text{diff}}}}$.

Néanmoins, quelle que soit la méthode de détermination de la zone de linéarité choisie, les coefficients directeurs des différentes droites sont proportionnels à la transconductance

$G_M = \sqrt{4\beta_{\text{diff}} \frac{I_0}{2}}$. Donc plus celle-ci est faible, plus la zone de linéarité sera importante.

Dans la pratique, pour obtenir un domaine de linéarité important, il est souhaitable d'utiliser une forte densité de courant et une paire différentielle de type PMOS (le produit μC_{ox} d'un transistor PMOS étant environ trois fois plus faible que celui d'un NMOS).

IV.2.4.4. Domaine admissible de la tension de mode commun

Selon le schéma de la figure IV-h, la tension admissible de mode commun intervient :

- par valeur inférieure lorsque les transistors M_{N02} et M_{N11} sont à la limite de la zone source de courant,

- par valeur supérieure lorsque les transistors M_{N11} - M_{N12} de la paire différentielle sont à la limite de la zone ohmique, les tensions V_{GS} des transdiodes canal P demeurant quasi constantes.

Les tensions admissibles de mode commun inférieure et supérieure sont donc données par (figure IV-k) :

$$v_{M\text{mini}} = V_{GS11\text{mini}} + \Delta V_{G02} \cong V_{th} + 2\Delta V_G \quad (\text{IV-22})$$

$$v_{M\text{maxi}} = V_{GS11} + (V_{CC} - |V_{GS22}|) - \Delta V_{G11} \cong V_{CC} - \Delta V_G \quad (\text{IV-23})$$

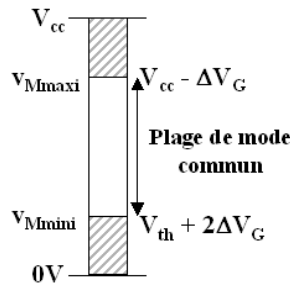


Figure IV-k : Plage admissible de la tension de mode commun v_M du circuit figure IV-h, avec l'hypothèse d'une très faible amplitude de la tension différentielle v_d .

Dans ces relations, nous n'avons pas pris en compte le fait que les tensions de seuil V_{th} des transistors de la paire différentielle augmentent par effet de substrat avec la tension de mode commun v_M ([Geiger, 1990]). Les relations (IV-22) et (IV-23) montrent que pour optimiser la plage admissible de la tension de mode commun, il est nécessaire de réduire la valeur ΔV_G des transistors M_{N02} , M_{N11} et M_{P22} :

$$\Delta V_G = V_{GS} - V_{th} = \sqrt{\frac{I_D}{\beta}} = \sqrt{\frac{I_D}{\mu C_{ox} \frac{W}{L}}} \quad (\text{IV-24})$$

Dans la pratique, cette condition est respectée pour une faible densité de courant par μm de largeur de canal et pour une paire différentielle de type NMOS. Il faut donc que le transistor source de courant M_{N02} et les transistors M_{P22} et M_{P23} soient larges avec un courant relativement faible.

IV.2.4.5. Circuit mono-tension

Une fois intégré, le circuit neuronal sera alimenté en mono-tension positive ($V_{CC} = 3.3V$), le potentiel de référence étant imposé par la masse. Il est particulièrement intéressant de faire en sorte que les tensions d'entrée soient appliquées à liaison directe par rapport à 0V. On réalise ce décalage par l'intermédiaire de deux sources suiveuses canal P appliquées sur les grilles des transistors M_{N11} et M_{N12} (figure IV-1).

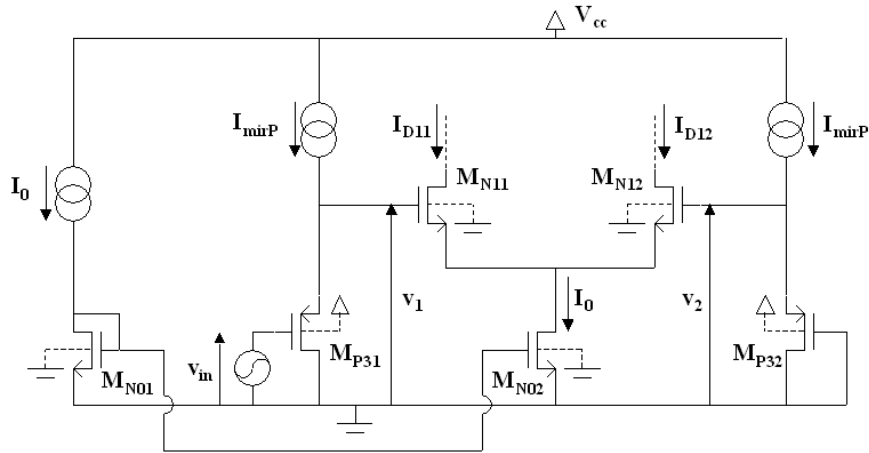


Figure IV-1: Sources suiveuses canal P associées à au circuit différentiel NMOS à sources couplées. Les sources suiveuses M_{P11} et M_{P12} permettent d'obtenir une tension de mode commun admissible de 0V.

Deux limitations interviennent au niveau de la valeur minimale v_{in_mini} de la plage admissible de mode commun:

- celle liée au transfert :

$$v_{in_mini} + |V_{GS31}| = V_{GS11} + \Delta V_{G02} \quad (IV-25)$$

celle liée à la condition de fonctionnement en zone source de courant du transistor M_{P31} :

$$v_{in_mini} + |V_{GS31}| = \Delta V_{G31} \quad (IV-26)$$

La première limitation montre qu'une forte densité de courant au niveau des sources suiveuses permet de réduire la valeur de la tension v_{in_mini} . D'après l'équation (IV-26), cette tension ne pourra être inférieure à $-V_{th}$. Notons qu'une paire différentielle canal P associée à deux sources suiveuses canal P aurait permis d'obtenir une tension v_{in_mini} encore plus négative.

Ce dispositif permet de réaliser facilement la conversion asymétrique-symétrique du signal de contrôle. Avec la grille de M_{P32} au potentiel de référence, l'expression de la tension différentielle v_d est très proche de celle de la tension v_{in} appliquée à l'entrée :

$$v_d = v_1 - v_2 = \frac{g_{M11}}{1 + \frac{g_{ds11} + g_{dsmirP}}{g_{M11}}} v_{in} \cong v_{in} \quad (IV-27)$$

IV.2.5. Multiplication par un poids référencé à la masse

IV.2.5.1. Principe

La cellule utilisée pour effectuer la multiplication est une double paire différentielle (figure IV-m). Ces deux paires différentielles sont connectées de la même façon que celles utilisées pour une cellule de Gilbert repliée ([Gray, 2001]).

L'objectif est d'obtenir un courant différentiel de sortie $\Delta I_{Dmult} = I_{Dmult1} - I_{Dmult2}$ proportionnel au courant différentiel d'entrée $\Delta I_{in} = I_{in1} - I_{in2}$ et à la tension différentielle w , correspondant au poids. Le courant différentiel ΔI_{in} provenant de la cellule précédente est multiplié par 5 par l'intermédiaire des miroirs de courants $M_{N01}-M_{N02}$ et $M_{N03}-M_{N04}$ ([Gatet, 2007(b)]). La figure IV-n présente le schéma bloc d'un multiplieur avec l'ensemble des cellules utilisées pour mettre en forme les signaux d'entrée.

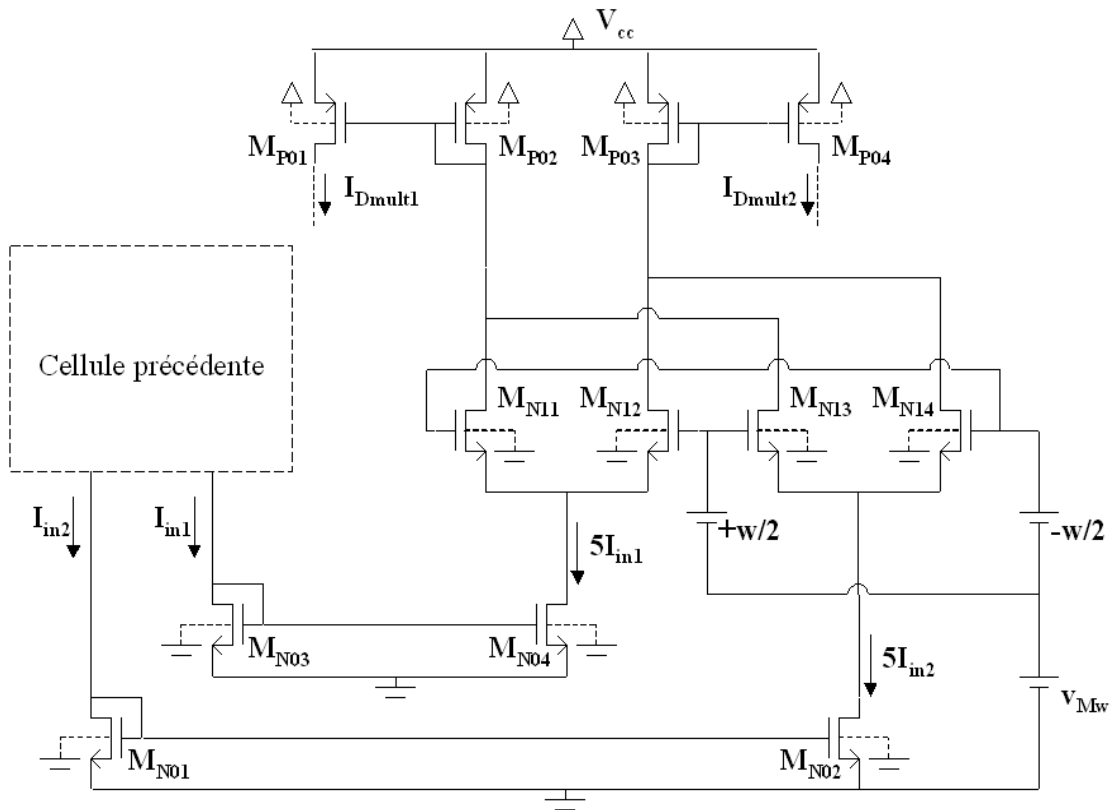


Figure IV-m : Schéma du multiplieur.

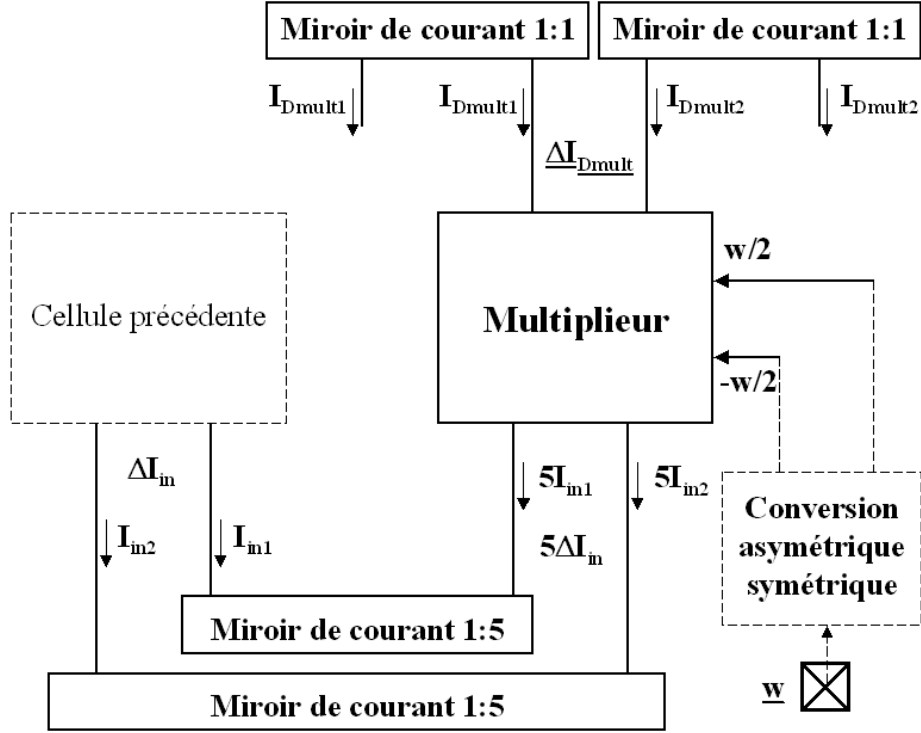


Figure IV-n : Schéma-bloc du multiplieur. Le poids w est converti en poids différentiel $\pm w/2$. Le courant différentiel d'entrée ΔI_{in} correspond au courant différentiel de sortie de la cellule précédente.

D'après l'équation (IV-19), en supposant que les quatre transistors M_{N11} , M_{N12} , M_{N13} et M_{N04} sont identiques, le courant différentiel de sortie s'écrit :

$$\Delta I_{Dmult} = I_{Dmult1} - I_{Dmult2} = w \left[\sqrt{2\beta_{diff} 5I_{in1}} \sqrt{1 - \frac{\beta_{diff} w^2}{10I_{in1}}} + \sqrt{2\beta_{diff} 5I_{in2}} \sqrt{1 - \frac{\beta_{diff} w^2}{10I_{in2}}} \right] \quad (IV-28)$$

Le courant différentiel d'entrée ΔI_{in} du multiplieur est multiplié par 5 de manière à obtenir pour un même poids w une amplitude du courant différentiel plus élevée.

Le courant différentiel $\Delta I_{in} = I_{in1} - I_{in2}$ étant généré par la cellule précédente, posons, comme dans le cadre de l'étude de la paire différentielle (figure IV-1):

$$I_{in1} + I_{in2} = I_0 \quad (IV-29)$$

avec I_0 le courant de polarisation de la cellule précédente. En se plaçant dans l'hypothèse de linéarité où $\beta_{diff} w^2 / 10I_0 \ll 1$, l'équation (IV-28) devient :

$$\Delta I_{Dmult} = w \sqrt{10\beta_{diff}} \left[\sqrt{I_{in1}} + \sqrt{I_{in2}} \right] \quad (IV-30)$$

A partir des équations (IV-29) et (IV-30), le courant différentiel de sortie "petit signal" ΔI_{Dmult} peut s'écrire sous la forme ([Gatet, 2007(b)]) :

$$\Delta I_{Dmult} = K_{mult} w \Delta I_{in} \quad (IV-31)$$

avec

$$K_{\text{mult}} = \sqrt{\frac{5\beta_{\text{diff}}}{I_0}} \quad (\text{IV-32})$$

Ainsi, pour de faibles amplitudes, le courant différentiel de sortie Δi_{Dmult} est proportionnel au produit du courant différentiel d'entrée Δi_{in} et du poids w . Pour des facilités de calcul au niveau du test, on s'impose d'avoir un courant différentiel de sortie Δi_{Dmult} égal au courant différentiel d'entrée Δi_{in} pour un poids w égal à 0.1V, soit, un coefficient de proportionnalité $K_{\text{mult}}=10\text{V}^{-1}$.

Parmi les trois "zones de linéarité" décrites dans la paragraphe IV.2.2.2 (figure IV-j), celle qui a été choisie est la troisième. En effet, l'objectif à court terme est d'entraîner le réseau à l'aide du logiciel Matlab et de transposer les valeurs des paramètres issus de la simulation au réseau de neurones intégré. Pour cela, il faut extraire du test expérimental du réseau intégré l'équation linéaire la plus "proche" de la caractéristique de transfert de la cellule, afin de pouvoir entraîner le réseau avec l'équation du réseau intégré. La zone de linéarité des sorties de chaque cellule par rapport aux entrées doit être la plus grande possible, de manière à pouvoir élargir au maximum les plages de variations des paramètres utilisés pour calibrer le réseau. La troisième façon de définir la zone de linéarité explicitée dans le paragraphe IV.2.2.2 permet à l'utilisateur de définir la précision souhaitée et d'avoir une dynamique associée aux tensions d'entrée et aux paramètres plus importante.

IV.2.5.2. Application au multiplieur quatre quadrants

La figure IV-o représente les résultats de simulation associés à un multiplieur situé entre une entrée du réseau de neurones et un neurone de la couche cachée. Le courant différentiel d'entrée Δi_{in} du multiplieur provient de la conversion de la tension d'entrée v_{in} en courant différentiel (figure IV-l). D'après les équations (IV-20) et (IV-27), en régime petit signal :

$$\Delta i_{\text{in}} \cong G_M v_{\text{in}} \quad (\text{IV-33})$$

Les tailles des transistors NMOS et le courant de polarisation I_0 sont respectivement choisis égaux à $2\mu\text{m}/1\mu\text{m}$ et $100\mu\text{A}$. Les variations du courant différentiel Δi_{in} ont été tracées en fonction de la tension d'entrée v_{in} . La zone de linéarité associée est de l'ordre de $1.4\text{V}_{\text{càc}}$ pour une précision de $5\mu\text{A}$ (soit une précision relative de 10%).

En tenant compte des équations (IV-31) et (IV-33), en régime petit signal, le courant différentiel de sortie Δi_{Dmult} est proportionnel au produit de la tension d'entrée v_{in} et du poids w :

$$\Delta i_{\text{Dmult}} = K_{\text{mult}} G_M w v_{\text{in}} \quad (\text{IV-34})$$

Les courbes en pointillés sont relatives à l'équation linéaire (IV-31). Elles résultent de la multiplication du courant différentiel d'entrée ΔI_{in} par le poids w , au facteur K_{mult} près. Pour un poids w égal $0.1V$, les caractéristiques de transfert relatives aux courants différentiels ΔI_{in} et $\Delta I_{Dmult}(w = 0.1V)$ sont relativement proches l'une de l'autre. Le petit écart observé entre les deux courbes est dû à la modification de la taille des transistors des paires différentielles de façon à optimiser la zone de linéarité. Cette modification entraîne la choix d'un facteur de proportionnalité $K_{mult} = 8V^{-1}$ en simulation, au lieu de $10V^{-1}$.

Pour les autres courbes représentées, relatives à d'autres valeurs de poids w choisis entre $-0.5V$ et $0.5V$, les courants différentiels de sortie simulés sont proches de ceux attendus pour une tension d'entrée variant de $-0.5V$ à $0.5V$ pour une précision d'environ $15\mu A$ (soit une précision relative de 10%).

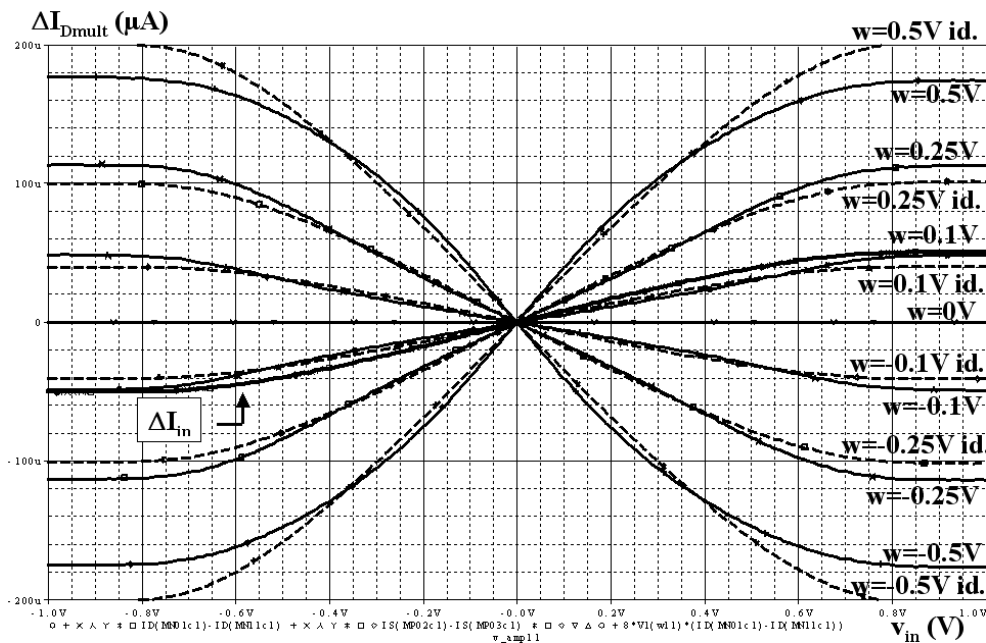


Figure IV-o : Représentation (en trait plein) de la variation du courant différentiel de sortie ΔI_{Dmult} en fonction de la tension v_{in} pour différentes valeurs de poids w . Les courbes en pointillés sont relatives à la multiplication du courant différentiel ΔI_{in} issu de la cellule précédente avec les différents poids, selon l'équation (IV-31).

IV.2.6. Génération d'un courant de décalage additionnel contrôlé par la tension de biais

Chaque additionneur de la couche cachée fait suite à deux multiplieurs (figure III-m). Comme l'ensemble des transferts entre cellules est effectué à partir de courants différentiels, l'objectif de l'additionneur de biais est d'ajouter un courant différentiel ΔI_{Dadd} à la somme des courants différentiels de sortie des deux multiplieurs. Pour réaliser cette opération, le circuit

différentiel représenté figure IV-1 est utilisé, avec $v_{in} = b$, $(\frac{W}{L})_{diff} = \frac{2\mu m}{1\mu m}$ et $I_0 = 100\mu A$. Ainsi,

l'expression du courant différentiel de sortie ΔI_{Dadd} de l'additionneur se déduit des équations (IV-19) et (IV-27) :

$$\Delta I_{Dadd} = I_{Dadd1} - I_{Dadd2} = b \sqrt{2\beta_{diff} I_0} \sqrt{1 - \frac{\beta_{diff}}{2I_0} b^2} \quad (IV-35)$$

Au premier ordre, on obtient une relation de proportionnalité entre le courant différentiel de sortie "petit signal" Δi_{Dadd} et la tension de biais b (IV-20):

$$\Delta i_{Dadd} = G_{Madd} b \quad (IV-36)$$

avec G_{Madd} la transconductance associée à la paire différentielle. La valeur de la transconductance a été choisie de manière à ce que la plage de variation du biais b entraîne une forte variation du courant différentiel de sortie de l'additionneur. Ce critère permet de balayer l'intégralité de la caractéristique de transfert de la fonction tangente hyperbolique (paragraphe IV.2.5).

Le courant de décalage de sortie ΔI_{Dadd} est représenté en fonction du biais b en trait plein sur la figure IV-p. Deux zones de linéarité admissibles ont été tracées, chacune pour une précision relative différente. Pour une précision de $15\mu A$ (soit une précision relative de 15%), on obtient une zone de linéarité de l'ordre de $1.5V_{c\grave{a}c}$. Pour une précision de $5\mu A$ (soit une précision relative de 5%), on obtient une zone de linéarité de l'ordre de $1V_{c\grave{a}c}$.

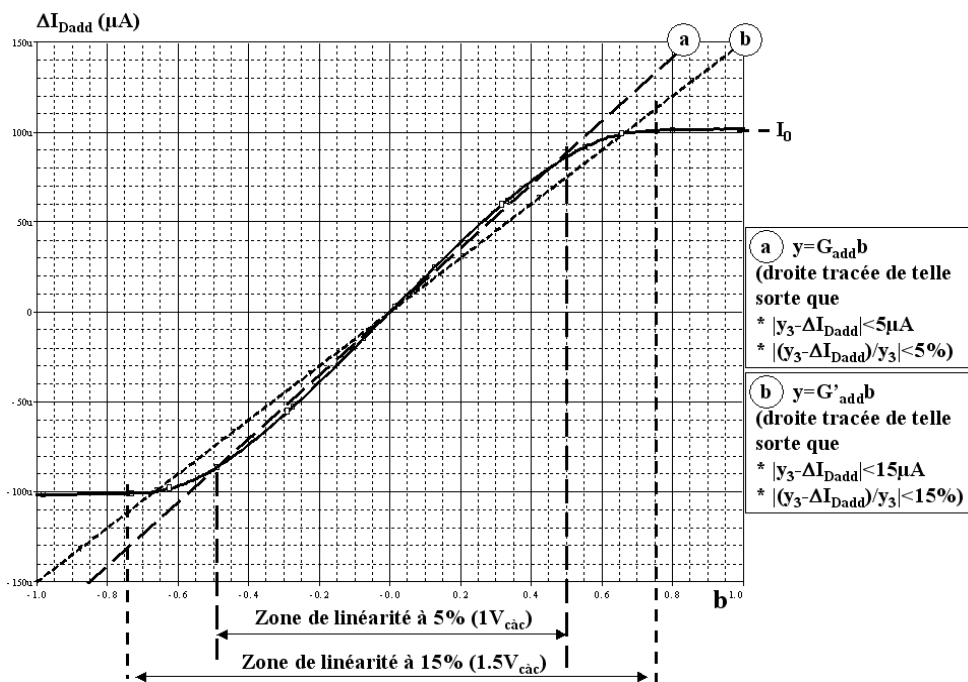


Figure IV-p : Observation du courant différentiel ΔI_{Dadd} en fonction du biais b .

IV.2.7. L'additionneur

Le courant différentiel de sortie de l'additionneur correspond au courant différentiel d'entrée de la fonction tangente hyperbolique. Il est égal à la somme des courants différentiels de sortie issus de deux multiplieurs et du courant de décalage additionnel contrôlé par la tension de biais. L'addition de ces courants est effectuée simplement à l'aide de miroirs de courants.

IV.2.8. Tangente hyperbolique

La tangente hyperbolique est la fonction non linéaire principale de ce circuit. C'est elle qui caractérise la fonction d'activation de ce circuit (en d'autres termes, la "compression" du transfert ou encore la non-linéarité). Elle intervient au niveau de chaque neurone de la couche cachée (figure IV-q).

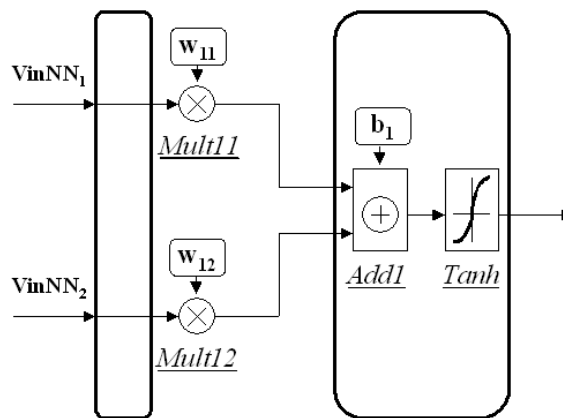


Figure IV-q : Schéma bloc de la fonction tangente hyperbolique faisant intervenir les cellules qui la précède. Pour alléger la figure, les transferts différentiels en courant entre cellules n'ont été représentés que par une flèche.

L'objectif premier de la fonction tangente hyperbolique est de limiter la valeur du courant différentiel de sortie à une valeur bien définie. Le principal inconvénient de la réalisation de cette fonction en circuit intégré analogique est liée au fait qu'on ne puisse plus modifier la non-linéarité appliquée une fois que le circuit a été conçu. Cela nous amène au deuxième objectif associé à cette cellule, celui de contrôler de la zone de non-linéarité par l'utilisateur, de manière à ce que le réseau de neurones puisse être paramétrable pour d'autres applications.

Pour réaliser la tangente hyperbolique, un circuit différentiel à sources couplées est à nouveau utilisé (figure IV-r).

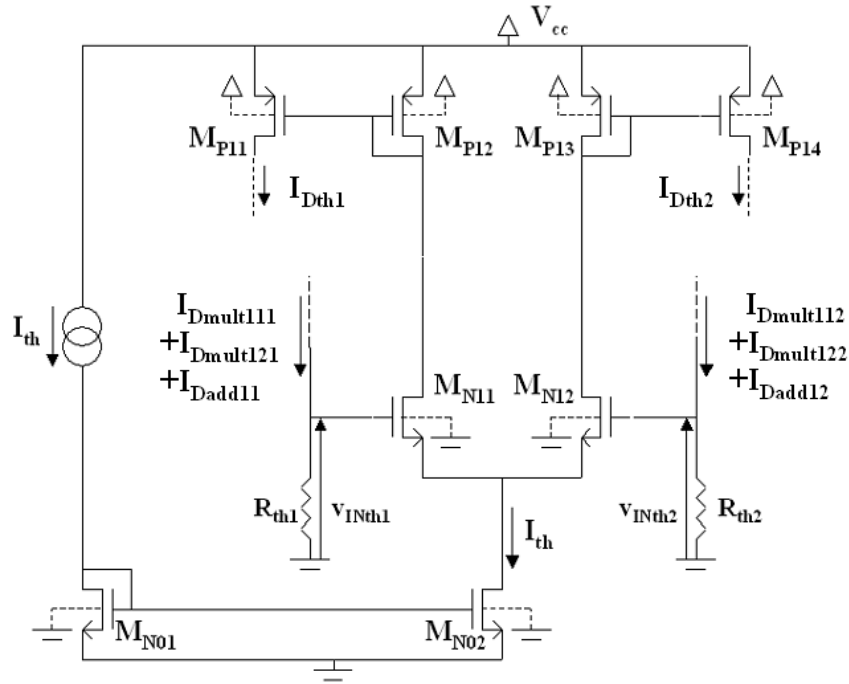


Figure IV-r : Schéma de la tangente hyperbolique.

A la différence des circuits multiplieurs et additionneurs, pour lesquels l'objectif était d'accroître au maximum la zone de linéarité, nous nous intéresserons ici à la caractéristique de transfert globale de la paire différentielle (IV-21) :

$$\Delta I_{Dth} = I_{th} \tanh\left(\frac{G_{Mth} \Delta v_{INth}}{I_{th}}\right) \quad (IV-37)$$

avec I_{th} le courant de polarisation de la cellule, $\Delta v_{INth} = v_{INth1} - v_{INth2}$ la tension différentielle d'entrée, proportionnelle au courant différentiel d'entrée ΔI_{INth} par l'intermédiaire des résistances R_{th1} et R_{th2} de valeur R_{th} , et G_{Mth} la transconductance associée au circuit différentiel.

$$G_{Mth} = \sqrt{4\beta_{diff} I_{th}} \quad (IV-38)$$

Ainsi, le transfert global en courant différentiel de la fonction tangente hyperbolique peut s'écrire sous la forme suivante :

$$\Delta I_{Dth} = I_{th} \tanh\left(G_{Mth} R_{th} \frac{\Delta I_{Dmult11} + \Delta I_{Dmult12} + \Delta I_{Dadd1}}{I_{th}}\right) \quad (IV-39)$$

Ainsi, à partir des équations (IV-34) et (IV-36), en régime petit signal :

$$\Delta I_{Dth} = I_{th} \tanh\left(G_{Mth} R_{th} \frac{K_{mult} G_M w_{11} v_{inNN1} + K_{mult} G_M w_{12} v_{inNN2} + G_{Madd1} b_1}{I_{th}}\right) \quad (IV-40)$$

La valeur de la résistance R_{th} a été choisie égale à $6k\Omega$ afin d'assurer une tension de mode commun suffisante. Les tailles des transistors de la paire différentielle ont été optimisés de

manière à ce que la transconductance G_{Mth} permette un transfert en courant différentiel unitaire dans la zone linéaire pour un courant I_{th} donné ([Gatet, 2007(b)]).

Le courant de polarisation I_{th} est paramétrable à l'extérieur de l'ASIC, de manière à pouvoir contrôler la zone de linéarité de la fonction par l'intermédiaire d'un courant de référence.

Remarque : les résistances R_{th1} et R_{th2} reliées à la masse permettent d'assurer une tension de mode commun suffisante. La solution proposée figure IV-s permet de décaler la tension commune aux deux résistances d'une tension V_{GS} .

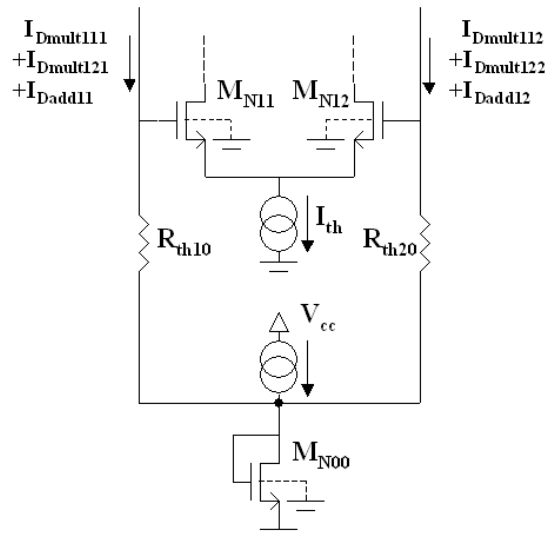


Figure IV-s : Décalage de la tension commune aux deux résistances d'un V_{GS} .

De ce fait, les valeurs des résistances R_{th10} et R_{th20} sont plus faibles, ce qui permet de réduire :

- la tension de décalage induite par les défauts d'appairage $\Delta R_{th}/R_{th}$ des deux résistances,
- la valeur de la constante de temps "vue" par l'entrée de la paire différentielle, et donc la bande passante associée,
- la surface de silicium occupée par ces deux résistances.

Les figures IV-t et IV-u montrent des comparaisons entre le courant différentiel d'entrée ΔI_{INth} (en pointillés) et le courant différentiel de sortie ΔI_{Dth} (en trait plein) en fonction de la valeur de la tension d'entrée v_{inNN1} pour différentes valeurs de poids w_{11} ou différentes valeurs de biais b_1 .

Le courant de polarisation I_{th} a été choisi égal à $60\mu A$. On note que le transfert en courant en mode "petit signal" est bien égal à 1, puisque pour un courant différentiel d'entrée ΔI_{INth} inférieur à I_{th} , le courant différentiel de sortie ΔI_{Dth} est superposé avec le courant différentiel d'entrée. Ainsi, pour $w_{11}=0.1V$ et $w_{12}=b_1=0V$, le courant différentiel de sortie de la tangente

hyperbolique est égal au courant différentiel ΔI_{in1} issu de la conversion de la tension d'entrée V_{inNN1} en courant différentiel.

Pour des fortes valeurs de courant différentiel d'entrée ΔI_{INth} (régime "forts signaux"), on note la compression de la caractéristique jusqu'à la valeur $I_{th}=60\mu A$ fixée par l'utilisateur.

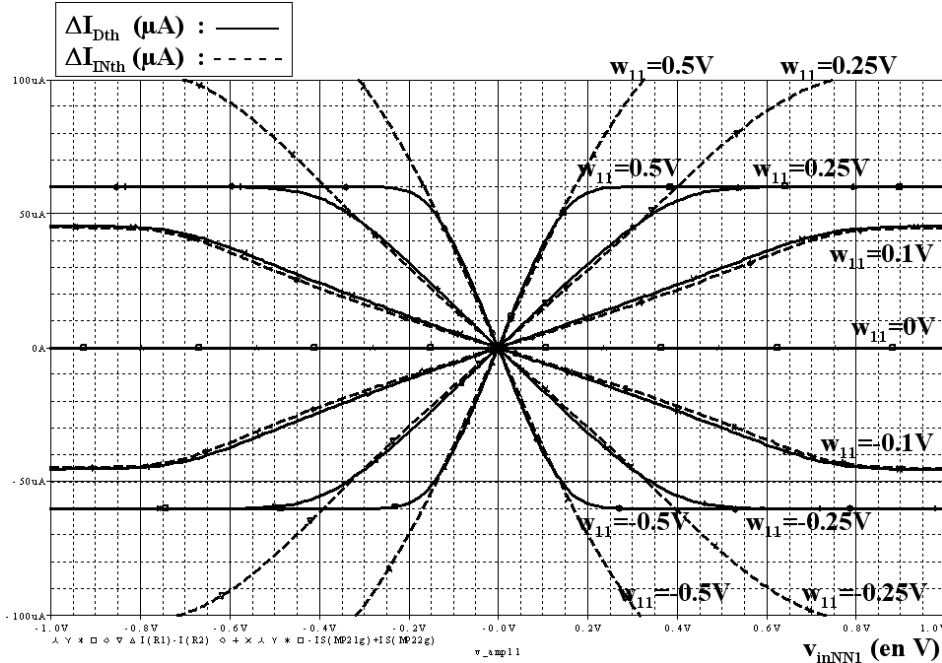


Figure IV-t : Variations des courants différentiels d'entrée et de sortie de la fonction tangente hyperbolique en fonction de la tension d'entrée v_{inNN1} pour différentes valeurs de poids w_{11} . La tension d'entrée v_{inNN2} et le biais b_1 ont été pris égaux à 0V.

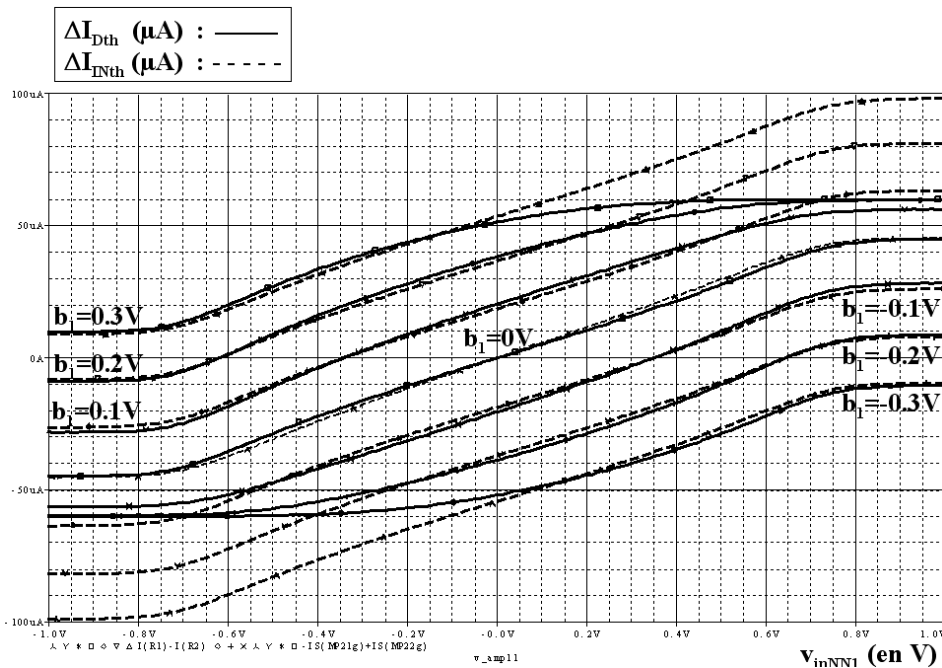


Figure IV-u : Variations des courants différentiels d'entrée et de sortie de la fonction tangente hyperbolique en fonction de la tension d'entrée v_{inNN1} pour différentes valeurs de biais b_1 . La tension d'entrée v_{inNN2} et le poids w_{11} ont été pris respectivement égaux à 0V et 0.1V.

IV.3. CONCEPTION DE L'ASIC

Les simulations effectuées sur la structure neuronale globale sont conformes aux spécifications fonctionnelles attendues. La consommation en courant du réseau est de 15mA (soit 50mW avec $V_{CC} = 3.3V$). Chaque neurone élémentaire (figure IV-v) consomme 4.35mA. Les performances du réseau en termes de dynamique, de précision et de consommation sont suffisantes pour les deux applications au sein desquelles celui-ci doit être implanté.

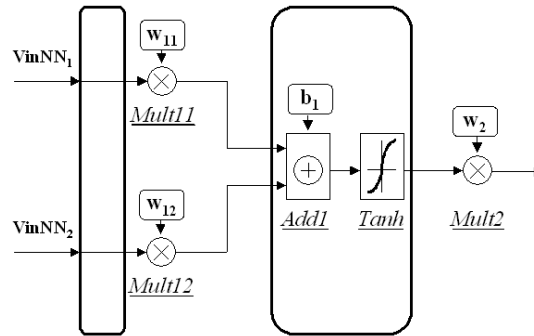


Figure IV-v : Structure d'un neurone élémentaire.

Pour valider expérimentalement le bon comportement du réseau observé en simulation, celui-ci a été intégré dans un ASIC. Comme le coût de fonderie dépend du nombre d'entrées/sorties du boîtier, celui-ci a été limité. Ne pouvant se priver des sorties intermédiaires pour valider chaque cellule, seul un neurone a été implanté. La figure IV-w représente le schéma du neurone intégré. Le multiplieur Mult2 a été conçu de manière à pouvoir être testé en tension et en courant. La figure IV-x présente le layout du neurone.

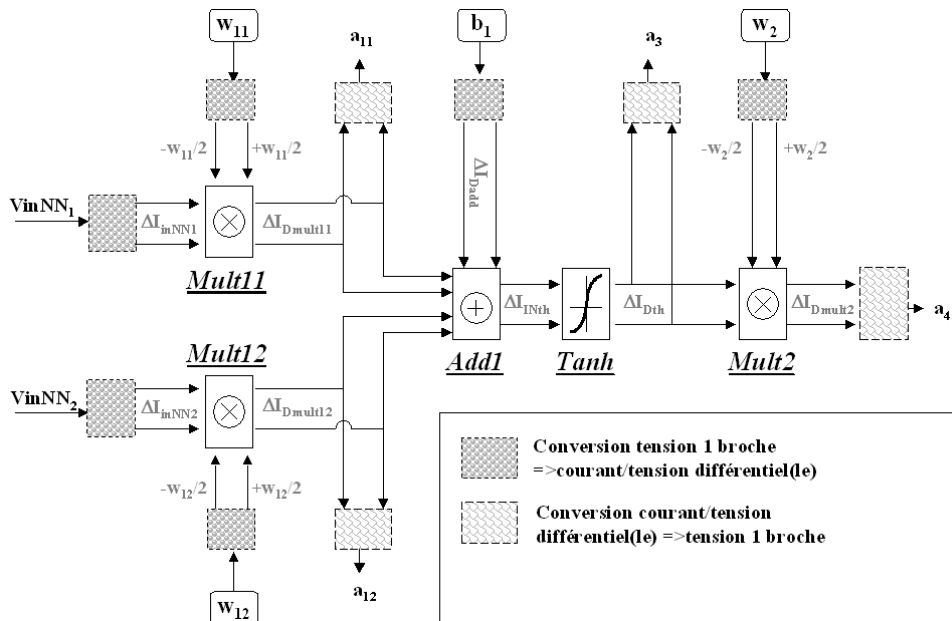


Figure IV-w : Schéma bloc du neurone élémentaire réalisé.

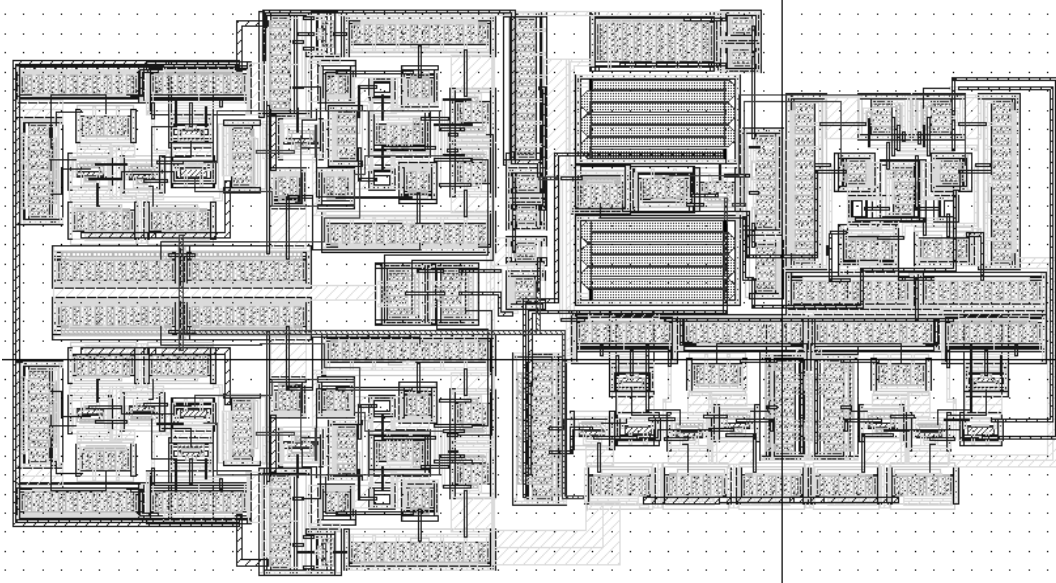


Figure IV-x : Layout du neurone élémentaire.
Taille du neurone hors plots d'entrée/sortie : $276\mu\text{m} \times 148\mu\text{m} = 0.04\text{mm}^2$.

Un soin tout particulier a été apporté au dessin des miroirs de courants et des paires différentielles. Les tensions de grille ont été réparties de manière homogène sur toute la largeur des transistors, en les divisant en plusieurs transistors mis en parallèle.

Pour les paires différentielles, l'objectif est de minimiser les erreurs de défauts d'appariage (*mismatches*) entre les deux transistors impliqués. Si deux transistors ne sont pas parfaitement appariés, une paire différentielle sera déséquilibrée au repos. Une conception systématique centroïde a été réalisée ([Hastings, 2005]).

De même, pour les miroir de courant, un défaut d'appariage entraînerait une erreur de copie. Pour éviter tout écart en relatif, les deux transistors impliqués sont imbriqués l'un dans l'autre.

L'annexe D décrit l'attribution des broches allouées au neurone, ainsi que le layout de l'ASIC complet.

La partie suivante présente les différents tests effectués à partir du neurone élémentaire intégré.

IV.4. VALIDATION EXPERIMENTALE

Les caractéristiques de transfert de chaque cellule appartenant au neurone élémentaire ont été tracées afin de les comparer avec celles obtenues en simulation. Chaque caractéristique a été tracée plusieurs fois afin de vérifier la reproductibilité du comportement dans des conditions variables. La principale fonction de ces courbes est d'en extraire l'équation fonctionnelle en fonction de chacun des paramètres ([Gatet, 2007(c)]). Pour la présentation des résultats, les courants de polarisation I_0 et I_{th} ont été fixés respectivement à $100\mu\text{A}$ et $60\mu\text{A}$. Le PCB réalisé pour tester le neurone est représenté sur la figure IV-y.

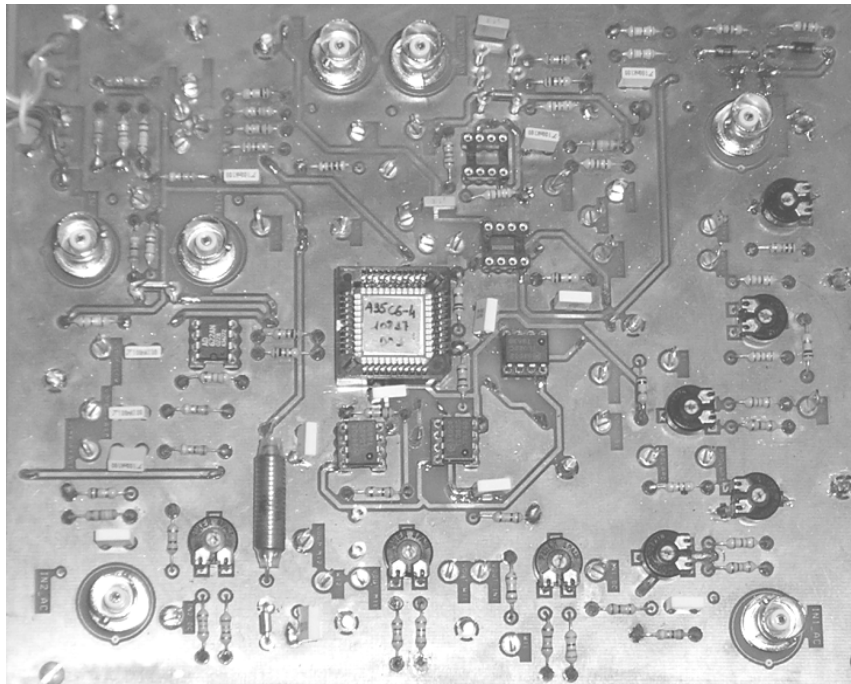


Figure IV-y : PCB réalisé pour tester le neurone élémentaire et les autres circuits implantés dans l'ASIC.

IV.4.1. Premier multiplieur

La figure IV-z représente l'ensemble des tensions de sortie continues mesurées à la sortie a_{11} du multiplieur Mult11 pour 11 valeurs de poids w_{11} uniformément réparties entre -0.5V et $+0.5\text{V}$. Pour valider localement le bon fonctionnement de chaque cellule élémentaire, des buffers de sortie intermédiaires ont été réalisés de manière très simple (à l'aide de miroirs de courant). Ils n'ont pas été optimisés en terme de dynamique afin de limiter la consommation globale du neurone.

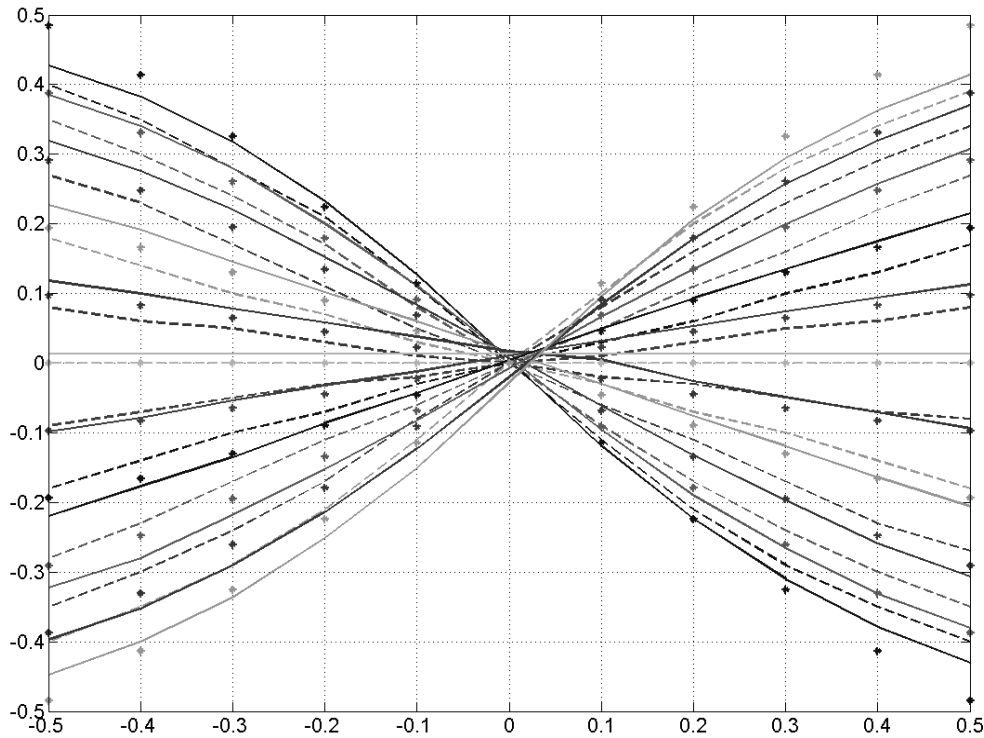


Figure IV-2 : Comparaison entre la simulation (pointillés), le test (trait plein) et l'équation de la caractéristique de transfert la plus proche ((IV-41) - points) du premier multiplieur en fonction de V_{inNN_1} pour différentes valeurs de poids $w_{11} \in [-0.5V; 0.5V]$ par pas de 0.1V.

Le buffer de sortie du multiplieur Mult11 décale la caractéristique pour des valeurs de tension d'entrée et de poids proches de 0.5V en valeur absolue.

La tension de décalage de 10mV observée au voisinage de 0V peut être attribuée au buffer de sortie, mais également aux inévitables défauts d'appairage des transistors des paires différentielles et des miroirs de courants, ainsi qu'aux limites de linéarité atteintes pour des poids proches de 0.5V en valeur absolue. Cette dernière hypothèse est en accord avec l'évolution de cette tension de décalage avec l'amplitude du poids. Certes, cette erreur se répercute sur les cellules suivantes, mais dans de faibles proportions en terme de précision et d'erreur relative.

L'équation du transfert en tension a_{11} du multiplieur se rapprochant le plus des caractéristiques de test de la cellule est de la forme :

$$a_{11} = 2,3w_{11}V_{inNN_1}(1-0,63V_{inNN_1}^2) \quad (IV-41)$$

Le critère utilisé pour déterminer les coefficients associés à cette équation de transfert est la minimisation de la somme des valeurs absolues des erreurs constatées entre les différents points de tests et cette équation. L'équation de transfert associée au test du multiplieur Mult12, mettant en relation la tension de sortie a_{12} en fonction de la tension d'entrée V_{inNN_2} et du poids w_{12} , est identique.

Le terme du troisième ordre est attribué à la distorsion due au buffer. On remarque que le coefficient de proportionnalité au premier ordre K_{a11} entre la tension de sortie a_{11} observée et le produit " $w_{11}V_{inNN_1}$ " est plus élevé en test qu'en simulation ($2.3V^{-1}$ contre $2V^{-1}$). Cette différence est attribué à des erreurs de recopie du courant de polarisation I_0 dans l'ASIC. Cette différence entre simulation et test peut être facilement compensée en ré-ajustant la valeur du poids w_{11} .

La somme des valeurs absolues des écarts constatés entre chaque point de test (courbe tracée en trait plein figure IV-z) et la valeur correspondante déduite de l'équation (IV-41) (points) a été moyennée de manière à déterminer l'écart moyen $e_{MEAN_{multi_i}}$ associée au multiplieur. Cet écart permet de quantifier la proximité de l'équation de transfert du multiplieur par rapport à la caractéristique de transfert "réelle". Il a été évalué à 20mV pour chacun des deux multiplieurs Mult11 et Mult12, et ce pour toute valeur de tension d'entrée V_{inNN_i} ou de poids w_{1i} appartenant à l'intervalle $[-0.5V;0.5V]$. Néanmoins, il faut rappeler que cette approximation tient compte des défauts du buffer utilisé.

IV.4.2. Fonction tangente hyperbolique

La figure IV-aa représente l'ensemble des tensions de sortie continues mesurées à la sortie a_3 de la fonction tangente hyperbolique en fonction de la tension d'entrée V_{inNN_1} pour 11 valeurs de poids w_{11} uniformément réparties entre $-0.5V$ et $+0.5V$. La tension d'entrée V_{inNN_2} , le poids w_{12} et le biais b_1 ont été fixés à 0V de façon à n'observer l'influence que d'une seule entrée.

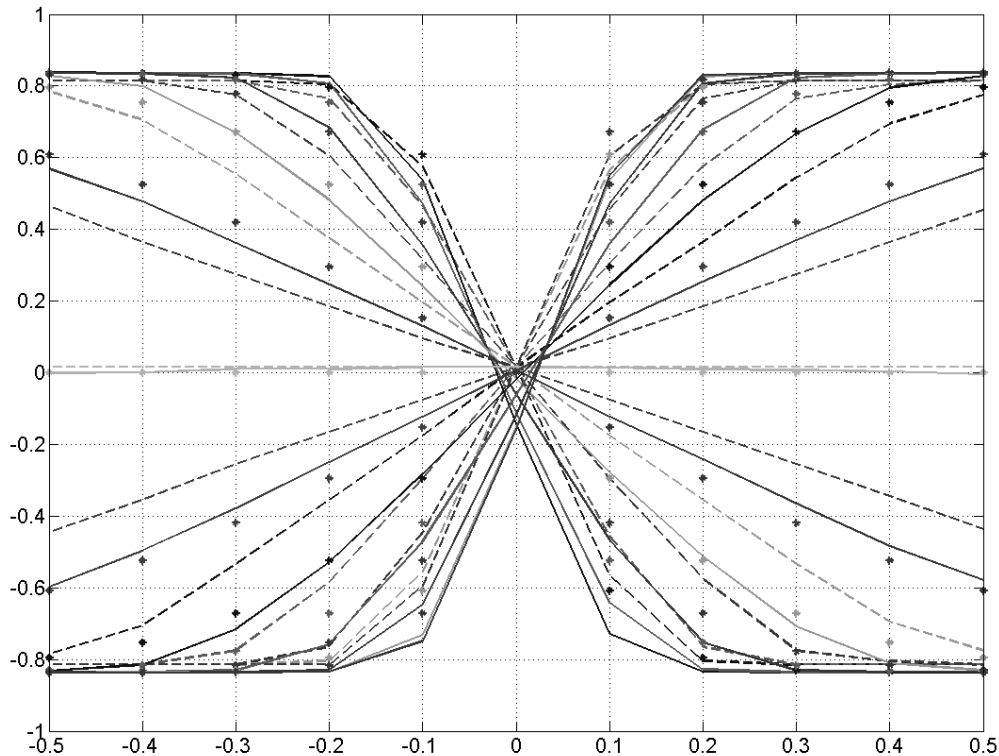


Figure IV-aa : Comparaison entre la simulation (pointillés), le test (trait plein) et l'équation de la caractéristique de transfert la plus proche ((IV-42) - points) de la fonction tangente hyperbolique en fonction de V_{inNN_1} pour différentes valeurs de poids $w_{11} \in [-0.5V; 0.5V]$ par pas de $0.1V$. La tension d'entrée V_{inNN_2} , le poids w_{12} et le biais b_1 sont fixés à $0V$.

On note que pour des valeurs de poids w_{11} proches de $0.5V$ en valeur absolue, la tension d'erreur observée au voisinage de $0V$ est importante au premier abord. Cependant, elle doit être attribuée à la difficulté de la prise de mesure pour des valeurs de tension d'entrée V_{inNN_1} proches de 0 à cause des fortes variations de la tension de sortie. Elle n'apparaît quasiment plus à la sortie de la cellule multiplieur suivant Mult2 (paragraphe IV.4.3).

Les caractéristiques de sortie de la tangente hyperbolique sont identiques lorsque l'on fait varier la tension d'entrée V_{inNN_1} ou V_{inNN_2} . Ainsi, la "superposition non linéaire" est admise. Pour un courant de polarisation I_{th} de $60\mu A$, l'équation de transfert en tension a_3 de la fonction tangente hyperbolique est donné par la relation :

$$a_3 = 0,84 \tanh(8(2,3w_{11}V_{inNN_1} + 2,3w_{12}V_{inNN_2} + 0,6b_1)) \quad (IV-42)$$

Le choix d'un courant de polarisation I_{th} de $60\mu A$ impose une limitation de la tension de sortie a_3 à $0.84V$ en simulation et en test. On vérifie que pour un courant de polarisation différent, la limite de la zone de linéarité diffère aussi.

Comme le coefficient de proportionnalité K_{a11} du multiplieur Mult11 au premier ordre est plus élevé en test qu'en simulation, il est normal d'observer un coefficient de proportionnalité G_{\tanh} de la cellule tangente hyperbolique en zone linéaire plus important en test qu'en

simulation ($0.84 \times 8 \times 2.3 = 15.4V^{-1}$ contre $10V^{-1}$). Là encore, cet écart au niveau du gain peut être facilement compensé en ré-ajustant la valeur des poids w_{11} et w_{12} .

Le coefficient de proportionnalité K_{a2} relatif au biais b_1 a été choisi relativement élevé pour balayer l'intégralité de la caractéristique de la tangente hyperbolique quelle que soit la valeur du courant de polarisation I_{th} dans l'intervalle $[I_0/5; 5I_0]$. Pour la présentation des résultats, le courant de saturation I_{th} a été fixé à $60\mu A$, soit $3/5$ du courant de polarisation I_0 . En conséquence, la plage de linéarité associée au biais b_1 est relativement faible ($[-0.15V; 0.15V]$).

L'écart moyen $e_{MEAN_{\tanh}}$ constaté entre chaque point de test (courbe tracée en trait plein figure IV-aa) et la valeur correspondante déduite de l'équation (IV-42) (points) est évalué à $30mV$ pour toute valeur de tension d'entrée V_{inNN_1} ou V_{inNN_2} ou de poids w_{11} ou w_{12} appartenant à l'intervalle $[-0.5V; 0.5V]$ ou de biais b_1 appartenant à l'intervalle $[-0.15V; 0.15V]$. Cette erreur est essentiellement due aux difficultés de mesure observées pour des valeurs de poids proches de $0.5V$ en valeur absolue et pour des valeurs de tension d'entrée proches de $0V$.

IV.4.3. Multiplieur de sortie – Neurone élémentaire

La figure IV-ab représente l'ensemble des tensions de sortie continues mesurées à la sortie a_4 du neurone, pour 11 valeurs de poids w_2 uniformément réparties entre $-0.5V$ et $+0.5V$. Le poids w_{11} a été fixé à $0.1V$ de manière à se trouver dans la zone de linéarité de la fonction tangente hyperbolique. La tension d'entrée V_{inNN_2} , le poids w_{12} et le biais b_1 ont été fixés à $0V$ de façon à n'observer l'influence que d'une seule entrée.

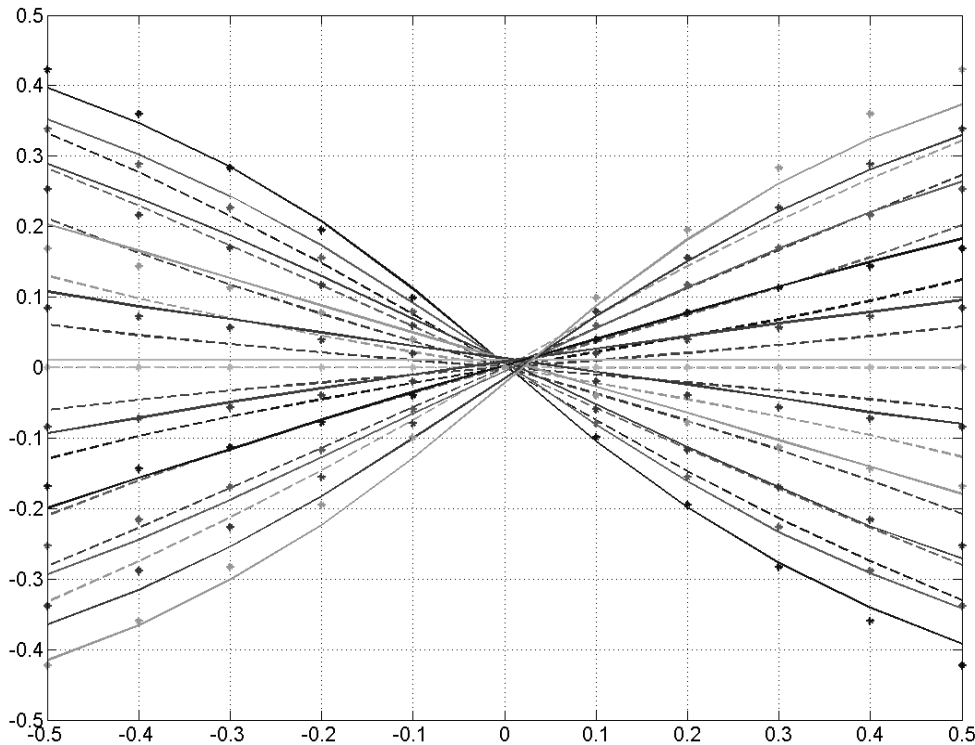


Figure IV-ab : Comparaison entre la simulation (pointillés), le test (trait plein) et l'équation de la caractéristique de transfert la plus proche ((IV-44) - points) du deuxième multiplieur en fonction de V_{inNN_1} pour différentes valeurs de poids $w_2 \in [-0.5V; 0.5V]$ par pas de $0.1V$. Le poids w_{11} est fixé à $0.1V$. La tension d'entrée V_{inNN_2} , le poids w_{12} et le biais b_1 sont fixés à $0V$.

On note que pour des valeurs de poids w_2 proches de $0.5V$ en valeur absolue, l'écart observé au voisinage de $0V$ est nettement moins important que celui observé en sortie de la tangente hyperbolique. Les deux principales raisons sont le soin apporté à la réalisation du buffer de sortie du multiplieur Mult2 et la variation moins importante de l'amplitude des signaux par rapport à la sortie de la tangente hyperbolique.

L'équation de transfert en tension a_4 du multiplieur Mult2 peut se limiter au premier ordre :

$$a_4 = 1,4w_2a_3 \quad (IV-43)$$

Ainsi, on peut en déduire l'équation de transfert en tension du neurone élémentaire de test :

$$a_4 = 1,17w_2 \tanh[8(2,3w_{11}V_{inNN_1} + 2,3w_{12}V_{inNN_2} + 0,6b_1)] \quad (IV-44)$$

Comme pour le premier multiplieur, le multiplieur Mult2 présente au premier ordre un coefficient de proportionnalité K_{a4} plus élevé en test qu'en simulation, ce qui explique les différences de pente non négligeables pour des mêmes conditions de simulation et de test. Cet écart au niveau du gain peut être facilement compensée en ré-ajustant la valeur des poids w_{11} , w_{12} et w_2 .

L'écart moyen $e_{MEAN_{mult2}}$ constaté entre chaque point de test (courbe tracée en trait plein figure IV-ab) et la valeur correspondante déduite de l'équation (IV-44) (points) est évalué à

10mV pour toute valeur de tension d'entrée V_{inNN_1} ou V_{inNN_2} ou de poids w_{11} ou w_{12} ou w_2 appartenant à l'intervalle $[-0.5V;0.5V]$ ou de biais b_1 appartenant à l'intervalle $[-0.15V;0.15V]$.

Il est à noter que cet écart moyen $e_{MEAN_{mult2}}$ est déterminé à partir de l'équation (IV-44), équation au premier ordre par rapport à chacune de ses variables. On peut donc facilement rapprocher cette équation des équations (III-7) à (III-11) et (III-15) à (III-19) utilisées pour valider comportementalement la mise en oeuvre du réseau de neurones au sein des deux applications télémétriques. La même procédure peut donc être ré-utilisée, avec un calcul de mise à jour des paramètres similaire lors de la phase d'entraînement. Il suffit juste de veiller à modifier les coefficients K_{a11} , K_{a12} , K_{a2} , K_{a3} , K_{th} et K_{a4} associés à chaque opération. Il faut néanmoins tenir compte de l'erreur moyenne $e_{MEAN_{mult2}}$ au moment du calibrage du réseau de neurones à l'aide du logiciel Matlab de façon à ce que les résultats obtenus à l'issue des phases d'entraînement et de validation soient en adéquation avec ceux observés en sortie du réseau implanté.

Si les résultats issus de la phase d'entraînement du réseau diffèrent trop par rapport à ceux issus des tests expérimentaux, il sera nécessaire de tenir compte des termes d'ordre supérieurs associés à chaque cellule, de façon à approcher avec plus de précision le comportement du neurone en test. La contre-partie sera la difficulté de programmation de la mise à jour des poids et biais à chaque itération lors de la phase d'entraînement qui tient compte des dérivées première et seconde de la tension de sortie par rapport à chacune de ses variables.

La consommation du neurone intégré est égale à 4.25mA, valeur en adéquation avec celle attendue en simulation (4.35mA, soit une erreur de 2%). Ainsi, le comportement du neurone élémentaire intégré a été validé par rapport aux simulations. Les plages de variation associées à chacun des paramètres sont $[-0.5V;0.5V]$ pour les entrées et les poids et $[-0.15V;0.15V]$ pour le biais. Sur ces plages de variation, la tension a_4 observée en sortie du neurone est quasiment linéaire par rapport à chacun de ces paramètres (à 10mV près en valeur moyenne).

IV.4.4. Réalisation et test du réseau de neurones complet

Comme le neurone élémentaire intégré a été validé fonctionnellement et qualifié, l'étape suivante consiste à réaliser le réseau de neurones complet ([Gatet, 2007(c)]). Plutôt que de réaliser un nouvel ASIC contenant la structure globale du réseau (figure IV-ac), il a été décidé, pour la validation du principe, d'associer trois ASIC contenant chacun un neurone élémentaire (figure IV-ad). La tension d'alimentation des trois ASIC est la même

($V_{CC} = 3.3V$). Les entrées V_{inNN1} et V_{inNN2} communes aux trois ASIC ont été reliées entre elles (figure IV-ac et IV-ae), ainsi que les masses.

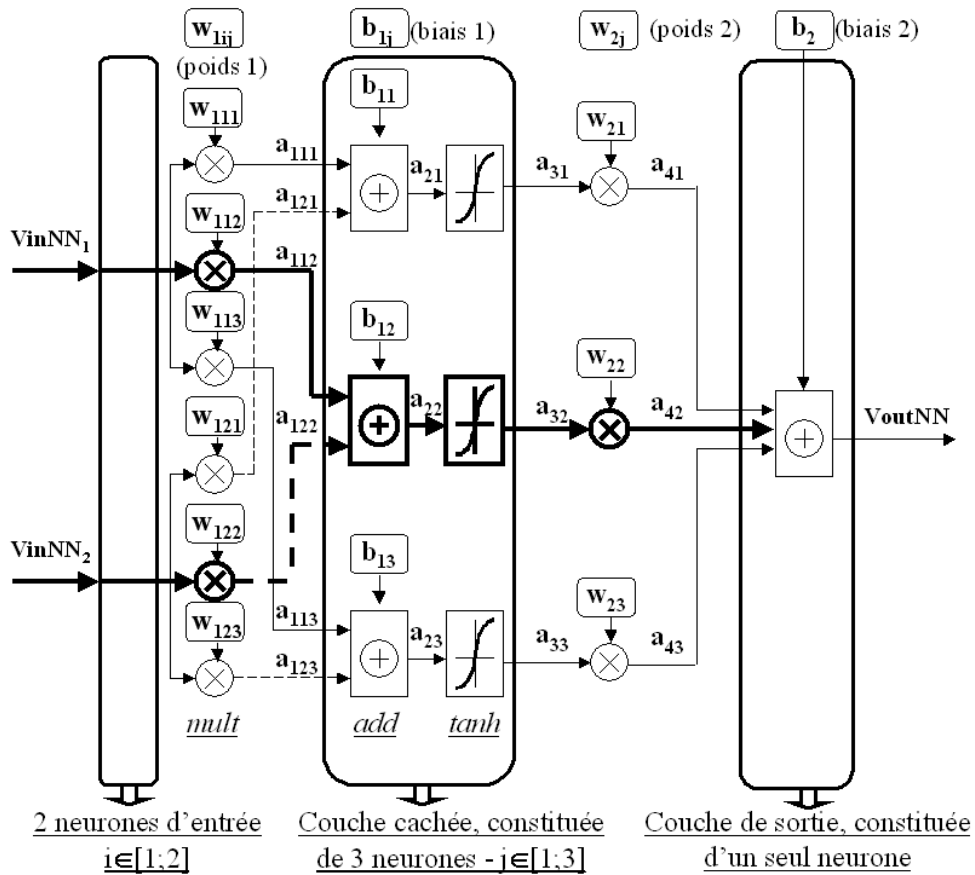


Figure IV-ac: Structure du réseau de neurones comprenant trois neurones élémentaires, chacun intégré dans un ASIC. En gras, les cellules neuronales contenues dans un ASIC.

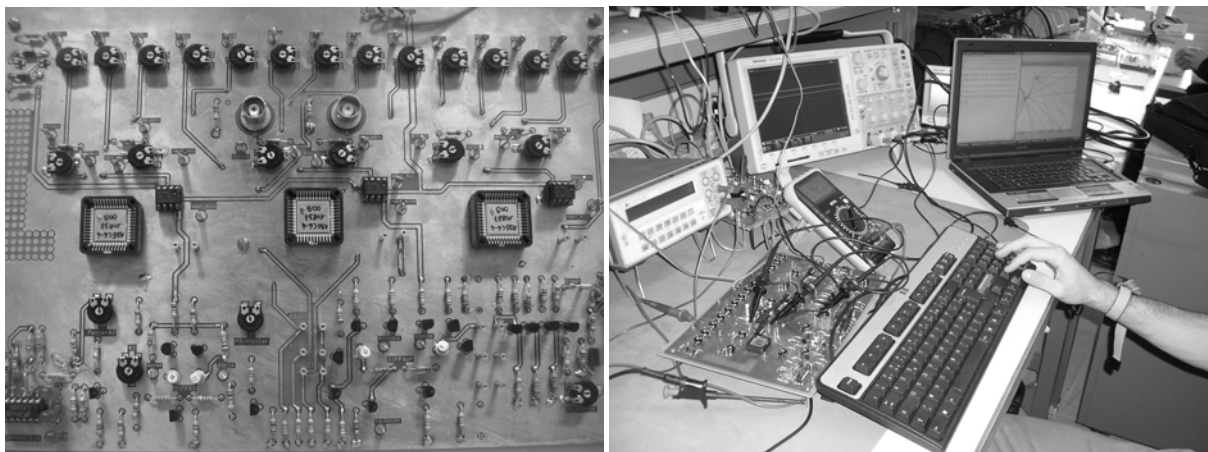


Figure IV-ad: Représentation du réseau de neurones implanté sur PCB et environnement expérimental.

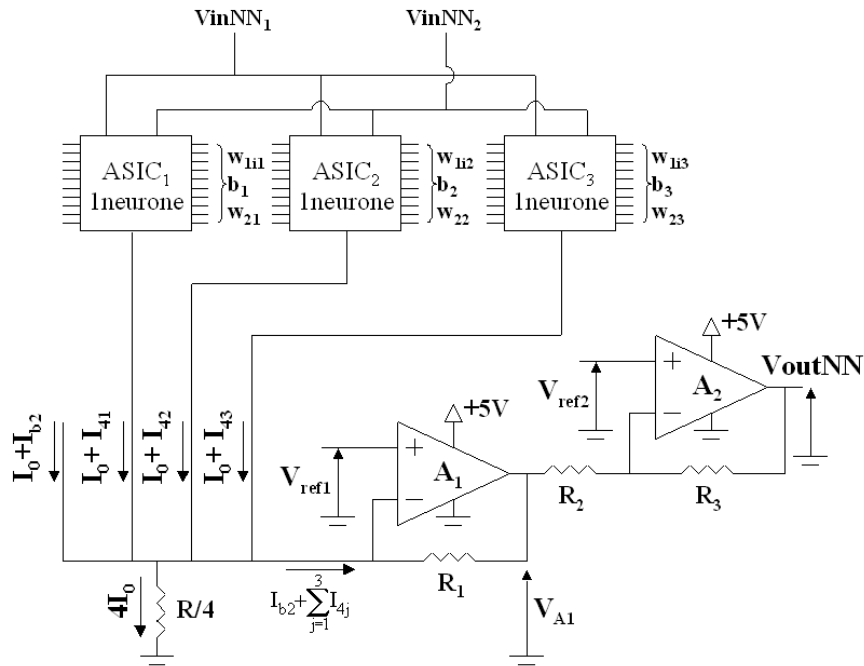


Figure IV-ae : Schéma électrique du réseau constitué de 3 neurones intégrés sur la couche cachée. Le neurone de sortie, réalisant la somme des courants issus des 3 neurones avec un courant proportionnel au biais b_2 , est directement implanté sur le PCB.

L'indice j est affecté aux grandeurs relatives à chaque neurone ($j \in \{1;2;3\}$). Ainsi, la sortie de chaque neurone élémentaire dont on a étudié l'équation de transfert en tension dans le paragraphe IV.4.3, est maintenant notée a_{4j} au lieu de a_4 .

Les courants de polarisation I_0 et I_{th} sont respectivement égaux à $100\mu A$ et $60\mu A$. La couche de sortie est réalisée directement sur PCB à l'aide de composants discrets. La tension d'alimentation associée aux différents circuits du neurone de sortie est prise égale à 5V de manière à accroître la dynamique des signaux de sortie.

Comme le buffer de sortie de chaque neurone a été réalisé de manière à utiliser la sortie "en tension" (a_{4j}) ou "en courant" (I_{4j}), les courants de sortie I_{4j} issus de chaque neurone sont sommés avec un quatrième courant I_{b2} proportionnel à une tension de biais b_2 référencée à la masse (figure IV-ac et IV-ae). Le courant total est converti en tension et bufferisé de manière à observer la tension de sortie du réseau V_{outNN} .

Enfin, un comparateur à trois seuils a été implanté sur le PCB à la sortie du réseau de neurones pour l'application relative à la détection de surfaces. Une DEL a été placée à chaque sortie des comparateurs, chacune s'allumant dès que la tension de sortie du réseau de neurones franchit un seuil. Le nombre de diodes allumées permet de corréliser plus facilement l'intervalle de tension auquel appartient la tension de sortie du réseau au type de surface détectée.

La tension de sortie a_{4j} du $j^{\text{ème}}$ neurone issu de l'ASIC $_j$ a été déterminée par l'intermédiaire du courant I_{4j} traversant une résistance de valeur R :

$$a_{4j} = RI_{4j} = 1,17w_{2j}\tanh(8(2,3w_{11j}V_{inNN_1}+2,3w_{12j}V_{inNN_2}+0,6b_{1j})) \quad (IV-45)$$

L'expression du courant de décalage additionnel I_{b2} contrôlé par la tension de biais b_2 est :

$$I_{b2} = G_{add2}b_2 \quad (IV-46)$$

L'amplificateur A_1 fonctionne en convertisseur courant-tension. Pour prendre en compte les différents courants de mode commun I_0 générés, on introduit une résistance supplémentaire $R/4$ telle que $V_{ref1} = \frac{R}{4} 4I_0$. Ainsi, la tension de l'entrée inverseuse reste constante. La tension de sortie V_{A1} de l'amplificateur opérationnel A_1 s'écrit :

$$V_{A1} = V_{ref1} - R_1(I_{b2} + \sum_{j=1}^3 I_{4j}) \quad (IV-47)$$

La tension de sortie V_{outNN} de l'amplificateur opérationnel A_2 s'écrit :

$$V_{outNN} = (1 + \frac{R_3}{R_2})V_{ref2} - \frac{R_3}{R_2}V_{A1} \quad (IV-48)$$

Ainsi, d'après les équations (IV-47) et (IV-48),

$$V_{outNN} = (1 + \frac{R_3}{R_2})V_{ref2} - \frac{R_3}{R_2}V_{ref1} + \frac{R_3}{R_2}R_1(I_{b2} + \sum_{j=1}^3 I_{4j}) \quad (IV-49)$$

Les tensions de référence V_{ref1} et V_{ref2} qui polarisent les bornes "+" des deux amplificateurs opérationnels sont fixées de telle sorte que la tension de sortie V_{outNN} du réseau de neurones soit égale à $V_{outNN_0} = 2.5V$ lorsque l'ensemble des paramètres du réseau est pris égal à $0V$. Les résistances R_1 , R_2 et R_3 sont choisies de manière à accroître d'un facteur 3 la plage de variation de la tension de sortie du réseau. Ainsi, d'après les équations (IV-45), (IV-46) et (IV-49), l'équation de transfert expérimentale du réseau de neurones complet est :

$$V_{outNN} = V_{outNN_0} + \frac{R_3}{R_2}R_1G_{add2}b_2 + \frac{R_3}{R_2}\frac{R_1}{R}\sum_{j=1}^3 a_{4j} = 2,5 + 6,75b_2 + 2,9\sum_{j=1}^3 a_{4j} \quad (IV-50)$$

Le coefficient de transfert $K_{b2} = \frac{R_3}{R_2}R_1G_{add2}$ associé au deuxième additionneur a été choisi relativement élevé de manière à pouvoir balayer l'intégralité de la plage des tensions admissibles en sortie du réseau. L'équation de transfert expérimentale du réseau de neurones complet en fonction de l'ensemble des paramètres s'écrit donc :

$$V_{outNN} = 2,5 + 6,75b_2 + 2,9\sum_{j=1}^3 [1,17w_{2j}\tanh(8(2,3w_{11j}V_{inNN_1} + 2,3w_{12j}V_{inNN_2} + 0,6b_{1j}))] \quad (IV-51)$$

L'écart moyen $e_{MEAN_{nn}}$ constaté est environ égal à $30mV$ pour toute valeur de tension d'entrée V_{inNN_1} ou V_{inNN_2} ou de poids w_{1ij} ou w_{2j} appartenant à l'intervalle $[-0.5V; 0.5V]$ ou

de biais b_{1j} appartenant à l'intervalle $[-0.15V;0.15V]$ ou de biais b_2 appartenant à l'intervalle $[-0.4V;0.4V]$.

Les consommations de chaque neurone intégré, de la couche de sortie réalisée sur PCB, et du réseau de neurones global ont été regroupées dans le tableau IV- β .

Tableau IV- β : Consommation en courant des cellules et puissance dissipée.

	Consommation en courant simulée	Consommation en courant mesurée	Puissance dissipée mesurée
Neurone j (ASICj, 3.3V)	4.35mA	4.25mA	14mW
Neurone de sortie (PCB, 5V)	2.15mA	2.25mA	11.3mW
Réseau de neurones global	15.2mA	15mA	53.3mW

IV.5. CONCLUSION

Le test fonctionnel du neurone élémentaire a permis la qualification du réseau de neurones complet pour le type d'application dans lequel il doit être intégré. Les plages admissibles de variation des entrées et des paramètres sont suffisantes pour calibrer le réseau pour les deux applications télémétriques. A partir de l'équation analytique extraite des caractéristiques continues du réseau de neurones, il est possible de l'entraîner à l'extérieur de l'ASIC (à l'aide du logiciel Matlab) en suivant les mêmes procédures que celles décrites dans le chapitre III. Néanmoins, à la vue des valeurs finales des paramètres observés à la fin des phases de validation comportementale (figure III-h et III-q), une restriction des intervalles de mesure associés s'impose. La phase de calibrage du réseau intégré relative à chaque application est détaillée dans le chapitre V.

Le réseau de neurones ainsi conçu présente un aspect générique. A priori, il peut être utilisé tel quel pour d'autres types d'applications nécessitant une basse consommation et une plage de variation des entrées et des poids de l'ordre d' $1V_{c\grave{a}c}$. De plus, il peut aussi être adapté en modifiant ses caractéristiques par l'intermédiaire des courants de polarisation I_0 et I_{th} fixés à l'extérieur de l'ASIC. On peut en particulier jouer sur la fonction de transfert associée à la tangente hyperbolique, en augmentant ou en diminuant la plage de linéarité.

Ce réseau de neurones peut donc être utilisé au sein d'une application dont on peut limiter le nombre de données d'entrées à deux (voire une) et ne nécessitant qu'une sortie, dont la résolution moyenne en tension requise est de 30mV sur une plage de $5V_{c\grave{a}c}$. Il peut servir de

démonstrateur pour tout type d'application où un réseau de neurones répondant à ces caractéristiques est potentiellement utilisable.

Un réseau de neurones intégré utilisant l'ensemble des cellules développées dans ce chapitre ne pose pas de difficultés en terme de conception. En revanche, le nombre de broches nécessaires au niveau du boîtier sera d'autant plus important qu'il y aura d'entrées et de neurones sur la couche cachée. Une solution consiste à multiplexer et à mémoriser les poids et de biais lentement variables au cours du temps.

L'ensemble des cellules peut être amélioré en terme de précision et de linéarité, en faisant appel par exemple à des structures cascodées, des amplificateurs rail-to-rail, etc. Ces améliorations se feront néanmoins au détriment de la simplicité de conception. Une discussion sera menée dans le chapitre V autour du type d'implantation à privilégier par l'intermédiaire de l'étude de la conception d'un neurone intégré en numérique.

IV.6. REFERENCES

- [Antognetti, 1993] : ANTOGNETTI P., MASSOBRIO G., *Semiconductor Device Modelling with SPICE*, McGraw-Hill, New York, 1993.
- [Babanezhad, 1987] : BABANEZHAD J.N., GREGORIAN R., "A Programmable Gain/Loss Circuit", *IEEE Journal of Solid State Circuits*, Vol. 22, n°6, pp1082-1090, Dec 1987.
- [Gatet, 2006(a)] : GATET L., TAP-BETEILLE H., LESCURE M., ROVIRAS D., MALLET A., Design and Test of a CMOS MLP Analog Neural Network for Fast On-Board Signal Processing, *13th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2006)*, Nice, France, Dec. 11-13, 2006, Pages: 922-925.
- [Gatet, 2006(b)] : GATET L., TAP-BETEILLE H., LESCURE M., Design of an analog MLP-type neural network to increase the distance measurement range of a phase-shift laser rangefinder, *WSEAS TRANSACTIONS ON CIRCUITS AND SYSTEMS*, Issue 6, Vol. 5, June 2006, Pages: 821-828.
- [Gatet, 2007(a)] : GATET L., TAP-BETEILLE H., LESCURE M., ROVIRAS D., MALLET A., Functional tests of a 0.6 μ m CMOS MLP Analog Neural Network for fast on-board signal processing, *ICECS 2006 Special Issue of Analog Integrated Circuits and Signal Processing*, Springer Science & Business Media (première lecture).
- [Gatet, 2007(b)] : GATET L., TAP-BETEILLE H., LESCURE M., Analog Neural Network Design for Real-Time Surface Detection with a Laser Rangefinder, *2007 IEEE Instrumentation and Measurement Technology Conference (IMTC 2007)*, Warsaw, Poland, May 1-3, 2007, *en cours de publication*.
- [Gatet, 2007(c)] : GATET L., TAP-BETEILLE H., LESCURE M., Test of a CMOS Analog Neural Network for Real-Time Surface Detection with a Laser Rangefinder, *IEEE Special issue on Instrumentation and Measurement*, première lecture.
- [Geiger, 1990] : GEIGER R., ALLEN P., STRADER N., *Analog Integrated Circuits*, 3th ed., John Wiley and sons, 1990.
- [Gray, 2001] : GRAY P., HURST P., LEWIS S., MEYER R., *Analysis and design of analog integrated circuits*, 4th ed., John Wiley and sons, 2001.
- [Hastings, 2005] : HASTINGS A., *The Art of Analog Layout*, 2nd ed., Prentice Hall, 2005.
- ([Josse, 2003]) : JOSSE S, *Transportabilité de Fonctions Analogiques CMOS en Technologies Submicroniques. Application : Contrôle du Retard des Fronts d'Horloges d'un Imageur CCD*, thèse de doctorat de l'INP, 2003.
- [Sooch, 1985] : SOOCH N.S., MOS Cascode Current Mirror, U.S. patent n° 4550284, Oct. 1985.

Chapitre V

Exploitation des résultats

Ce chapitre est destiné à l'exploitation des tests des circuits neuronaux analogiques au sein des deux applications pour lesquels ils ont été implantés. Une version numérique de la même architecture est ensuite étudiée. Enfin, les performances et les limitations des deux systèmes développés sont résumées en fin de chapitre.

V.1. LEVEE D'INDETERMINATION SUR LA MESURE DE DISTANCE

V.1.1. Entraînement du réseau de neurones

Rappelons brièvement la problématique. Pour une cible donnée, avec un angle d'incidence donné, il n'est pas possible, avec le télémètre seul, d'avoir une très grande résolution sur la mesure de la distance sur un grand intervalle de distance. L'objectif est de garder le même système de mesure permettant d'avoir une grande résolution, et d'insérer un réseau de neurones afin de lever l'indétermination en $2K\pi$ introduite par la mesure du déphasage. Maintenant que le réseau de neurones a été qualifié et que son équation de transfert a été déterminée, il est possible de l'entraîner suivant la méthode décrite dans le paragraphe III.2.2. Pour cela, on injecte les coefficients K_{a1} , K_{a2} , K_{a3} , K_{th} , K_{a4} , K_{sum} et K_{b2} déterminés à partir de l'équation expérimentale (IV-51) dans les équations (III-7) à (III-11). Comme cette application ne nécessite qu'une seule entrée, la deuxième est connectée à la masse ($V_{inNN2} = 0V$) ([Gatet, 2007(a)]).

Une fois les phases d'entraînement et de validation achevées, les valeurs finales des poids et biais trouvés peuvent être directement injectées sous forme de tensions dans le réseau de neurones implanté par l'intermédiaire de potentiomètres. On pourra ainsi vérifier la capacité du réseau de neurones intégré à inverser la loi en " $1/D^2$ ", c'est à dire d'avoir une tension de sortie proportionnelle à la distance. L'intervalle de mesure a été limité à [0.5m ; 3.2m] compte

tenu de la limitation des intervalles de tensions des poids et biais et de la précision de la tension de sortie par rapport à l'équation expérimentale (IV-51) au premier ordre.

Les différentes valeurs de l'ensemble d'entraînement, uniformément réparties en échelle logarithmique, sont au nombre de 9 (figure V-a). Une fois la phase d'entraînement achevée, le réseau de neurones sera testé avec un jeu de valeurs plus important appartenant à l'intervalle de mesure de manière à vérifier sa faculté de généralisation.

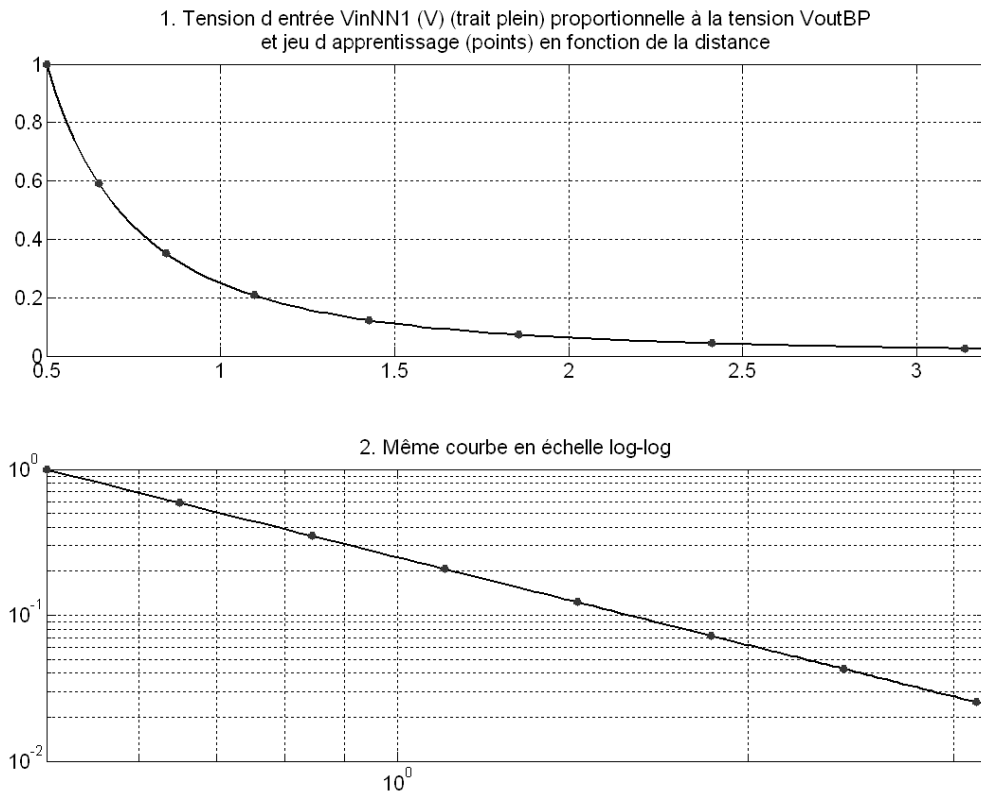


Figure V-a : Courbes représentant la répartition des valeurs des amplitudes d'entrée du réseau de neurones, fonction des valeurs de distance dans l'intervalle [0.5m;3.2m].

Ainsi, à chaque itération k , est choisie une valeur aléatoire de distance $D(k)$, correspondant à une des 9 valeurs de distance de l'ensemble d'entraînement. Cette valeur de distance correspond à une valeur de tension d'entrée expérimentale $V_{inNN1}(k)$ du réseau de neurones proportionnelle à l'amplitude du signal observée en sortie du filtre passe bas. Afin de profiter de la dynamique de $1V_{c\grave{a}c}$ offerte par le circuit neuronal au niveau de chacune de ses entrées, on fait varier la tension d'entrée entre $-0.5V$ et $0.5V$ plutôt qu'entre 0 et 1 V :

$$V_{inNN1}(k) = K'_1 \frac{\rho(k)\cos(\theta_0(k))}{D(k)^2} - 0,5 \quad (V-1)$$

L'objectif de la phase d'entraînement est de faire en sorte que le réseau de neurones réalise l'inverse de la loi "1/D²". La procédure d'entraînement, similaire à celle décrite dans la partie III.2.2, est décrite sur la figure V-b.

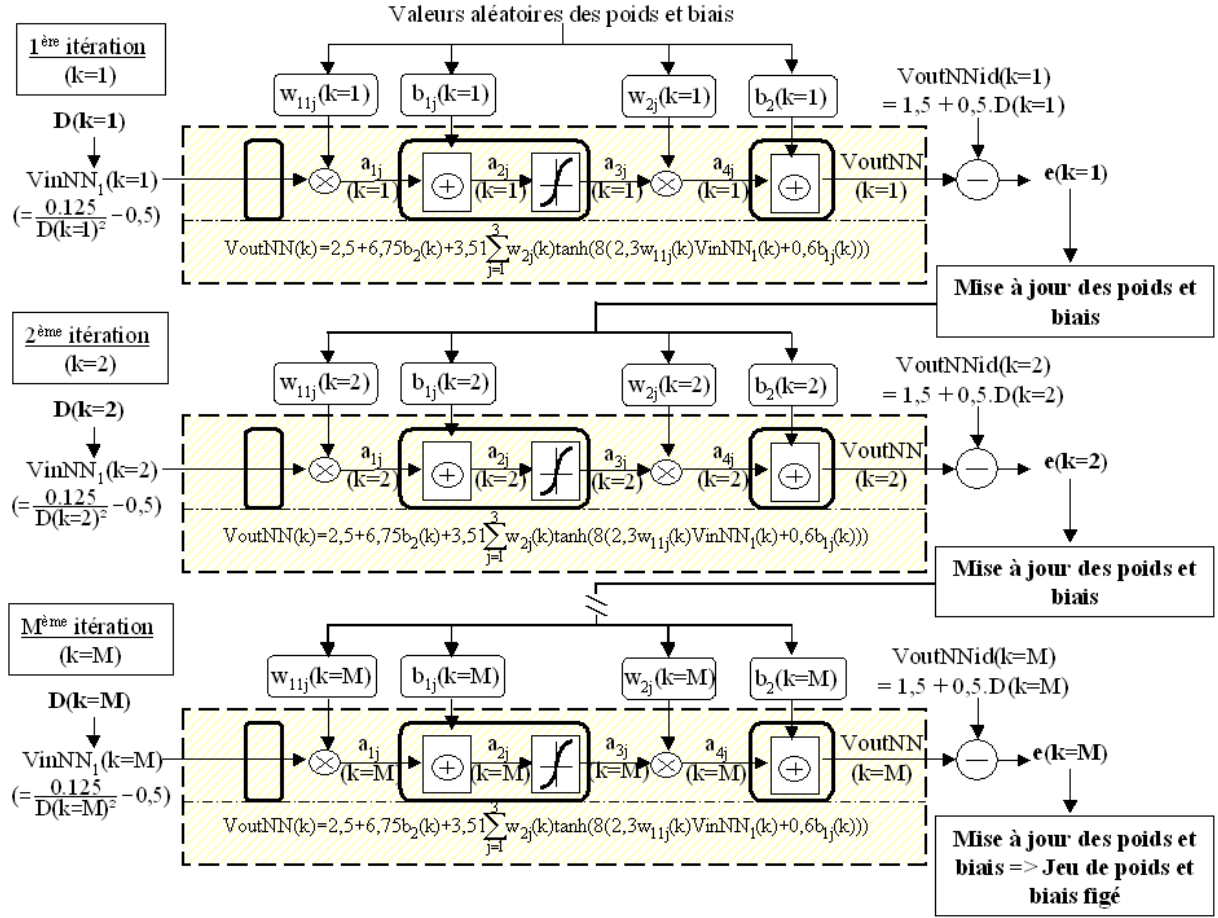


Figure V-b : Description de la phase d'entraînement.

La relation entre la tension de sortie du réseau, son entrée et ses paramètres est :

$$V_{outNN}(k) = 2,5 + 6,75b_2(k) + 3,51 \sum_{j=1}^3 w_{2j}(k) \tanh(8(2,3w_{11j}(k)V_{inNN_1}(k) + 0,6b_{1j}(k))) \quad (V-2)$$

A chaque itération, la tension de sortie $V_{outNN}(k)$ est comparée à la tension idéale de sortie $V_{outNNid}(k)$ proportionnelle à la distance $D(k)$ (figure V-b). De la même façon que pour les entrées, la valeur centrale 2.5V de la tension de sortie a été choisie comme tension idéale obtenue pour une distance de 2m, valeur de distance située à peu près au milieu de l'intervalle de mesure. Pour des question de dynamique, le facteur de proportionnalité entre la tension de sortie idéale souhaitée et la distance a été choisi égal à $0,5V m^{-1}$:

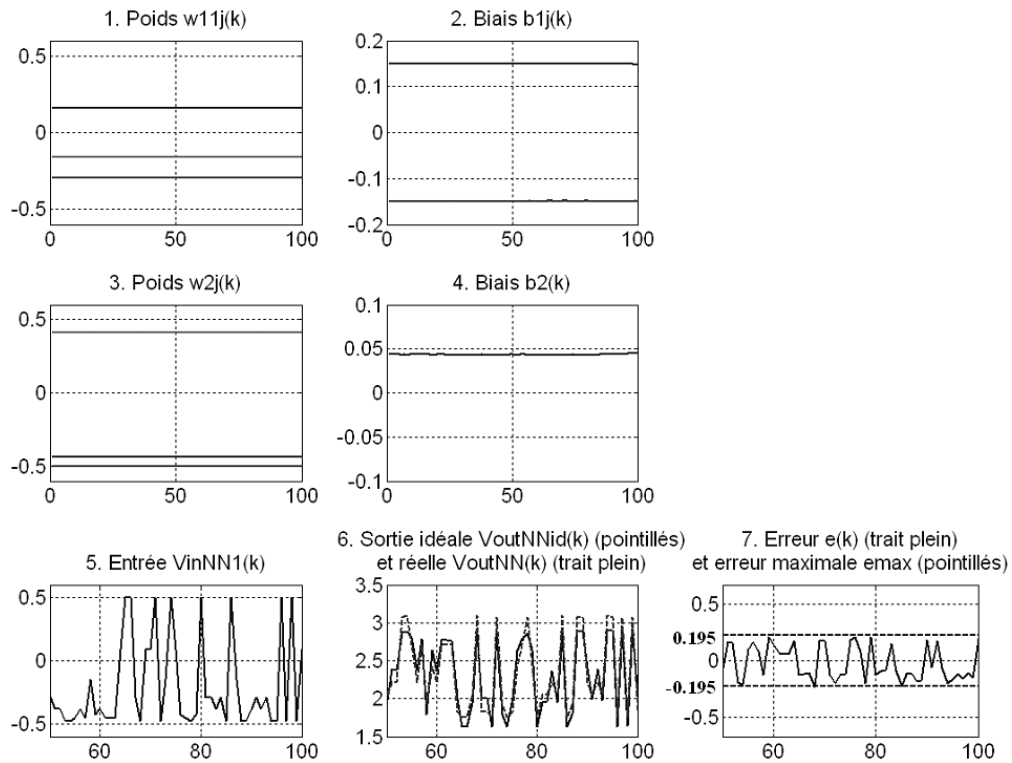
$$V_{outNNid}(k) = 1,5 + 0,5.D \quad (V-3)$$

La différence entre $V_{outNN}(k)$ et $V_{outNNid}(k)$ conduit à une erreur $e(k)$. Cette erreur est comparée à la valeur maximale de l'erreur e_{max} relative au facteur de proportionnalité choisi entre la tension de sortie idéale souhaitée et la distance. Le domaine de mesure Λ est égal à 0.9m. Pour distinguer deux valeurs de distance modulo Λ , il suffit d'avoir une évaluation de la distance à $\pm\Lambda/2$ près, soit $\pm 0,45m$. Or, le facteur de proportionnalité entre la tension de sortie

idéale souhaitée et la distance a été choisi égal à $0.5V\text{m}^{-1}$. Donc la valeur maximale de l'erreur tolérée de manière à évaluer la distance avec une précision suffisante pour lever l'indétermination modulo 2π sur la mesure de phase est égale à $0.5\Lambda/2$, soit 225mV. En tenant compte de l'écart moyen $e_{\text{MEAN}_{\text{nm}}}=30\text{mV}$ constaté entre l'équation (V-2) et la sortie expérimentale du réseau de neurones, l'erreur maximale de sortie e_{max} durant la phase d'entraînement est choisie égale à 195mV.

Les poids et biais sont mis à jour en fonction de la valeur de l'erreur $e(k)$ en utilisant la méthode de rétropropagation de l'erreur ([Rumelhart, 1986], [Werbos, 1974]) associé à l'algorithme de Levenberg-Marquardt ([Levenberg, 1944], [Marquardt, 1963]). La mise à jour des poids et biais s'arrête à partir du moment où l'erreur constatée sur un certain nombre d'itérations est inférieure à 0.195m. Ce nombre d'itérations est choisi cinq fois plus élevé que le nombre de valeurs de distance appartenant à l'ensemble d'entraînement (soit 45 itérations). Ainsi, la probabilité que le réseau effectue l'approximation de la fonction désirée pour l'ensemble des valeurs de distance appartenant à l'ensemble d'apprentissage est proche de 1.

La figure V-c montre un exemple de convergence de l'algorithme. Les variations des poids et biais lors des 100 dernières itérations sont représentées sur les figures V-c.1 à V-c.4. On note que les paramètres n'évoluent pas sur ces dernières itérations, ce qui confirme qu'ils ont bien convergé vers une valeur définitive. La figure V-c.5 montre l'ensemble des valeurs d'entrée $V_{\text{inNN}}(k)$ choisies aléatoirement à chaque itération. La figure V-c.6 représente la tension de sortie $V_{\text{outNN}}(k)$ du réseau de neurones (courbe en trait plein) et la tension de sortie idéale $V_{\text{outNNid}}(k)$ correspondante. La différence $e(k)$ entre les deux courbes a été tracée en trait plein sur la figure V-c.7. On remarque que sur les 50 dernières itérations, l'erreur $e(k)$ ne franchit pas les deux droites tracées en pointillés, relatives à l'erreur maximale $e_{\text{max}}=\pm 195\text{mV}$. Ainsi, pour les 9 valeurs de distance appartenant à l'ensemble d'entraînement comprises entre 0.5m et 3.2m, la tension de sortie du réseau de neurones est proportionnelle à la distance, avec une précision inférieure à 195mV. Comme les valeurs des poids et biais n'ont pas évolué durant ces 45 dernières itérations, cela signifie que le jeu de paramètres final, observé à la dernière itération, permet au réseau de neurones de réaliser la fonction inverse de la loi " $1/D^2$ ", avec une précision suffisante pour lever l'indétermination en $2k\pi$ due à la mesure de déphasage. Ce jeu de paramètres est donc sauvegardé, de manière à valider à l'étape suivante la capacité d'approximation du réseau de neurones pour toute valeur de distance appartenant à l'intervalle [0.5m ;3.2m].



*Figure V-c : De 1. à 4.: Variation des poids et biais lors des 100 dernières itérations.
 5.: Valeurs $V_{inNN}(k)$ présentées en entrée du réseau de neurones lors des 50 dernières itérations, correspondant aux valeurs aléatoires de distance choisies dans l'ensemble d'apprentissage.
 6.: Sorties idéale $V_{outNNid}(k)$ (pointillés) et réelle $V_{outNN}(k)$ (trait plein) lors des 50 dernières itérations.
 7. Erreur $e(k)$ (trait plein) et erreur maximale e_{max} (pointillés).*

V.1.2. Validation comportementale

La phase de validation est réalisée avec la même équation du réseau de neurones, issue de mesures expérimentales, et avec les poids et biais obtenus à la fin de la phase d'entraînement. 90 valeurs de distances, soit 10 fois le nombre de valeurs de distance appartenant au jeu d'entraînement, sont choisies uniformément réparties entre 0.5m et 3.2m, de façon à tester la faculté de généralisation du réseau de neurones. La figure V-d présente les résultats issus de cette phase de validation. Les différentes valeurs issues de la phase d'entraînement ont été mises en valeur par des points. La tension d'entrée V_{inNN1} du réseau a été représentée en fonction de la distance sur la figure V-d.1. La tension de sortie correspondante V_{outNN} est tracée sur la figure V-d.2. La courbe en pointillés est relative à la tension de sortie idéale $V_{outNNid}$, linéaire par rapport à la distance (V-3). La différence entre les deux courbes est représentée sur la figure V-d.3. Comme la courbe en trait plein ne franchit pas les limites, tracées en pointillés, relatives à l'erreur maximale $e_{max} = \pm 195mV$, cela signifie que le réseau de neurones est capable d'approcher la fonction inverse de la loi "1/D²" avec une précision

suffisante sur l'intervalle entier [0.5m;3.2m]. La capacité de généralisation du réseau de neurones est donc vérifiée.

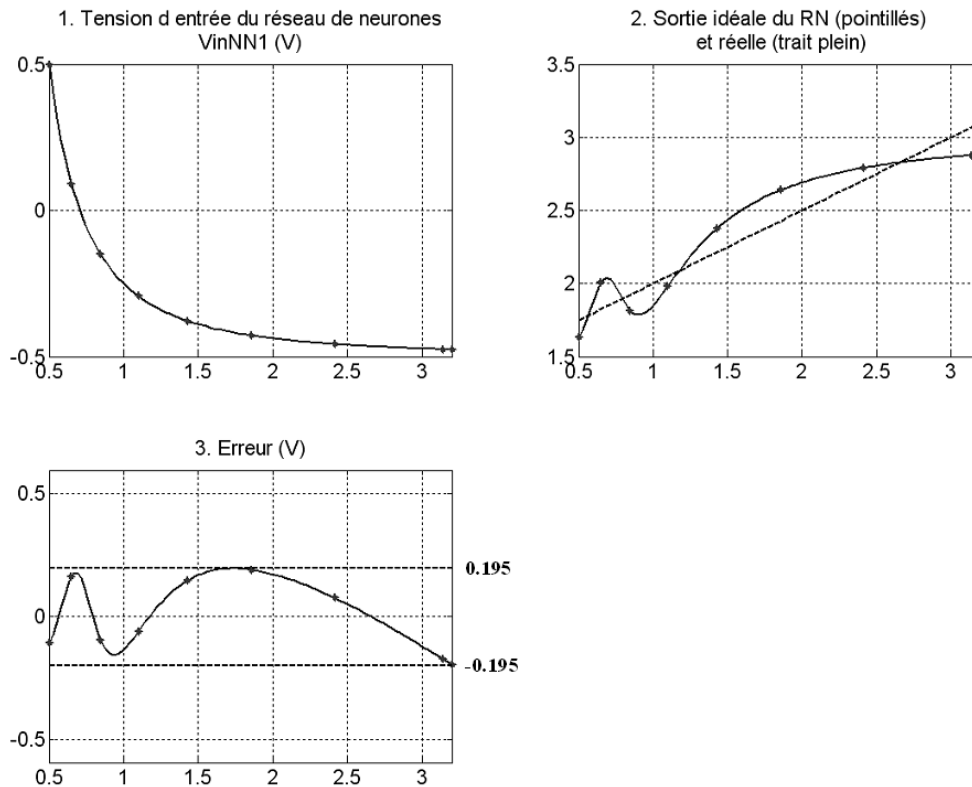


Figure V-d : 1.: Ensemble des valeurs d'entrée V_{inNN1} du réseau. Les valeurs relatives à l'ensemble d'apprentissage sont représentées par des points.
 2.: Tensions de sortie idéale $V_{outNNid}$ (pointillés) et réelle V_{outNN} (trait plein) du réseau de neurones.
 3.: Erreur (trait plein) ; l'erreur maximale $\pm e_{max} = \pm 195mV$ est représentée en pointillés.

V.1.3. Calibrage du réseau de neurones sur PCB – Tests

Comme cela a été mentionné en début du paragraphe V.1, le réseau de neurones a été entraîné à l'aide de l'équation du réseau de neurones déterminée à partir de mesures expérimentales ([Gatet, 2007(a)]). Les poids et biais issus des phases d'entraînement et de validation peuvent donc être directement fixés sur le PCB sur lequel le réseau a été implanté (figure IV-ad). L'amplitude V_{outBP} du signal issu de la sortie du filtre passe-bande du télémètre laser est convertie en une tension continue de manière à être transmise à l'entrée V_{inNN1} du réseau de neurones (figure V-e.1). Les valeurs issues de la phase d'entraînement ont été mises en valeur par des points. La figure V-e.2 présente une comparaison entre la tension de sortie idéale du réseau de neurones (pointillés), les tensions de sortie issues des phases d'entraînement et de validation (respectivement les points et la courbe en trait plein) et la tension de sortie expérimentale du réseau de neurones implanté (croix). On note que les

valeurs expérimentales sont quasiment confondues avec les valeurs issues de la simulation. La figure V-e.3 montre la différence entre simulation et test expérimental. Les deux droites en pointillés sont relatives à l'écart moyen $e_{MEAN_{nn}} = \pm 30\text{mV}$ constaté entre l'équation (IV-51) et l'ensemble des points expérimentaux ayant servi à tracer les caractéristiques d.c. du réseau de neurones. On note que pour cette application, l'erreur entre simulation et test ne dépasse pas 30mV en valeur absolue. La figure V-e.4 représente les erreurs observées en simulation (trait plein) et en test (croix) par rapport à l'idéalité. Comme les erreurs relatives aux points expérimentaux sont toutes inférieures à 225mV en valeur absolue, cela signifie que le réseau de neurones analogique implanté sur PCB est capable d'approcher la fonction inverse de la loi "1/D²" avec une précision suffisante sur l'intervalle [0.5m;3.2m].

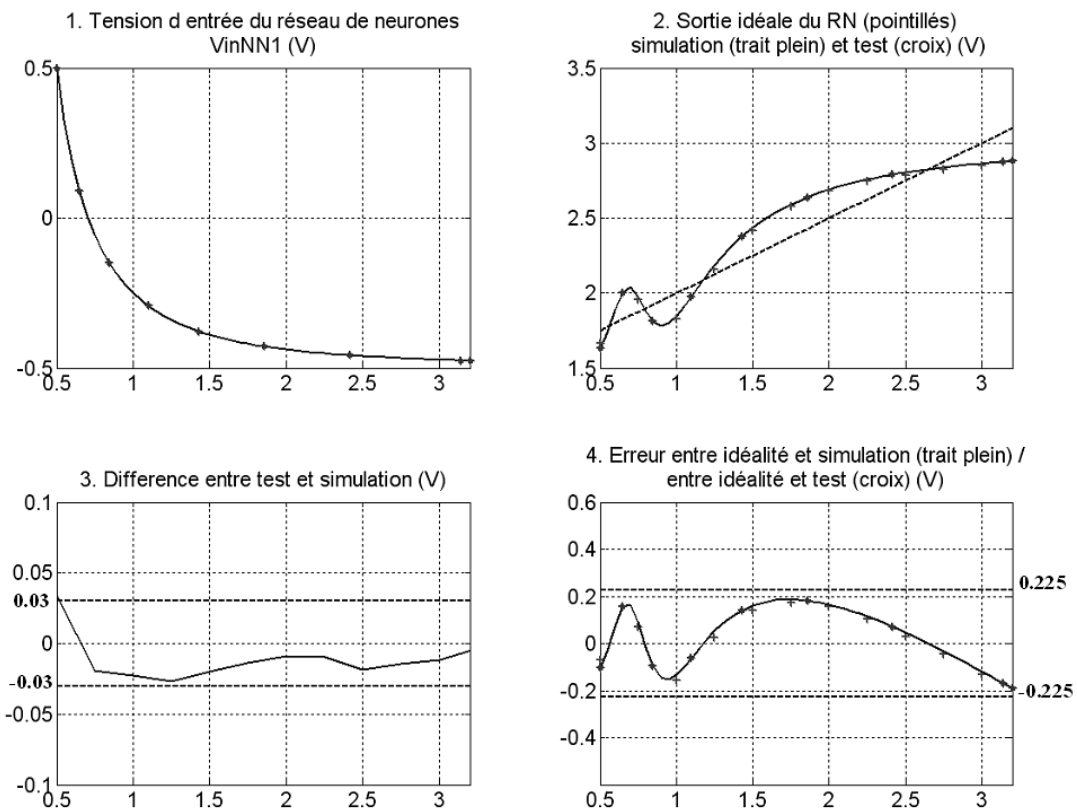


Figure V-e : Comparaison entre résultats expérimentaux, simulation et idéalité.

Ainsi, il est possible d'obtenir à partir de la tension de sortie du réseau de neurones une valeur de la distance approximative, mais suffisamment précise pour lever l'indétermination sur la mesure de distance issue de la mesure du déphasage sur l'intervalle [0.5m;3.2m]. En combinant les informations sur la distance issues des sorties du télémètre et du réseau de neurones, il est donc possible d'obtenir une mesure de distance de précision (50 μm) sur un intervalle de mesure 3 fois plus important que celui limité à la mesure du déphasage.

V.2. CLASSIFICATION DE SURFACES

V.2.1. *Entraînement du réseau de neurones*

Rappelons brièvement la problématique. La détection de surface à l'aide du seul signal de sortie du filtre passe bande V_{outBP} issu du télémètre s'avère très peu intéressante du fait de la limitation de l'intervalle de mesure, et ce même pour des surfaces très éloignées les unes des autres d'un point de vue coefficient de réflexion diffusant et pour un angle d'incidence fixe. L'objectif est donc de se servir de l'information contenue dans l'amplitude du signal V_{outBP} quant à la nature de la surface testée, et d'étendre l'intervalle de mesure par l'intermédiaire du signal de sortie $V_{outRANG}$ directement proportionnel à la distance, tout en conservant la même structure du télémètre laser par déphasage et en se permettant une variation de l'angle d'incidence du télémètre par rapport à la cible. Maintenant que le réseau de neurones a été qualifié et que son équation de transfert a été déterminée, il est possible de l'entraîner suivant la méthode décrite dans le paragraphe III-3.2 appliquée au réseau de neurones conçu ([Gatet, 2007(b)]). Pour cela, on injecte les coefficients K_{a11} , K_{a12} , K_{a2} , K_{a3} , K_{th} , K_{a4} , K_{sum} et K_{b2} déterminés à partir de l'équation expérimentale (IV-51) dans les équations (III-15) à (III-19).

Une fois les phases d'entraînement et de validation achevées, les valeurs finales des poids et biais trouvés peuvent être directement injectées sous forme de tensions dans le réseau de neurones implanté par l'intermédiaire de potentiomètres. On pourra ainsi vérifier la capacité du réseau de neurones intégré à discriminer les quatre types de surface choisis pour une distance et un angle d'incidence variables. Les intervalles de mesure de distance et d'angle d'incidence ont été limités respectivement à $[0.5m;1.25m]$ et $[-\pi/6;\pi/6]$ compte tenu de la limitation des intervalles de tensions des poids et biais et de la précision de la tension de sortie par rapport à l'équation expérimentale (IV-51) au premier ordre. Le réseau de neurones est calibré de manière à ce que sa tension de sortie appartienne à l'intervalle relatif au type de surface testé (tableau V- α).

Tableau V- α : Tensions de sortie analogiques idéales du réseau de neurones, intervalles de sortie et tensions de sortie idéales des différents détecteurs de seuil associés.

Surface	Type	Tension de sortie idéale	Intervalle de sortie associé	Sortie du premier détecteur de seuil V_{out_1} (2.9V)	Sortie du second détecteur de seuil V_{out_2} (2.5V)	Sortie du troisième détecteur de seuil V_{out_3} (2.1V)
S_1	Surface plastique	3.1]2.9;3.1['1'	'1'	'1'
S_2	Papier glacé	2.7]2.5;2.9['0'	'1'	'1'
S_3	Mur peint	2.3]2.1;2.5['0'	'0'	'1'
S_4	Surface poreuse	1.9]1.9;2.1['0'	'0'	'0'

Dans l'absolu, le signal de sortie du réseau de neurones peut varier entre 0 et 5V (tension d'alimentation de la couche de sortie). On note qu'à toute tension supérieure à 2.9V est associé le code '111' relatif aux trois sorties à l'état haut des comparateurs de seuils, et qu'à toute tension inférieure à 2.1V est associé le code '000' relatif aux trois sorties à l'état bas des comparateurs de seuils. On peut donc limiter l'intervalle associé à la tension de sortie à [1.9V;3.1V].

Les tensions idéales de sortie associées aux quatre types de surface S_1 à S_4 (surface plastique, papier glacé, mur peint et surface poreuse) sont respectivement égales à 3.1V, 2.7V, 2.3V et 1.9V. Ces tensions sont situées au milieu de l'intervalle associé au type de surface, sauf pour S_1 et S_4 pour lesquelles la tension idéale est confondue avec une limite de l'intervalle de la tension de sortie du réseau de façon à accroître la dynamique des signaux (tableau V- α). Les tensions seuils associées aux trois détecteurs de seuil sont choisies à équidistance des tensions idéales associées à chaque surface, soit 2.9V, 2.5V et 2.1V. Ces tensions correspondent aux limites des intervalles communes à deux types de surfaces. A chaque type de surface est associé un code "binaire" relatif aux sorties des trois détecteurs de seuil. La sortie de chaque détecteur dépend de l'intervalle auquel appartient la sortie du réseau de neurones (tableau V- α).

L'objectif de cette application est de discriminer quatre types de surface. Chaque type de surface a un coefficient de réflexion diffusant ρ propre. Les coefficients associés à la surface plastique, au papier glacé, au mur peint et à la surface poreuse, déterminés expérimentalement, sont respectivement égaux à 0.9, 0.6, 0.35 et 0.15. A chaque itération k de l'algorithme, le type de surface est choisi aléatoirement. En d'autres termes, une valeur de coefficient de réflexion diffusant $\rho(k)$ est choisie à chaque itération parmi les 4 valeurs de

coefficient de réflexion diffusant apparent $\{0.15; 0.35; 0.6; 0.9\}$ constituant le jeu d'entraînement associé au type de surface. Seules deux valeurs d'angle d'incidence ont été choisies pour faire partie du jeu d'entraînement relatif à l'angle d'incidence : 0 et $|\theta_{\max}|=\pi/6$. A chaque itération, une valeur d'angle $\theta_0(k)$ est choisie parmi ces deux valeurs. Pour confirmer la capacité de discrimination du réseau pour tout angle appartenant à l'intervalle $[-\pi/6; \pi/6]$, des valeurs intermédiaires seront choisies au moment de la phase de validation. Concernant le jeu d'entraînement relatif à la distance, 7 valeurs de distance ont été réparties entre 0.5m et 1.25m. Cette répartition a été optimisée pour minimiser le nombre de valeurs de distance, afin de diminuer les temps de simulation. Le nombre de valeurs de distance sera augmenté lors de la phase de validation de façon à vérifier que le réseau classe bien les différentes surfaces pour toute valeur de distance appartenant à l'intervalle de mesure.

La figure V-f présente l'ensemble des tensions V_{inNN_1} associées aux quatre types de surface, à cinq valeurs d'angles d'incidence comprises entre 0 et $\pi/6$ et à l'ensemble des valeurs de distance comprises entre 0.5m et 1.25m. Les points mettent en valeur l'ensemble des tensions V_{inNN_1} correspondant aux différentes valeurs d'entraînement associées aux quatre types de surfaces et aux deux variables D et θ_0 .

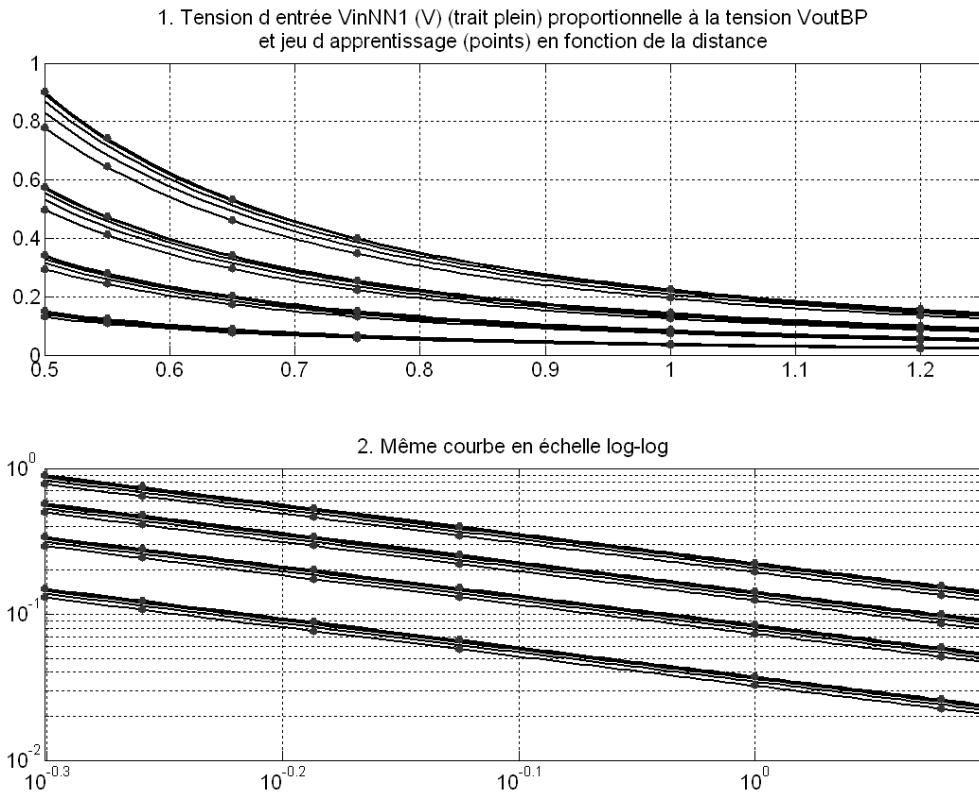


Figure V-f : Courbes représentant la répartition des valeurs des tensions d'entrée V_{inNN1} du réseau de neurones, fonction du type de surface, de l'angle d'incidence variant de $-\pi/6$ à $\pi/6$ et des mesures de distance choisies dans l'intervalle $[0.5m; 1.25m]$. Les points correspondent aux tensions d'entrée associées aux trois ensembles d'entraînement.

Ainsi, à chaque itération k , trois valeurs aléatoires sont choisies, correspondant à une des 4 valeurs de coefficient de réflexion diffusant apparent de la cible non coopérative testée, à un des 2 angles d'incidence 0 ou $\pi/5$, et à une des 7 valeurs de distance appartenant à l'ensemble d'entraînement correspondant. Ces trois valeurs aléatoires permettent de définir deux valeurs de tensions d'entrée du réseau $V_{inNN1}(k)$ et $V_{inNN2}(k)$, respectivement proportionnelles aux signaux de sortie du filtre passe bande et du télémètre. Afin de profiter de la dynamique de 1V crête à crête offerte par le circuit neuronal au niveau de chacune de ses entrées, on fait varier ces deux tensions d'entrée entre $-0.5V$ et $0.5V$ plutôt qu'entre 0 et 1 V :

$$V_{inNN1}(k) = K'_1 \frac{\rho(k)\cos(\theta_0(k))}{D(k)^2} - 0,5 \quad (V-4)$$

$$V_{inNN2}(k) = K'_2 D(k) - 0,5 \quad (V-5)$$

L'objectif de la phase d'entraînement est de faire en sorte que le réseau de neurones classe chaque jeu d'entrée relatif à un type de surface dans l'intervalle de sortie qui lui a été attribué. La procédure d'entraînement, similaire à celle décrite dans la partie III.3.2, est décrite sur la figure V-g.

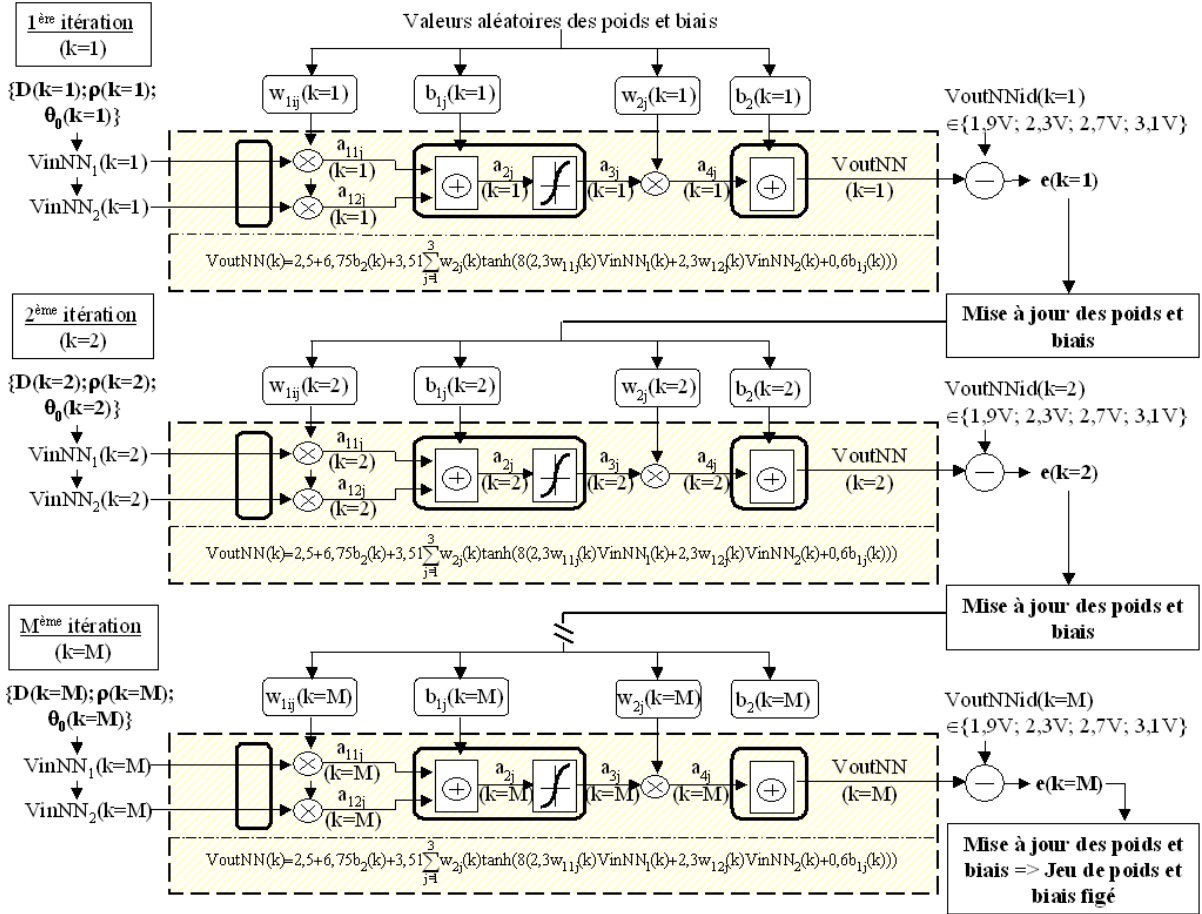


Figure V-g : Description de la méthode utilisée pour calculer les poids et biais.

La relation entre la tension de sortie du réseau, ses entrées et ses paramètres est :

$$V_{outNN}(k) = 2,5 + 6,75b_2(k) + 3,51 \sum_{j=1}^3 w_{2j}(k) \tanh(8(2,3w_{11j}(k)V_{inNN_1}(k) + 2,3w_{12j}(k)V_{inNN_2}(k) + 0,6b_{1j}(k))) \quad (V-6)$$

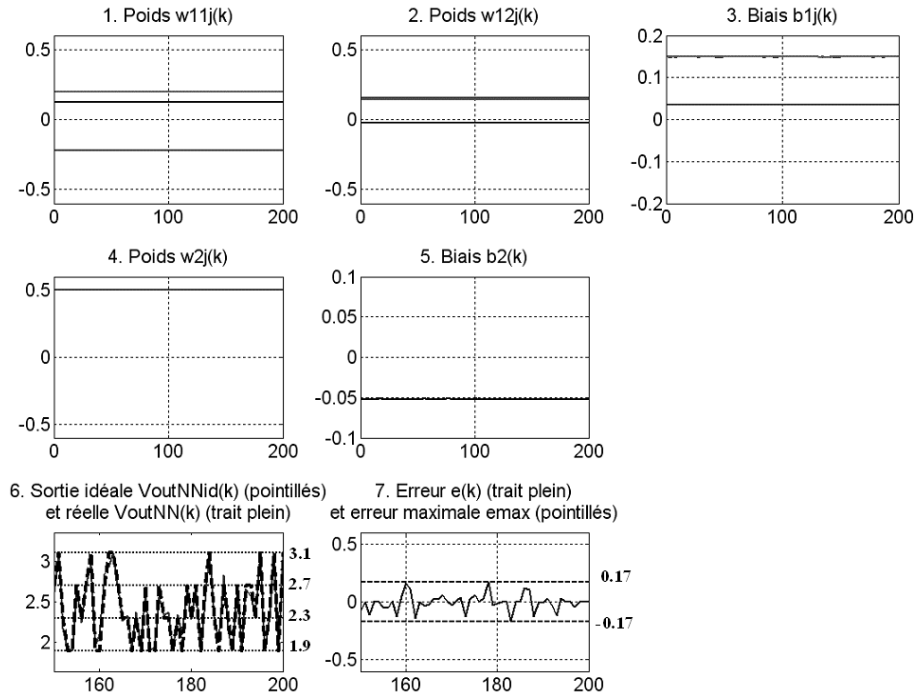
À chaque itération, la tension de sortie $V_{outNN}(k)$ est comparée à la tension idéale de sortie $V_{outNNid}(k)$ correspondant à la surface testée à l'itération k (figure V-g). La différence entre $V_{outNN}(k)$ et $V_{outNNid}(k)$ conduit à une erreur $e(k)$. Cette erreur est comparée à la valeur maximale de l'erreur e_{max} , correspondant à la différence entre la tension de sortie idéale et la tension de sortie limite commune à une autre surface. En tenant compte de l'écart moyen $e_{MEANnn} = 30mV$ constaté entre l'équation (V-6) et la sortie expérimentale du réseau de neurones, l'erreur maximale de sortie e_{max} durant la phase d'entraînement est choisie égale à 170mV. Une erreur binaire $e_{bin}(k)$ est déduite de la différence entre l'erreur $e(k)$ et e_{max} . Si cette différence est positive, l'erreur binaire est égale à '1', sinon, elle est égale à '0'. Cette erreur permet de déterminer facilement si la bonne surface a été détectée ou non. Les poids et biais sont mis à jour en fonction de la valeur de l'erreur $e(k)$. La méthode de mise à jour est la

méthode de rétropropagation de l'erreur ([Rumelhart, 1986], [Werbos, 1974]), utilisée avec l'algorithme de Levenberg-Marquardt ([Levenberg, 1944], [Marquardt, 1963]). La mise à jour des poids et biais s'arrête à partir du moment où l'erreur binaire constatée sur un certain nombre d'itérations est égale à '0'. Ce nombre d'itérations est choisi égal à 200, soit un peu plus de trois fois le nombre de combinaisons entre le nombre de surfaces, le nombre d'angles d'incidence et le nombre de distance appartenant à leurs jeux d'entraînement respectifs. Ainsi, la probabilité que le réseau effectue sa tâche de discrimination pour l'ensemble des valeurs des 3 ensembles d'apprentissage est proche de 1.

Tant que l'erreur binaire n'a pas été constatée égale à '0' durant 200 itérations successives, la simulation se poursuit. Ce critère est déterminant de manière à être sûr que le réseau de neurones effectue bien sa tâche de discrimination pour toutes les valeurs associées aux différents jeux d'entraînement, dans l'optique que cette faculté de classification soit bien confirmée durant la phase de validation. Le temps de convergence associé à l'application de détection de surface est donc nettement plus important que celui associé à la levée d'indétermination sur la mesure de distance.

Pour améliorer le temps de convergence, seules les 200 dernières valeurs mises en jeu durant la phase d'entraînement sont sauvegardées à chaque itération. La figure V-h montre un exemple de convergence de l'algorithme. Les variations des poids et biais lors des 200 dernières itérations ont été représentées sur les figures V-h.1 à V-h.5. On note que leurs valeurs n'évoluent pas sur ces dernières itérations, ce qui prouve qu'elles ont atteint leur valeur définitive. Un zoom sur les 50 dernières itérations a été effectué sur les courbes V-h.6 et V-h.7. La courbe V-h.6 représente les variations de la sortie idéale $V_{outNNid}(k)$ (en pointillés) et réelle $V_{outNN}(k)$ (en trait plein) du réseau, en fonction des valeurs des entrées $V_{inNN_1}(k)$ et $V_{inNN_2}(k)$, pour un type de surface, un angle d'incidence et une distance aléatoires. Sur la figure V-h.7, on note que la différence entre les deux courbes (trait plein) ne dépasse pas la valeur limite $e_{max}=\pm 0.17V$. Ainsi, sur les 50 dernières itérations, le réseau de neurones est capable de discriminer chaque surface des trois autres sans erreur pour les 2 valeurs d'angles d'incidence et les 7 valeurs de distance appartenant à leur jeu d'entraînement respectif. Ce résultat peut être étendu aux 200 dernières itérations, puisque les courbes ne s'affichent qu'à partir du moment où les 200 dernières valeurs de $e_{bin}(k)$ ont été constatées égales à '0'. Comme les valeurs des poids et biais n'ont pas évolué durant ces 200 dernières itérations, cela signifie que le jeu de paramètres final, observé à la dernière itération, permet au réseau de neurones de discriminer les 4 surfaces les unes des autres sans erreur. Ce jeu de paramètres est donc sauvegardé, de manière à valider à l'étape suivante la capacité de classification du réseau de

neurones pour tout angle d'incidence et pour toute valeur de distance appartenant respectivement aux intervalles $[-\pi/6 ; \pi/6]$ et $[0.5m ; 1.25m]$.



*Figure V-h : De 1. à 5.: Variation des poids et biais lors des 200 dernières itérations.
6.: Sorties idéale $V_{outNNid}(k)$ (pointillés) et réelle $V_{outNN}(k)$ (trait plein) lors des 50 dernières itérations. Les deux courbes sont quasiment superposées.
7. Erreur $e(k)$ (trait plein) et erreur maximale $e_{max} = \pm 170mV$ (pointillés).*

V.2.2. Validation comportementale

La phase de validation est réalisée avec la même équation du réseau de neurones, issue de mesures expérimentales, et avec les poids et biais obtenus à la fin de la phase d'entraînement. Chaque surface est testée pour 5 valeurs d'angles d'incidence uniformément réparties entre 0 et $\pi/6$. Pour chaque angle d'incidence, 35 valeurs de distance (soit 5 fois plus que durant la phase d'entraînement) sont choisies réparties uniformément entre 0.5m et 1.25m de façon à tester la faculté de généralisation du réseau de neurones. La figure V-i présente un exemple de résultats issus de cette phase de validation pour chaque type de surface. Les différentes valeurs issues de la phase d'entraînement ont été mises en valeur par des points. Les tensions d'entrée V_{inNN_1} et V_{inNN_2} du réseau ont été représentées en fonction de la distance pour les différentes valeurs d'angle d'incidence sur les figures V-i.1 et V-i.2. Les courbes sont toutes superposées sur les figures V-i.2 car la tension de sortie du télémètre laser est indépendante de l'angle d'incidence. La tension de sortie correspondante V_{outNN} est tracée sur les figures V-i.3. Chaque courbe en trait plein est relative à un angle d'incidence. La droite en pointillés gras montre la tension idéale $V_{outNNid}$ relative à la surface testée. Les autres courbes

pointillées montrent les frontières entre chaque intervalle relatif à un type de surface (tableau V- α). On note que l'ensemble des courbes appartient à l'intervalle relatif au type de surface testé. Pour le confirmer de manière simple, la valeur absolue de la différence entre chaque courbe V_{outNN} et la tension idéale $V_{outNNid}$ est représentée sur les figures V-i.4. La courbe en pointillés est relative à l'erreur maximale $e_{max}=0.17V$ qui ne doit pas être franchie de manière à ce que chaque surface soit bien discriminée des autres. Comme chaque courbe en trait plein, relative à un angle d'incidence, est située en dessous de l'erreur maximale pour toute valeur de distance entre 0.5m et 1.25m, l'erreur binaire représentée sur les figures V-i-5 est égale à '0' pour chacune des surfaces. Cela signifie que le réseau de neurones est capable de discriminer chaque type de surfaces des autres pour un angle d'incidence variant entre $-\pi/6$ et $\pi/6$, dans l'intervalle de distance [0.5m;1.25m]. La capacité de généralisation du réseau de neurones à partir des résultats de la phase d'apprentissage est donc vérifiée.

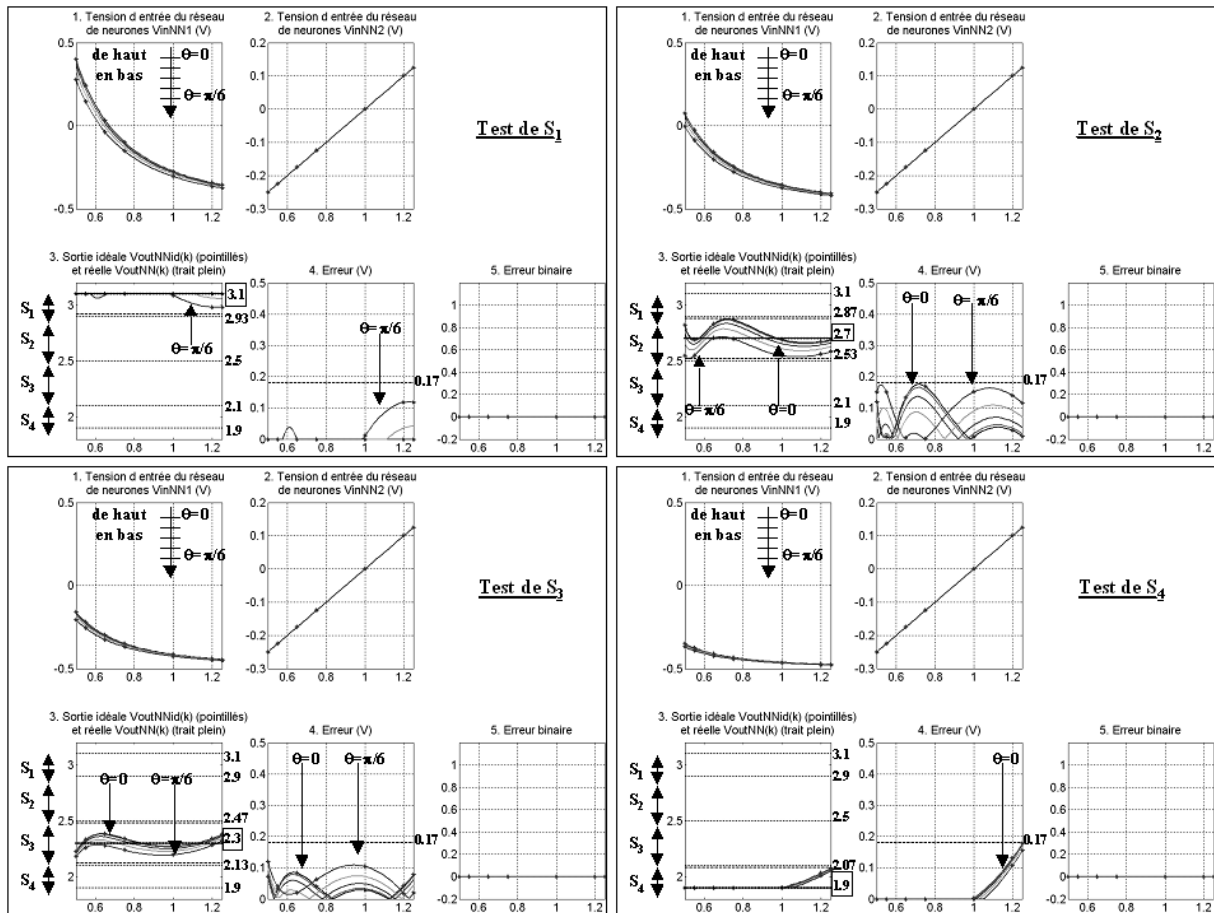


Figure V-i : 1. et 2.: Ensemble des valeurs d'entrée V_{inNN1} et V_{inNN2} du réseau relatives aux 35 valeurs de distance et aux 5 valeurs d'angle d'incidence appartenant à leur jeu de validation respectif. Les valeurs appartenant à l'ensemble d'apprentissage sont représentées par des points.

3.: Tensions de sortie idéale relative à la surface testée ($V_{outNNid}$ - trait gras) et réelle (résultat de l'équation (V-6)) du réseau de neurones.

4.: Erreur en valeur absolue (trait plein) et erreur maximale $e_{max}=0.17V$.

5.: Erreur binaire.

V.2.3. *Calibrage du réseau de neurones sur PCB – Tests*

Comme cela a été mentionné en début du paragraphe V.2, le réseau de neurones a été entraîné à l'aide de l'équation du réseau de neurones déterminée à partir de mesures expérimentales ([Gatet, 2007(b)]). Les poids et biais issus des phases d'entraînement et de validation peuvent donc être directement fixés sur le PCB sur lequel le réseau a été implanté (figure IV-ad). L'amplitude V_{outBP} du signal issu de la sortie du filtre passe-bande du télémètre laser et le signal $V_{outRANG}$ obtenu en sorti du télémètre laser sont convertis en une tension continues de manière à être transmis aux deux entrées du réseau de neurones (figures V-j.1 et V-j.2). Les valeurs issues de la phase d'entraînement ont été mises en valeur par des points. Les figures V-j.3 présentent une comparaison entre la tension de sortie idéale du réseau de neurones (pointillés gras), les tensions de sortie issues des phases d'entraînement et de validation (respectivement les points et la courbe en trait plein) et la tension de sortie expérimentale, relative au réseau de neurones implanté (croix). On note que les valeurs expérimentales sont quasiment confondues avec les valeurs issues de la simulation. Les figures V-j.4 montrent la différence entre simulation et test expérimental. Les deux droites en pointillés correspondent à l'erreur moyenne $e_{mean} = \pm 30mV$ constatée entre l'équation (IV-51) et l'ensemble des points expérimentaux ayant servi à tracer les caractéristiques du réseau de neurones. On note que pour cette application, certaines valeurs d'erreurs sont supérieures à 30mV en valeur absolue. Néanmoins, cela n'a aucune conséquence sur l'aptitude du réseau à classer les surfaces sans erreur. En effet, les figures V-j.5 représentent les erreurs observées en simulation (trait plein) et en test (croix) par rapport à l'idéalité. Comme les erreurs relatives aux points expérimentaux sont toutes inférieures à 200mV en valeur absolue, l'erreur binaire relative au test représentée sur les figures V-i-6 est égale à '0' pour chacune des surfaces. En d'autres termes, le "code" indiqué par les DEL en sortie des détecteurs de seuil correspond au type de surface testé. Cela signifie que le réseau de neurones analogique implanté sur PCB est capable de discriminer chaque type de surfaces des autres pour un angle d'incidence variant entre $-\pi/6$ et $\pi/6$, dans l'intervalle de distance [0.5m ; 1.25m].

Chap V : Exploitation des résultats

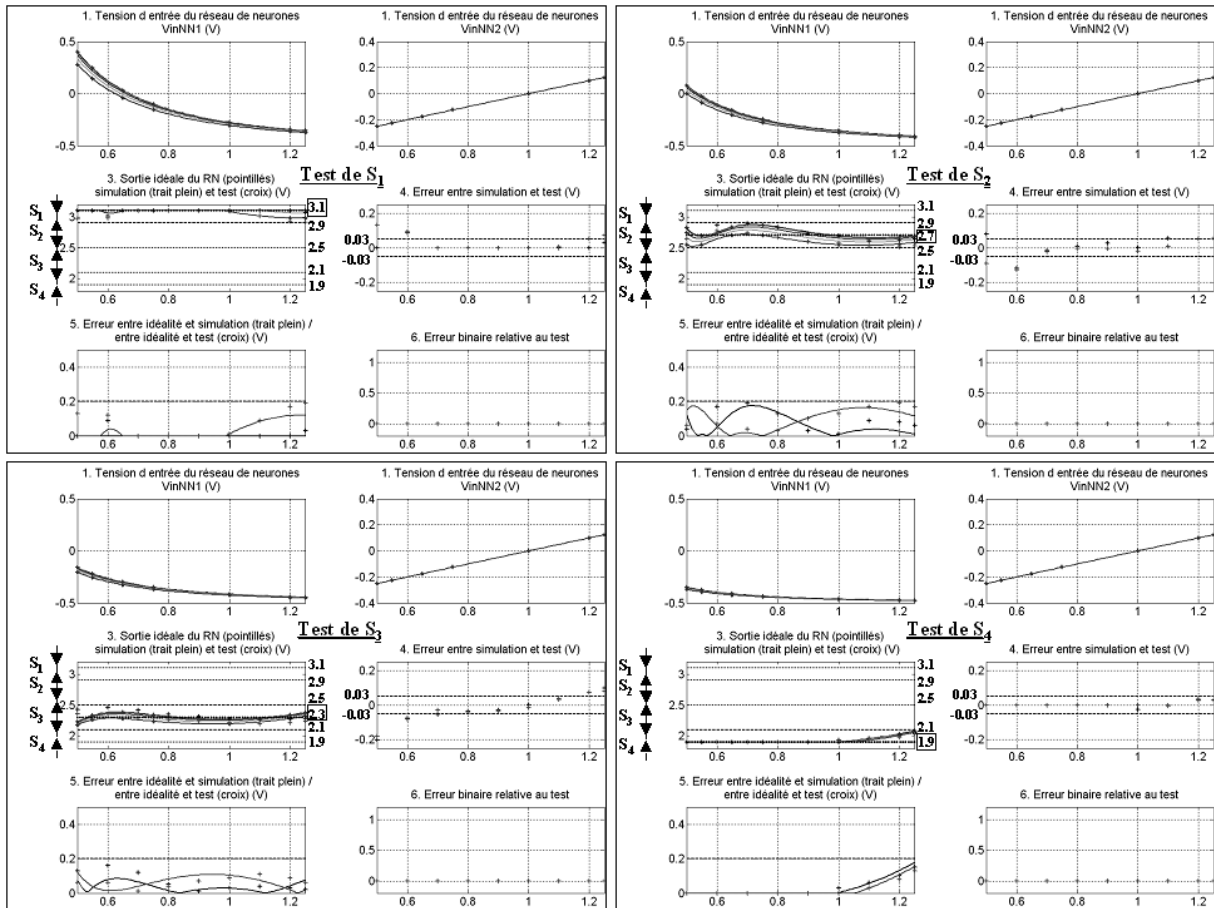


Figure V-j : Comparaison entre résultats expérimentaux, simulation et idéalité.

V.3. COMPARAISON ANALOGIQUE - NUMERIQUE

V.3.1. Conception numérique

L'intégration d'un neurone numérique de type perceptron multi-couches a été étudiée afin de déterminer les avantages et les inconvénients associés aux deux techniques de conception analogiques et numériques. Cette étude fait suite à celle déjà réalisée dans les paragraphes I.5.2.1 et I.5.2.2 à partir d'exemples issus de la littérature.

Il existe trois types de réalisation d'un neurone numérique : logicielle, matérielle, et mixte (de type *co-design*).

La solution logicielle, plus souple en terme de programmation de l'algorithme, permet d'étudier plusieurs configurations du neurone sans modification du support matériel. Elle inclue au moins un processeur de calcul, type DSP ou disposant d'une unité assignée aux nombres réels. Cette solution nécessite le pilotage des composants d'interfaces tels que les convertisseurs analogiques-numériques et numériques-analogiques. L'utilisation d'une mémoire est nécessaire de manière à implanter le programme.

La solution matérielle présente un inconvénient majeur lié au codage de l'information. En effet, généralement, du fait de leur complexité, les blocs numériques réalisés ne traitent pas les nombres en virgule flottante. Cela se traduit par un codage en virgule fixe, entraînant des erreurs de troncatures tout au long du calcul. En revanche, cette solution est la plus rapide en terme de performances dynamiques. En effet, tous les blocs arithmétiques étant réalisés sous forme de composants combinatoires, le retard engendré n'est dû qu'à la propagation des signaux.

La solution mixte n'a pas été envisagée ici car le système numérique à concevoir n'impose pas de fortes contraintes au niveau temporel ou en terme de précision, nécessitant l'utilisation du codage en virgule flottante.

Pour effectuer une comparaison avec le neurone analogique intégré, nous nous sommes donc orientés vers la solution matérielle de type FPGA. Le neurone numérique développé est constitué des mêmes fonctions élémentaires que le neurone analogique précédemment implanté (figure IV-v). La phase de conception peut se décomposer en plusieurs parties :

- Choix du composant,
- Synthèse des différentes fonctions – Simulations pré-routage,
- Placement – Routage – Simulations post-routage,
- Conception de la carte de test - Implantation du neurone – Test fonctionnel.

- Le choix du composant a été effectué en fonction des performances requises relatives au fonctionnement du neurone, du nombre de cellules impliquées (appelées blocs logiques, ou *slices*) et du nombre d'entrées-sorties nécessaires. L'ensemble de ces paramètres dépend de la complexité des structures des différentes fonctions synthétisées et du codage choisi. Pour pouvoir effectuer une comparaison avec le neurone analogique conçu, la dynamique associée à chacune des entrées et chacun des paramètres est égale à 1V. Pour avoir une précision de l'ordre de 1mV, les entrées et les paramètres du neurone ont été codés en 10 bits. Ce choix implique l'intervention de convertisseurs analogiques – numériques 10 bits au niveau des deux entrées du neurone, dont il faudra tenir compte dans les performances de la structure globale.

- La synthèse des différents composants a été réalisée en VHDL. Les différentes opérations sont effectuées de manière asynchrone, de façon à maîtriser le retard global de la structure neuronale. Une structure "pipeline" pourra être intégrée dans un second temps de manière à optimiser la bande passante. Les entrées/sorties de chaque multiplieur et de chaque additionneur ont été codées en 16 bits. L'augmentation de la résolution en interne du réseau est motivée par la perte de précision en sortie de chaîne, causée par les troncatures faites lors de chaque opération élémentaire. Le nombre de blocs logiques élémentaires impliqués dans les synthèses du multiplieur et de l'additionneur 16 bits dans un FPGA Xilinx sont respectivement égaux à 50 et 8, répartis respectivement sur 93 et 16 LUTs (*look-up tables* - table de mémoire ROM à 4 entrées réalisant des fonctions combinatoires).

A noter que le signal de sortie de chaque multiplieur a été multiplié par 4 (2^2), à l'image de la multiplication par 5 du courant différentiel ΔI_{in} de chaque multiplieur analogique, afin d'accroître la dynamique des signaux. Le signal issu de la somme des signaux de sortie des multiplieurs Mult11 et Mult12 et de l'additionneur Add1 a donc une dynamique de 2.5V. Compte tenu des valeurs d'entrée V_{inNN_1} et V_{inNN_2} et des paramètres w_{11} , w_{12} et b_1 dans des cas pratiques, la probabilité que la dynamique du signal d'entrée de la fonction tangente hyperbolique dépasse 1V est faible.

La méthode choisie pour synthétiser la fonction d'activation sigmoïdale est celle du développement limité, mettant en œuvre une série de multiplications et d'addition :

$$a_3 = a_2 - \frac{a_2^3}{3} + \frac{2a_2^5}{15} - \frac{17a_2^7}{315} + o(a_2^8) \quad (V-7)$$

Comme les termes d'ordre supérieur à 3 ne deviennent prépondérant qu'à partir de 1V, pour optimiser la consommation et le nombre de blocs logiques utilisés, le développement a été

limité à l'ordre 3. La figure V-k représente l'arrangement des cellules multiplieurs et additionneur sous forme de schéma bloc et en vue schématique.

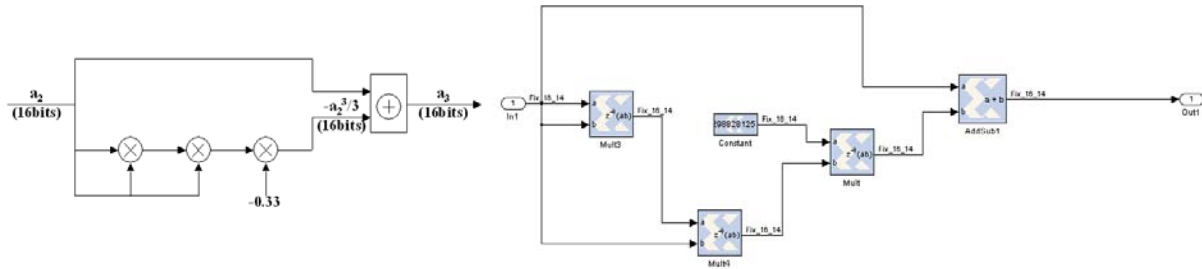


Figure V-k : Approximation de la tangente hyperbolique à l'ordre 3.

$$a_3 = a_2 - \frac{a_2^3}{3}$$

La figure V-l représente le neurone élémentaire. Après synthèse, le nombre de *slices* est égal à 104, impliquant 204 LUTs. Le temps de propagation est évalué à 35ns.

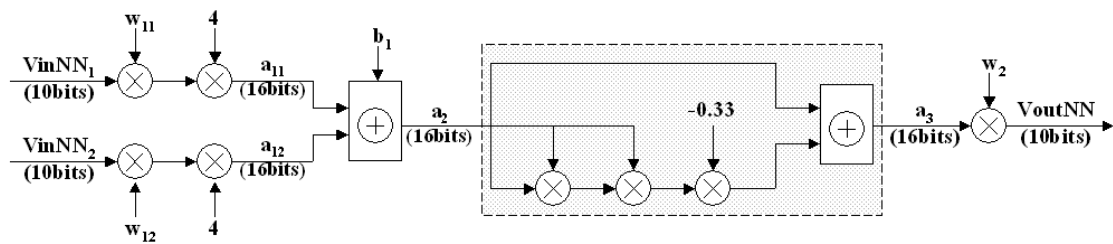


Figure V-l : Schéma d'implantation numérique du neurone.

$$a_4 = w_2 \left(a_2 - \frac{a_2^3}{3} \right) \text{ avec } a_2 = 4w_{11} \text{VinNN}_1 + 4w_{12} \text{VinNN}_2 + b_1$$

Pour une dynamique de 1V sur chacune des entrées, la tension de sortie est évaluée avec une précision de 977 μ V. Compte tenu du nombre de *slices* et de broches impliquées, le type de FPGA choisi est un Xilinx Spartan 3 (XC3S200 ; FPGA à 200000 portes équivalentes, ayant 12 multiplieurs câblés).

- Les caractéristiques de transfert du multiplieur (figure V-m) et de la fonction tangente hyperbolique (figure V-n), issues de simulation post-routage, sont très proches des caractéristiques idéales attendues. L'écart relatif constaté entre les caractéristiques idéale et simulée de la fonction tangente hyperbolique à partir de 0.3V crête, est lié à la limitation du développement limité à l'ordre 3. Une approximation à un ordre supérieur permettrait de se rapprocher de la caractéristique idéale, mais pour une consommation et un nombre de blocs logiques mobilisés plus importants.

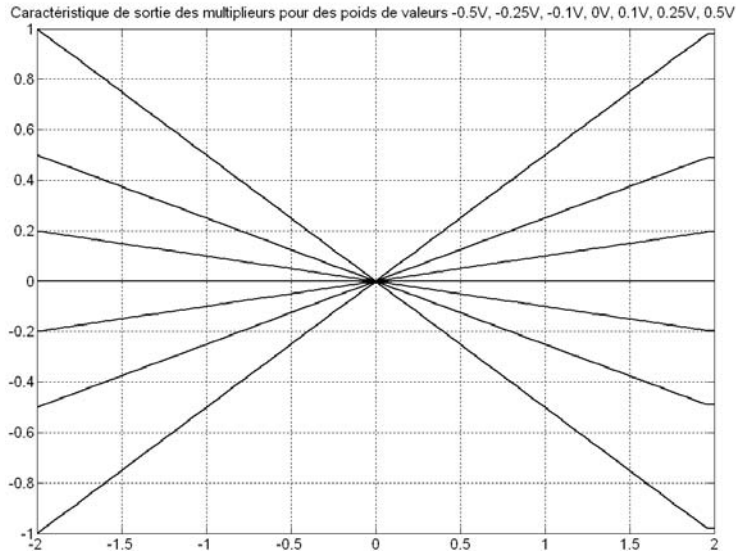


Figure V-m : Caractéristique du multiplieurMult11 en fonction du signal V_{inNN_1} pour différentes valeurs de poids w_{11} .

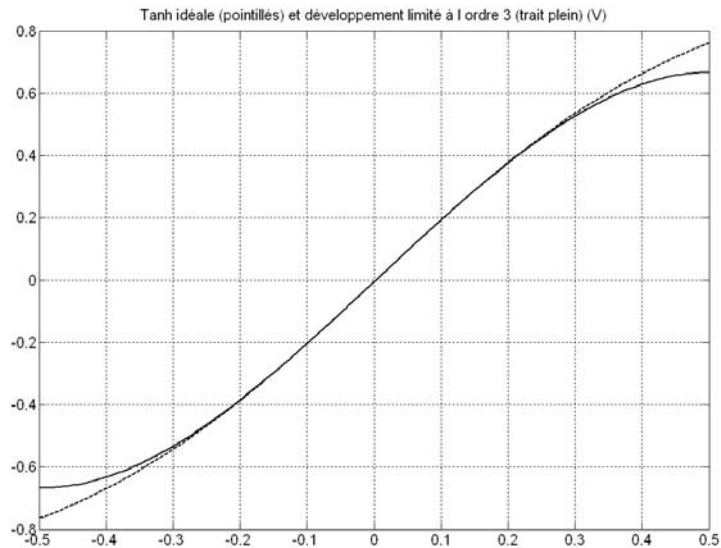


Figure V-n : Caractéristique de la fonction tangente hyperbolique.

La figure V-o représente les puissances consommées par la fonction tangente hyperbolique et par le neurone complet. On note que celles-ci dépendent de la fréquence. De plus, même à basse fréquence, le composant dissipe une puissance statique de 37.2mW, liée à la mémoire statique SRAM de configuration du FPGA choisi.

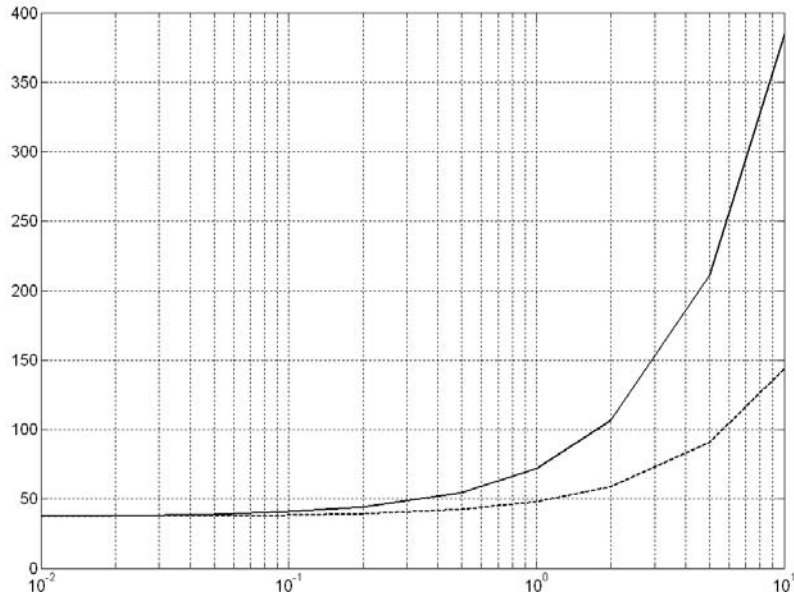


Figure V-o : Puissances consommées en mW par la fonction tangente hyperbolique (pointillés) et par le neurone (trait plein) en fonction de la fréquence (en MHz).

- Une fois les simulations post-routages validées, le neurone a été synthétisé dans le FPGA. Les tensions d'entrée V_{inNN_1} et V_{inNN_2} analogiques sont converties en un signal numérique 10 bits. Le signal de sortie a_4 numérique (10 bits) est converti en analogique de manière à pouvoir l'observer plus facilement (Figure V-p). Ce même convertisseur permet aussi de visualiser chaque signal intermédiaire a_1 , a_2 ou a_3 , afin de valider fonctionnellement chaque cellule élémentaire. La photographie ci-dessous montre le PCB de test du neurone numérique réalisé.

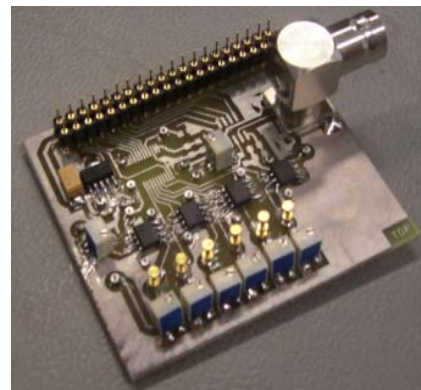
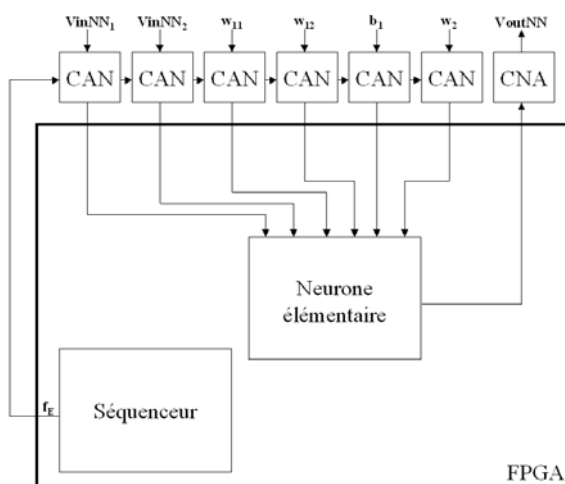


Figure V-p : Photo du PCB de test du neurone numérique.

V.3.2. Avantages liés à chaque méthode d'intégration

Effectuons maintenant un bilan comparatif entre les deux méthodes de conception analogiques et numériques du neurone élémentaire du réseau de neurones amené à être intégré dans la tête optique du télémètre. Ce bilan est effectué à partir d'extrapolation de valeurs issues des simulations numériques.

- En terme de complexité de la phase de conception, l'approche numérique est nettement plus avantageuse, sachant que les cellules numériques constituant le réseau de neurones de type PMC sont relativement simples à mettre en œuvre ou extractibles de bibliothèques existantes.

- En terme de temps de développement, la conception numérique est donc beaucoup plus rapide. Le temps estimé pour passer de la phase de simulation post-routage au test du circuit sur PCB est évalué à 1 mois environ pour le neurone numérique contre 6 mois pour le neurone analogique.

- La précision du système numérique dépend du codage choisi, de la résolution en terme de nombre de bits et de la complexité de la caractéristique de transfert à réaliser. Si l'on prend l'exemple de la cellule tangente hyperbolique, plus l'ordre du développement limité de la fonction sera élevé, plus la précision sera grande, mais plus le nombre de cellules impliquées et la consommation associée seront importants. En revanche, pour un nombre de bits d'entrée égal à 10, la précision associée à un multiplieur ou à un additionneur numérique 16 bits est bien inférieure au mV (valeur moyenne du bruit constaté lors du test du neurone analogique).

Si un type d'intégration est à privilégier en terme de précision, celui-ci dépendra de la fréquence d'utilisation, des contraintes imposées en terme de consommation, et de la nécessité ou non de mettre à jour les paramètres au sein de l'application.

- En terme de consommation, l'intégration analogique du réseau non bouclé est nettement plus intéressante. En effet, elle ne dépend pas de la puissance dissipée liée à l'utilisation d'une mémoire statique ou à la fréquence de fonctionnement. Ce critère est particulièrement important dans le cadre d'une intégration au sein d'un système embarqué.

Ce propos relatif à la consommation est à nuancer en cas d'intégration numérique de l'algorithme de mise à jour des paramètres du réseau, en fonction du nombre de convertisseurs utilisés.

- En terme de retard ou de temps de traitement des signaux d'entrée (tels que les conversions analogiques-numériques des grandeurs), la conception analogique est avantageuse. Bien que les technologies numériques présentent des longueurs de canaux plus faibles, les retards introduits par la sérialisation des cellules en terme de temps de propagation

limitent la bande passante d'utilisation du circuit global, et par là même, la fréquence d'échantillonnage. Une structure "pipeline" permettrait d'optimiser la bande passante.

- En terme de coût, l'implantation numérique d'un réseau relativement simple est nettement plus avantageuse. Le coût de développement limitant d'un ASIC analogique est lié à la technologie et à la surface de silicium utilisée. La rentabilité d'une intégration analogique n'intervient que dans le cadre d'une production importante du circuit implanté.

- Enfin, en termes de calibrage, de facilité de mise à jour et de reconfigurabilité de la structure, le neurone numérique est nettement plus avantageux.

Ainsi, le choix du type d'implantation d'un réseau de neurones tient essentiellement dans les caractéristiques que l'on souhaite mettre en avant au sein de l'application choisie. Dans le cadre de l'intégration de réseaux de neurones non bouclés au sein d'applications téléométriques, les avantages d'une implantation analogique sont une faible consommation, un faible encombrement, un traitement rapide des signaux d'entrée et la possibilité de n'effectuer aucune conversion analogique numérique.

V.4. PERFORMANCES - LIMITATIONS

Le principal avantage du système de levée d'indétermination sur la mesure de distance est l'augmentation de l'intervalle de distance d'un facteur 3, sans modifier la structure initiale du télémètre permettant d'effectuer des mesures de distance de précision.

Le système de discrimination de surfaces permet de réaliser de la classification de surfaces sur des plages de distance et d'angle d'incidence assez importantes. L'utilisateur a aussi le choix de privilégier la détection d'un grand nombre de surfaces sur des intervalles restreints.

Les performances liées à l'utilisation d'un réseau de neurones analogique non bouclé sont :

- une faible consommation ajoutée,
- une capacité d'intégration dans la tête optique du télémètre, de manière à permettre la levée d'indétermination directement au sein du système embarqué,
- un fonctionnement "temps réel" après entraînement (calibrage) du réseau,
- une capacité d'approximation du réseau à partir de mesures expérimentales, tenant compte du bruit et des défauts dus à la défocalisation ou à la diaphonie,
- une mise à jour des paramètres du réseau de neurones relativement simple d'un point de vue software après acquisition des valeurs d'entraînement, laissant présager une capacité de mise à jour assez rapide des paramètres par un système numérique intégré.

Les principales limitations de ces deux systèmes sont liées à la difficulté de "re-calibrage" du réseau de neurones si les conditions expérimentales changent (texture de la cible, angle d'incidence, température, lumière ambiante ou semi-obscurité, etc.).

Pour pallier ce problème, il est possible de sélectionner d'autres entrées pertinentes au sein du système, telles que l'angle d'incidence (par un système de mesure dont on peut extraire une valeur de tension continue proportionnelle à l'angle), la température (par un capteur de température), l'éclairement (par un capteur de mesure d'éclairement type luxmètre), etc. La phase d'entraînement devra tenir compte de la variation de l'ensemble des entrées choisies en fonction des conditions expérimentales. La finesse de l'approximation et la complexité de la structure du réseau de neurones associé seront fonctions du nombre de paramètres d'entrée.

Une autre possibilité consiste à entraîner le réseau à intervalles de temps réguliers sans changer sa structure actuelle. Ainsi, par un système de mise à jour des paramètres, le réseau sera calibré à partir d'un échantillon de couples {entrées; sortie}, les entrées étant déterminées à partir de mesures expérimentales. Par cette méthode, les conditions de test n'ont pas besoin d'être "quantifiées". En revanche, le système et la procédure d'acquisition des échantillons et de mise à jour des paramètres sont à mettre en œuvre.

V.5. REFERENCES

[Gatet, 2007(a)] : GATET L., TAP-BETEILLE H., LESCURE M., Measurement Range Increase of a Phase-Shift Laser Rangefinder using a CMOS Analog Neural Network, *IEEE Transactions on Instrumentation and Measurement* (première lecture).

[Gatet, 2007(b)] : GATET L., TAP-BETEILLE H., LESCURE M., Real Time Surface Discrimination using an Analog Neural Network implemented in a Phase-Shift Laser Rangefinder", *IEEE Sensors Journal* (en cours de publication).

[Levenberg, 1944] : LEVENBERG K., A method for the solution of certain non-linear problems in least squares, *Quarterly Journal of Applied Mathematics*, 2, p. 164-168.

[Marquardt, 1963] : MARQUARDT D.W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society of Industrial and Applied Mathematics*, 11, p. 431-441.

[Rumelhart, 1986] : RUMELHART D.E., HINTON G.E., WILLIAMS R.J., Learning internal representations by error backpropagation, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, p.318-362, MIT Press.

[Werbos, 1974] : WERBOS P.J., *Beyond regression : new tools for prediction and analysis in the behavioural sciences*, Ph. D. thesis, Harvard University.

Conclusion générale

Afin de valider l'intérêt d'utiliser des réseaux de neurones au sein de systèmes de mesure optoélectroniques, un travail de recherche approfondi a été effectué de manière à déterminer l'ensemble des propriétés des réseaux de neurones et les diverses problématiques associées à la télémétrie laser. Les aptitudes des réseaux de neurones, en particulier en terme d'approximation de fonctions et de classification à partir d'un nombre limité de données, ont été étudiées pour pouvoir être mises en exergue au sein d'une application télémétrique. Une veille technologique de l'ensemble des structures neuronales intégrées a été effectuée de façon à garder à l'esprit la possibilité d'une intégration au sein du système si l'étude comportementale s'avérait fructueuse.

La principale problématique associée à la télémétrie laser est la mesure de distance sur une plage importante avec une résolution élevée. Par ailleurs, la possibilité d'effectuer de la classification de surfaces à partir de ce même système offre une ouverture intéressante. Les domaines d'applications sont nombreux : robotique terrestre ou spatiale (évolution de systèmes robotisés dans un environnement inconnu), industrie (réalisation de tâches de précision et de contrôle d'homogénéité), surveillance (détection d'intrus), transport (cartographie de l'environnement), médical (imagerie de précision), etc.

Après avoir étudié les différentes techniques télémétriques de mesure de distance par faisceau laser, deux études comportementales ont été réalisées afin d'évaluer l'aptitude d'un réseau de neurones de type perceptron multi-couches à répondre à deux applications à partir de données expérimentales. La première consiste à lever l'indétermination sur la mesure de distance issue du télémètre laser par déphasage par l'intermédiaire d'un réseau de neurones. La seconde concerne l'aptitude de ce même réseau à discriminer plusieurs types de cibles de coefficient de réflexion diffusant apparent différent sous différents angles d'incidence et à une distance variable.

A l'issue de ces deux études, il a été démontré qu'un réseau de neurones est capable d'effectuer une approximation de la valeur de la distance avec une précision suffisante pour

lever l'indétermination modulo 2π apportée par le télémètre laser par déphasage, sur un intervalle de distance trois fois supérieur à celui limité par la mesure du déphasage. De plus, une architecture neuronale quasiment similaire permet d'effectuer la classification de quatre types de surfaces (une surface plastifiée, du papier glacé, un mur peint et une surface poreuse) sur une plage de distance de l'ordre du mètre et pour un angle d'incidence variable. Comme les deux structures neuronales sont relativement proches l'une de l'autre (au nombre d'entrées près), un même réseau de neurones peut être utilisé pour répondre aux contraintes des deux applications.

Ayant évalué l'intérêt majeur d'une intégration de ce réseau au sein du système télémétrique, il a été décidé de procéder à l'implantation de la structure neuronale de deux manières différentes, afin de comparer les performances de chaque type d'intégration. Ainsi, un réseau de neurones analogique à deux entrées, une couche cachée contenant trois neurones et un neurone de sortie a été réalisé à partir de neurones développés en technologie CMOS AMS 0.35 μ m basse tension. Une étude poussée des cellules élémentaires a été menée de manière à optimiser leur structure en terme de dynamique, de consommation et de vitesse de transfert. Une étude quasi-statique, petit signal et fort signal a permis de valider l'aspect paramétrable du neurone intégré dans l'ASIC.

A l'issue de cette étude, une équation de transfert du premier ordre par rapport aux deux entrées et à chaque paramètre a été extraite, de manière à pouvoir entraîner le réseau à l'aide d'un logiciel annexe. La capacité d'approximation du réseau relative à l'accroissement de l'intervalle de mesure de distance a donc été vérifiée expérimentalement, tout comme l'aptitude du réseau intégré à classer plusieurs types de surfaces pour une distance et un angle d'incidence variable.

La réalisation de la même structure neuronale par l'intermédiaire de cellules numériques a permis de justifier le choix de la conception analogique en termes de consommation, de précision et de vitesse de traitement des signaux pour une utilisation non bouclée du réseau.

Pour pallier les principales limitations de ces deux systèmes, liées à la difficulté de "re-calibrage" du réseau de neurones en fonction des conditions expérimentales, une amélioration notable pourrait être apportée par l'ajout d'une troisième entrée au réseau correspondant à la détermination de la valeur de l'angle d'incidence. Cette mesure peut être effectuée par

l'intermédiaire d'une barrette ou d'une matrice de photodiodes à avalanches (APD) à la place de la seule photodiode utilisée actuellement. La connaissance de l'angle d'incidence permettrait de réaliser la levée d'indétermination sur la mesure de distance pour un angle d'incidence variable et d'accroître considérablement les plages de mesure relatives à la classification de surfaces. De plus, cette barrette ou cette matrice permettrait de s'affranchir de possibles imperfections des surfaces testées (goutte d'eau, saleté, non homogénéité, fissure, etc.) en moyennant l'ensemble des signaux issus de chaque APD.

Un système de mise à jour des poids et biais à partir d'échantillons de couples {entrées; sortie} prélevés à intervalles de temps réguliers permettrait de s'affranchir des autres variations liées aux conditions de test du système. Cette méthode évite la "quantification" des diverses variables relatives aux conditions d'expérimentations, et ainsi une complexité de traitement du réseau liée à un nombre d'entrées trop important. Pour faciliter la mise à jour de ces paramètres, l'intervention d'un système numérique est préférable, nécessitant de ce fait des conversions analogiques - numériques et numériques - analogiques pour permettre le dialogue avec le réseau de neurones. La structure du réseau "non bouclé" peut alors être optimisée, en faisant appel à des multiplieurs et des additionneurs analogiques, dont les transferts en courant relatifs aux valeurs des poids et biais sont déterminés par l'intermédiaire d'un système d'interrupteurs (*switches*) contrôlé par une commande numérique.

Un nombre de neurones cachés plus important permettrait d'accroître les capacités de discrimination du système global. Un interrupteur placé à chaque sortie de neurone caché rendrait la structure du réseau plus modulable et adaptable à toute application portable nécessitant l'approximation d'une fonction ou la classification de données. Enfin, la structure analogique relative à la fonction d'activation sigmoïdale peut être améliorée en termes de contrôle de la zone de linéarité et de bande passante en utilisant des structures cascodées, des amplificateurs rail-to-rail, etc.

Annexe A

Modélisation statique

La propriété d'approximation parcimonieuse des réseaux de neurones est utile pour résoudre le problème de modélisation "boîte noire" ou "boîte grise". Pour le montrer, considérons d'abord le cas simple d'un modèle à une seule variable x_1 et à une seule sortie. On cherche à établir ainsi un profil très précis de la fonction relative au phénomène étudié, associant la mesure $z_1(x_1)$ au paramètre x_1 qui la fait varier.

Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, par un réseau de neurones. Il est donc possible d'utiliser un réseau de neurones pour effectuer une régression de la fonction à approximer à partir du moment où cette fonction est suffisamment régulière. La modélisation "boîte noire" du phénomène associé consiste donc à estimer les paramètres du réseau de neurones de manière à approximer la fonction avec la précision souhaitée ([Dreyfus, 2004]). Cette approximation est généralement faite au sens des moindres carrés.

Ainsi, si on appelle g la fonction réalisée par le réseau de neurones, relative à la composée des fonctions associées à chacun des neurones, et W_0 l'ensemble des paramètres mis en jeu au sein du réseau de neurones, on estime la fonction de coût des moindres carrés J . Si l'on considère toujours que l'on ne cherche à approximer que la variation d'une donnée du système qui n'est fonction que d'un paramètre d'entrée x_1 du système, la fonction de coût des moindres carrés J se calcule pour un nombre N de valeurs d'entrée x_1 réparties sur l'espace des entrées :

$$J = \frac{1}{2} \sum_{\alpha=1}^N [z_1(x_{1\alpha}) - g(x_{1\alpha}, W_0)]^2 \quad (\text{A-1})$$

avec $x_{1\alpha}$ ($\alpha \in [1 ; N]$) la $\alpha^{\text{ième}}$ valeur d'entrée du système et $z_1(x_{1\alpha})$ la donnée correspondante (figure A-a).

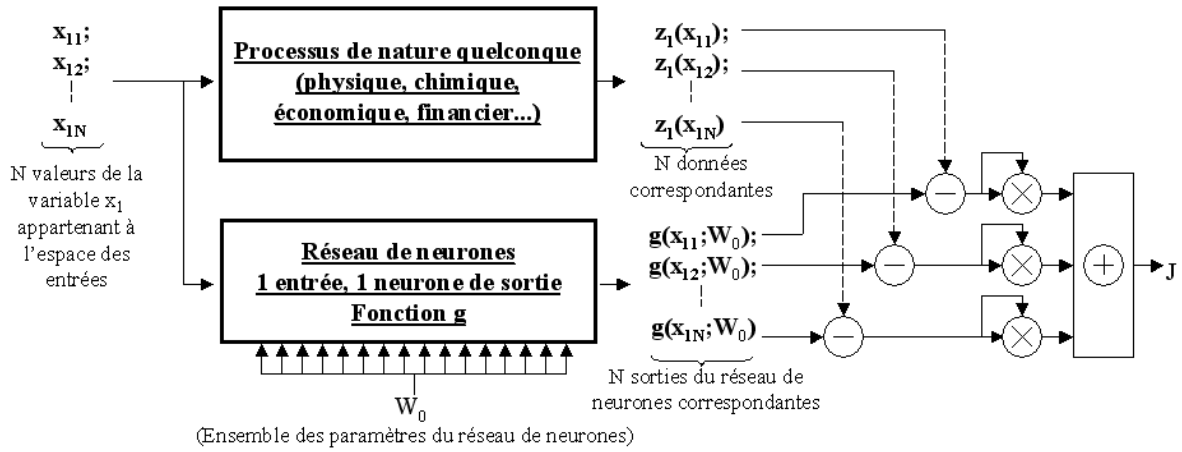


Figure A-a : Calcul de la fonction de coût des moindres carrés pour un système d'une donnée ne dépendant que d'une variable.

Dans le cas général, lorsque l'on souhaite approximer la variation de N_m données d'un système fonction de n variables d'entrée x_i ($i \in [1 ; n]$) du système, la fonction de coût des moindres carrés J se calcule pour un nombre N de jeux de valeurs d'entrée $\{x_{i\alpha}, i \in [1 ; n], \alpha \in [1 ; N]\}$ réparties sur l'espace des entrées :

$$J = \frac{1}{2} \sum_{\alpha=1}^N \left(\sum_{m=1}^{N_m} [z_m(x_{1\alpha}, x_{2\alpha}, \dots, x_{n\alpha}) - g_m(x_{1\alpha}, x_{2\alpha}, \dots, x_{n\alpha}; W_0)]^2 \right) \quad (A-2)$$

où $z_m(x_{1\alpha}, x_{2\alpha}, \dots, x_{n\alpha})$ est la $m^{\text{ième}}$ donnée que la $m^{\text{ième}}$ fonction de sortie g_m du réseau de neurones doit modéliser à partir du $\alpha^{\text{ième}}$ vecteur d'entrée $\{x_{i\alpha}, i \in [1 ; n]\}$ et de son jeu de paramètres W_0 (figure A-b).

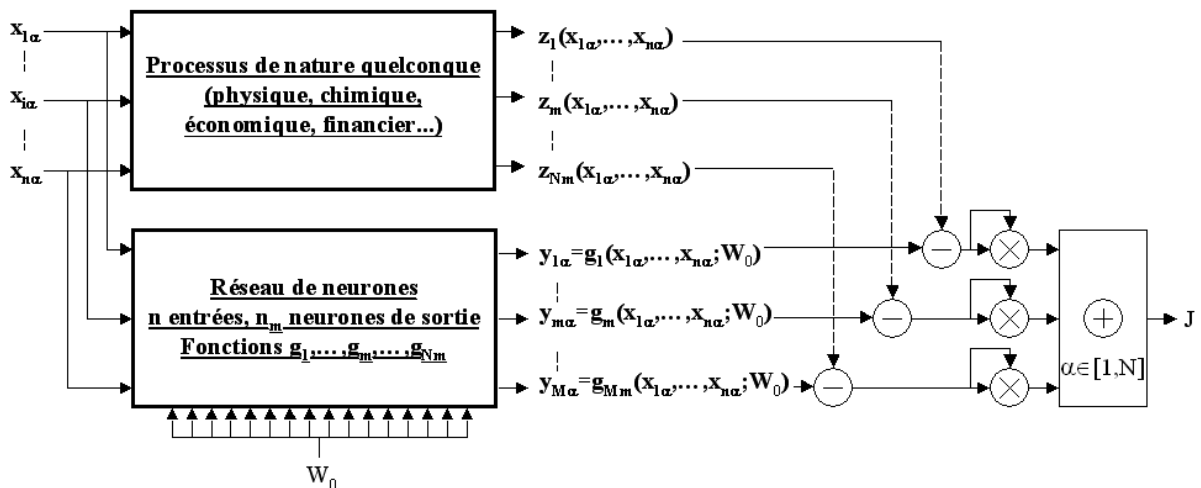


Figure A-b : Calcul de la fonction de coût des moindres carrés pour un vecteur d'entrée indicé α (avec $\alpha \in [1; N]$) dépendant de n variables dans un système où l'on veut modéliser la variation de N_m données.

Lors de la phase d'apprentissage d'un réseau, le nombre de neurones cachés est fixé dans un premier temps. L'approximation des fonctions de variation des données g_m ($m \in [1 ; N_m]$)

consiste donc en une procédure de minimisation de la fonction de coût des moindres carrés J , soit, en d'autres termes, à trouver un ensemble satisfaisant des paramètres W_0 mis en jeu au sein du réseau de neurones.

Voici un exemple concret de régression d'une fonction d'une donnée fonction d'une seule variable. La figure A-c représente un signal $z_1(x_1)$ que l'on voudrait modéliser, c'est-à-dire un signal dont on voudrait obtenir la meilleure approximation possible en fonction de l'entrée x_1 .

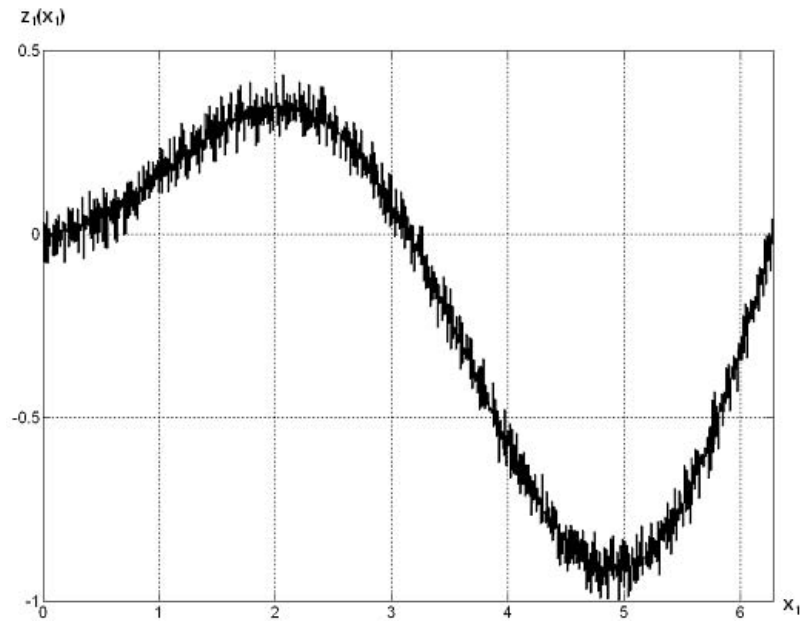


Figure A-c : Un signal que l'on voudrait modéliser.

N'ayant accès qu'à un nombre fini de valeurs, ou bien étant limités par le temps relatif à chaque prise de mesures du fait de la complexité de l'expérience, on n'effectue qu'un nombre fini N de mesures $z_1(x_1)$, comme illustré sur la figure A-d.

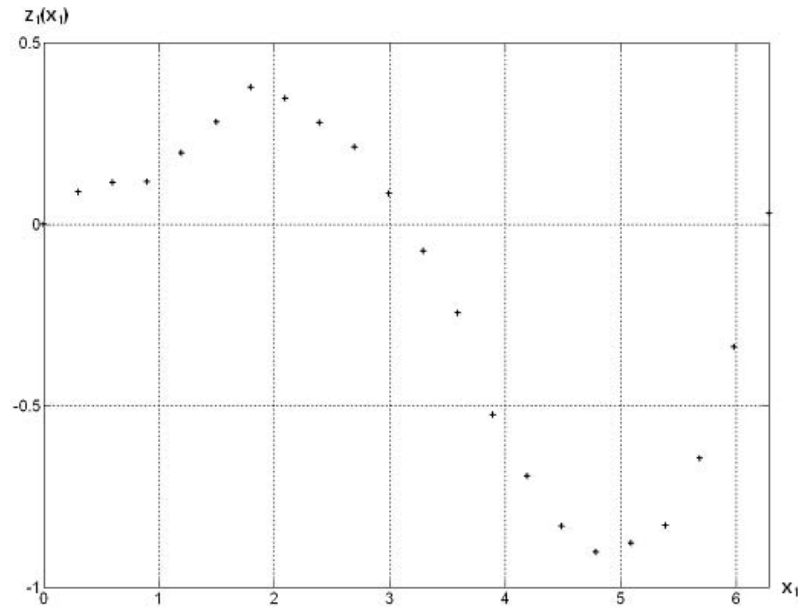


Figure A-d : Nombre fini N de mesures $z_1(x_1)$. Dans cet exemple, les N points de mesure sont répartis uniformément sur l'intervalle de mesure.

A partir de ces exemples, pour une structure de réseau de neurones et pour un nombre de paramètres donnés, l'objectif est de minimiser la fonction de coût des moindres carrés J en jouant sur l'ensemble W_0 des paramètres mis en jeu au sein du réseau de neurones (figure A-a). La figure A-e donne un exemple d'approximation de la fonction de régression $g(x_1, W_0)$ réalisée par le réseau de neurones.

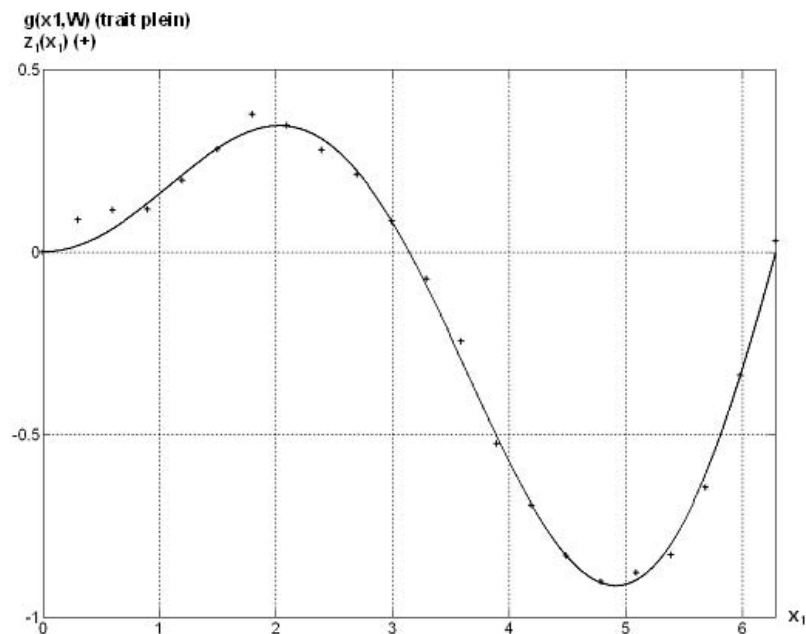


Figure A-e : Approximation de la fonction de régression $g(x_1, W_0)$, réalisée par un réseau de neurones, à partir des points expérimentaux $z_1(x_1)$ de la figure A-d.

Annexe B

Calibrage d'un réseau de neurones

Une fois le type de réseau de neurones, le nombre d'entrées et le nombre de neurones de sortie choisis, il reste à déterminer le nombre de couches intermédiaires dites "cachées" et le nombre de neurones par couche cachée.

Il n'existe pas, à l'heure actuelle, de résultat théorique permettant de prévoir le nombre de neurones cachés nécessaires pour obtenir une performance spécifiée du modèle, compte tenu des données disponibles. Certes, ces neurones sont apparentés à la partie "processeur" du réseau, donc leur nombre et leur arrangement en couches vont être responsables de la qualité de l'approximation effectuée. Il est évident qu'un nombre insuffisant de neurones cachés aura pour conséquence une approximation de la ou des fonctions de mauvaise qualité.

Un raisonnement basique conduirait au fait que plus le nombre de couches cachées et de neurones par couche cachée est important, meilleure sera l'approximation. De plus, comme nous l'avons vu précédemment, pour un réseau de neurones non bouclé, le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone d'un réseau non bouclé est négligeable. Chaque couche étant constituée de neurones en parallèle, le nombre de neurones par couche n'influencerait donc pas le temps de transit de l'information des entrées vers les sorties. Il serait donc tout indiqué à première vue de placer autant de couches cachées et de neurones par couches cachées que possible.

Malheureusement, un réseau de neurones n'effectue pas une approximation uniforme d'une ou plusieurs fonctions mais un ajustement d'une fonction à un nombre fini de points par composition de plusieurs fonctions élémentaires paramétrables. Il faut donc, non seulement que la fonction réalisée par le réseau de neurones passe "le plus près possible", au sens des moindres carrés, des points utilisés pour l'apprentissage (l'ensemble de ces points est appelé "ensemble d'apprentissage"), mais également qu'il soit capable d'approcher de manière satisfaisante d'autres points n'appartenant pas à l'ensemble d'apprentissage. Un nombre excessif de neurones et de paramètres associés sur une couche cachée entraîne une très grande précision au niveau de l'approximation de la ou des fonctions pour chaque valeur appartenant

à l'ensemble d'apprentissage (apprentissage "par cœur" de la valeur de sortie associée aux valeurs d'entrées) mais fournit des valeurs dépourvues de signification entre ces points. Ce phénomène, appelé "surajustement" (*overfitting*), est représenté sur la figure B-a dans le cadre d'une approximation de fonction à une variable.

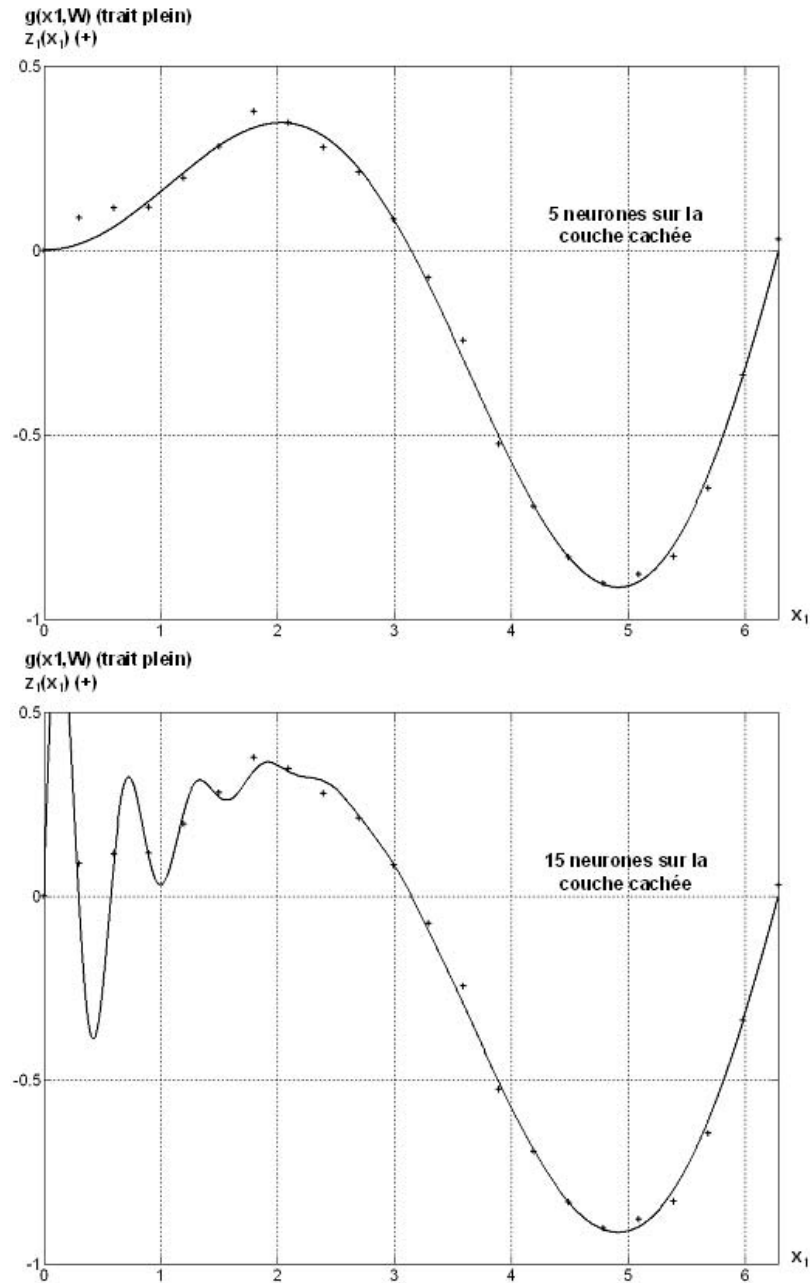


Figure B-a : Mise en évidence du phénomène de surajustement.

La structure finale du réseau de neurones va donc dépendre en particulier de l'arrangement et du nombre de neurones sur sa ou ses couches cachées, et des paramètres associés. Ainsi, structure et calibrage sont deux phases corrélées et indissociables ([Dreyfus, 2004]).

Il existe plusieurs méthodes pour déterminer les données les plus pertinentes parmi un jeu de valeurs (par exemple [Mc Quarrie, 1998]) pour constituer l'ensemble d'apprentissage. Néanmoins, une connaissance a priori de la fonction à approcher (modèle "boîte grise") aide grandement au choix des valeurs d'entrée constituant cet ensemble d'apprentissage.

Pour quantifier le caractère satisfaisant ou non du degré d'approximation de la fonction réalisée par le réseau de neurones, une des méthodes utilisées consiste à calculer l'erreur quadratique moyenne commise pour chacun des exemples appartenant à l'ensemble d'apprentissage ou de validation ([Dreyfus, 2004]). Ainsi, si l'on se place dans le cas où l'on souhaite approcher une fonction à partir de N données $z_1(x_{1\alpha})$ ($\alpha \in [1 ; N]$) relatives à la variation d'un seul paramètre $x_{1\alpha}$, et si on distingue, parmi ces N valeurs du paramètre $x_{1\alpha}$, N_A exemples $x_{1\alpha A}$ dans l'ensemble d'apprentissage et N_V exemples $x_{1\alpha V}$ dans l'ensemble de validation, les erreurs quadratiques moyennes relatives à chacun des deux ensembles sont respectivement :

$$e_A = \sqrt{\frac{1}{N_A} \sum_{\alpha A=1}^{N_A} [z_1(x_{1\alpha A}) - g(x_{1\alpha A}, W_0)]^2} \quad (\text{B-1})$$

$$e_V = \sqrt{\frac{1}{N_V} \sum_{\alpha V=1}^{N_V} [z_1(x_{1\alpha V}) - g(x_{1\alpha V}, W_0)]^2} \quad (\text{B-2})$$

avec g la fonction réalisée par le réseau de neurones à une entrée et un neurone de sortie et W_0 l'ensemble de ses paramètres.

Dans le cas général, si le réseau doit approcher la variation de N_m fonctions à partir de N vecteurs de données relatives à la variation d'un vecteur de n paramètres $\{x_{i\alpha}, i \in [1 ; n], \alpha \in [1 ; N]\}$, les erreurs quadratiques moyennes sont la somme des écarts quadratiques moyens relatifs à chacune des sorties.

$$e_A = \sum_{m=1}^{N_m} \sqrt{\frac{1}{N_A} \sum_{\alpha A=1}^{N_A} [z_m(x_{1\alpha A}, \dots, x_{n\alpha A}) - g_m(x_{1\alpha A}, \dots, x_{n\alpha A}, W_0)]^2} \quad (\text{B-3})$$

$$e_V = \sum_{m=1}^{N_m} \sqrt{\frac{1}{N_V} \sum_{\alpha V=1}^{N_V} [z_m(x_{1\alpha V}, \dots, x_{n\alpha V}) - g_m(x_{1\alpha V}, \dots, x_{n\alpha V}, W_0)]^2} \quad (\text{B-4})$$

avec g_m la $m^{\text{ième}}$ fonction de sortie réalisée par le réseau de neurones, fonction du $\alpha^{\text{ième}}$ vecteur d'entrée $\{x_{i\alpha}, i \in [1 ; n], \alpha \in [1 ; N]\}$ et du jeu de paramètres associé W_0 , de manière à approcher au mieux la variation des données $z_m(x_{1\alpha}, x_{2\alpha}, \dots, x_{n\alpha})$ (figure A-b).

Sur la figure B-a, l'erreur quadratique moyenne e_A relative à l'ensemble d'apprentissage du réseau possédant 15 neurones sur sa couche cachée est inférieure à celle du réseau possédant 5

neurones cachés. Ainsi, si l'on ne s'appuyait que sur la valeur de e_A , pour estimer le degré de satisfaction d'une approximation, le réseau possédant 15 neurones cachés et qui présente un phénomène de surajustement serait le meilleur approximateur. L'erreur quadratique moyenne e_A relative à l'ensemble d'apprentissage seule n'est pas un bon critère de la qualité d'approximation.

Maintenant, si l'on prend pour ensemble de validation les valeurs des entrées uniformément réparties entre les valeurs de l'ensemble d'apprentissage, l'erreur quadratique moyenne e_V du réseau à 15 neurones cachés est nettement supérieure à celle du réseau de neurones à 5 neurones cachés. Ainsi, à l'aide des deux erreurs quadratiques moyennes e_A et e_V , on peut déterminer que le réseau à 5 neurones cachés approche "le plus parcimonieusement" la fonction désirée. Il est donc nécessaire de déterminer deux jeux de valeurs à partir de l'ensemble des exemples disponibles et de bien choisir le jeu d'apprentissage de manière à valider la qualité d'approximation du réseau de neurones.

Ainsi, l'objectif de la modélisation peut se résumer à trouver l'approximation la plus parcimonieuse possible, de telle sorte que les erreurs quadratiques moyennes sur les ensembles d'apprentissage et de validation soient du même ordre de grandeur, et aussi petites que possible, c'est-à-dire inférieures à une valeur déterminée par le concepteur.

L'ensemble des paramètres W_0 permettant au réseau de neurones d'approcher la fonction de régression désirée est déterminée de manière itérative lors de la phase d'apprentissage. Pour une structure bien définie, on procède à une mise à jour de ces paramètres à chaque itération de manière à obtenir une erreur quadratique minimum pour l'ensemble des valeurs du jeu d'apprentissage. L'apprentissage est donc un problème d'optimisation. Cette phase consiste en d'autres termes à trouver l'ensemble des paramètres W_0 qui rendent la valeur de la fonction de coût au sens des moindres carrés J minimale sur l'ensemble d'apprentissage, ou, ce qui revient au même, qui minimisent la valeur de l'erreur quadratique moyenne e_A relative à l'ensemble d'apprentissage.

Dans le cas de la mise en oeuvre de réseaux de neurones dont la sortie n'est pas linéaire par rapport aux paramètres, tels que des perceptrons multi-couches ou les RBF à centres et écarts-types variables, les techniques utilisées consistent à modifier les paramètres du réseau, dont les valeurs de départ sont aléatoires, jusqu'à ce qu'un minimum de la fonction de coût soit atteint ou qu'un critère d'arrêt soit satisfait. Ces techniques sont toutes des méthodes de gradient : elles sont fondées sur le calcul, à chaque itération, du gradient de la fonction de coût par rapport aux paramètres, gradient qui est ensuite utilisé pour calculer une modification de

ceux-ci. Le calcul du gradient peut être effectué soit dans le sens direct, soit par "rétropropagation de l'erreur" ([Rumelhart, 1986], [Werbos, 1974]).

La méthode de rétropropagation de l'erreur est la plus utilisée, car elle est généralement plus économe en termes de nombres d'opérations arithmétiques à effectuer pour évaluer le gradient. Cette méthode consiste à propager le gradient de l'erreur aux différents paramètres en allant de la couche de sortie vers les couches d'entrées. Ainsi, plus on s'éloigne de la couche de sortie - ou plus on se rapproche des couches d'entrée -, plus les dérivées des paramètres par rapport à l'erreur commise sont complexes à calculer. Certains algorithmes associés à la méthode de rétropropagation ne font appel qu'à la dérivée première des paramètres, comme l'algorithme du gradient simple ([Haykin, 1994]).

D'autres utilisent aussi les dérivées secondes de manière à prendre aussi en compte le sens de variation de l'erreur, comme la méthode de Broyden-Fletcher-Goldfarb-Shanno (algorithme BFGS [Broyden, 1970]) ou l'algorithme de Levenberg-Marquardt ([Levenberg, 1944], [Marquardt, 1963]). Il est évident que plus le nombre de couches intermédiaires sera important, plus les formules impliquées pour mettre à jour les paramètres des premières couches seront compliquées. C'est pour cette raison que le nombre de couches intermédiaires est limité à deux dans la très grande majorité des cas. Concernant le nombre de neurones par couche cachée, comme les opérations impliquant les paramètres des neurones d'une même couche sont identiques, celles-ci sont effectuées en parallèle par l'intermédiaire de vecteurs. La limitation du nombre de neurones par couche provient du temps de calcul durant la phase d'apprentissage qui augmente linéairement avec le nombre de composantes du vecteur.

Annexe C

Exemple d'intégration d'un réseau de neurones analogique-numérique dans un système de télécommunication embarqué

C.a. Description succincte du système

Dans les satellites de télécommunication, il est important d'avoir un bilan en puissance optimisé et une efficacité spectrale accrue. Les amplificateurs haute puissance (*High Power Amplifier (HPA)*) tels que les tubes à ondes progressives (*Travelling Wave Tubes (TWT)*) ou les amplificateurs à semi-conducteurs (*Solid State Power Amplifiers (SSPA)*) sont utilisés près de la zone de saturation de manière à transmettre les données avec un maximum de puissance. Malheureusement, utiliser ces amplificateurs près de leur zone de saturation entraîne de fortes non-linéarités.

Plusieurs solutions existent pour limiter ces effets de distorsion ([Langlet, 2004]) : utiliser des égaliseurs non-linéaires au sein des terminaux ([Bigieri, 1984], [Bouchired, 1998], [Bouchired, 1999]), implémenter un système de linéarisation à bord du satellite ([Langlet (a), 2001], [Langlet (b), 2001], [Lazzarin, 1994]), etc. Les techniques de linéarisation peuvent être classées en deux catégories : les techniques directes (*feedforward*), où un système analogique réalise un ajustement linéaire de la caractéristique de l'amplificateur à partir du signal d'entrée, et les techniques "rétroactives" (*feedback*), où une correction est apportée en amont de l'amplificateur en fonction de la caractéristique de sortie observée. Cette correction s'effectue généralement sur le signal modulé, à la fréquence de la porteuse.

Dans les nouvelles générations de satellites à charges régénératives, les signaux modulés peuvent être démodulés de manière à les traiter directement au sein du satellite. Ces signaux sont disponibles en bande de base avant modulation. Cela signifie qu'il est possible d'effectuer la correction en amont de l'amplificateur sur des signaux bande de base avant les opérations de transposition en fréquence et d'amplification.

La technique de correction amont consiste à "prédistordre" le signal d'entrée de l'amplificateur de manière à ce que la caractéristique globale de sortie du système {prédistorsion+amplificateur} soit linéaire. En d'autres termes, la fonction réalisée par le système de prédistorsion doit être telle que la composée de cette fonction par la fonction caractéristique de l'amplificateur soit proportionnelle à l'identité. L'objectif est donc de réaliser en bande de base l'approximation de la fonction inverse de la caractéristique non-linéaire de l'amplificateur. Il existe plusieurs systèmes de prédistorsion, relatifs aux différents approximateurs de fonction connus, tels que les séries de polynômes, les séries de Volterra, les réseaux de neurones, les tables d'équivalence (*look-up tables*), etc. ([Tummla, 1997], [Langlet (a), 2001], [He, 2006]).

Néanmoins, l'approximateur doit pouvoir compenser directement à bord du satellite les lentes variations de la caractéristique de transfert de l'amplificateur en fonction de la température, du vieillissement, etc. Au delà de proposer une solution intéressante en terme de finesse d'approximation de fonctions non-linéaires et d'intégration, la prédistorsion de type neuronale offre une possibilité de mise à jour adaptative de ses paramètres.

Pour l'application présentée dans [Langlet (a), 2001], le système de prédistorsion est réalisé à l'aide d'un réseau de neurones de type PMC, de manière à compenser les fortes distorsions dues à un amplificateur SSPA. L'architecture du système de prédistorsion est présentée figure C-a.

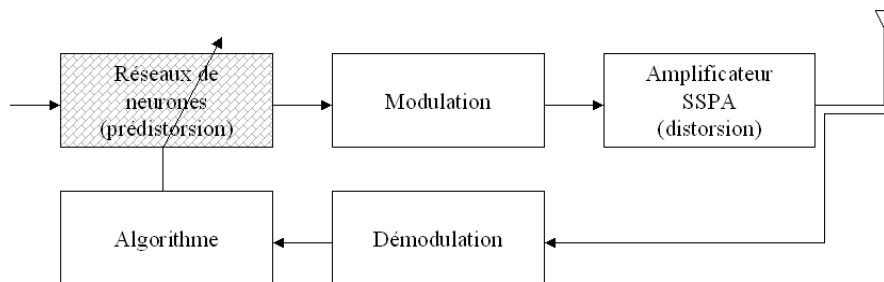


Figure C-a : Architecture du système de prédistorsion embarqué dans le satellite.

Le réseau de neurones réalise la prédistorsion du module des signaux I et Q relatifs à une modulation d'amplitude en quadrature à 16 états (16-MAQ). Le signal est ensuite modulé puis amplifié par l'amplificateur de puissance. Le signal amplifié est comparé au signal idéal non distordu par l'intermédiaire d'un algorithme de manière à ce que celui-ci remette à jour les poids et biais du réseau de neurones si une non-linéarité est constatée.

La distorsion relative à l'amplificateur de puissance peut se décomposer sous la forme de deux non-linéarités : une non-linéarité en amplitude (AM/AM) et une non-linéarité en phase (AM/PM). Le gain de l'amplificateur peut être modélisé par :

$$G(r) = A(r)\exp(j\phi(r)) \quad (C-1)$$

avec r le module du signal d'entrée, $A(r)$ le gain en amplitude et $\phi(r)$ le terme relatif à la phase. Les deux derniers termes, non linéaires, sont donnés par le modèle de Saleh [SALEH 1981] :

$$A(r) = \frac{\alpha_1}{1+\beta_1.r^2} \quad (C-2)$$

$$\phi(r) = \frac{\alpha_2.r^2}{1+\beta_2.r^2} \quad (C-3)$$

Les termes α_1 , α_2 , β_1 et β_2 sont des scalaires associés au modèle d'amplificateur choisi. Pour l'application donnée, les valeurs de ces termes sont évalués par le Centre National des Etudes Spatiales (CNES). La figure C-b représente l'amplitude de sortie normalisée et le déphasage en fonction du module du signal d'entrée.

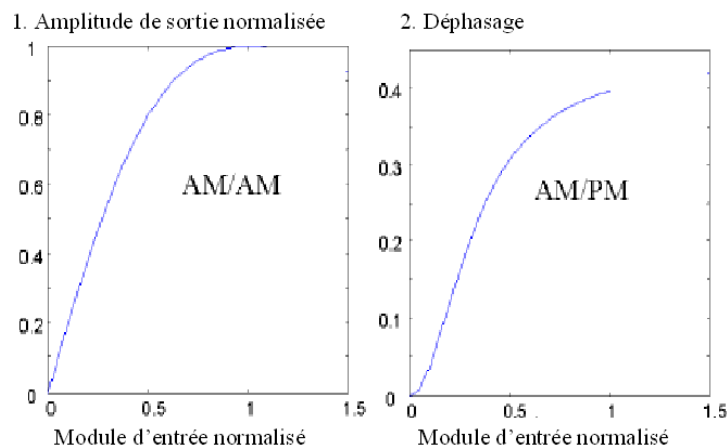


Figure C-b : Représentation des non-linéarités AM/AM normalisée (C-b.1) et AM/PM (en radians) (C-b.2) en fonction de l'amplitude du signal d'entrée normalisée.

C.b. Description du réseau de neurones choisi

Afin de compenser ces non-linéarités, un système à base de réseaux de neurones a été étudié de manière à prédistordre le signal en bande de base par l'intermédiaire de deux réseaux de neurones [Langlet (b), 2001]. La figure C-c présente le système de prédistortion développé, utilisant deux réseaux de neurones NN1 et NN2, chacun réalisant respectivement la prédistortion en amplitude AM/AM et en phase AM/PM.

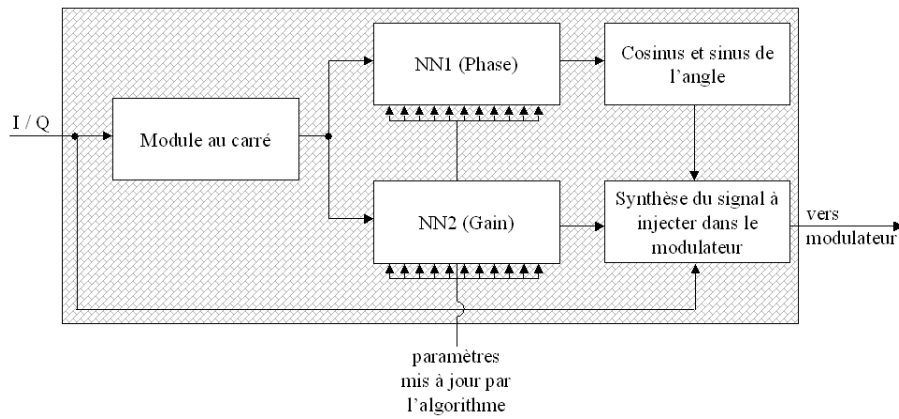


Figure C-c : Structure du système de prédistorsion.

Le type de réseau de neurones choisi est un perceptron mutli-couche (PMC), de manière à pouvoir réaliser l'approximation de la fonction de prédistorsion sur une plage de tension continue. La structure des deux réseaux de neurones utilisés est identique : un neurone d'entrée, un neurone de sortie et une seule couche cachée constituée de dix neurones en parallèle (figure C-d). Le nombre de neurones sur la couche cachée a été déterminé en incrémentant leur jusqu'à ce que l'approximation de la prédistorsion souhaitée soit suffisante. Pour entraîner le réseau de neurones, la mise à jour des poids et biais est effectuée à l'aide du logiciel Matlab par la méthode de rétropropagation de l'erreur. L'algorithme de mise à jour utilisé est celui du gradient simple [Haykin, 1994].

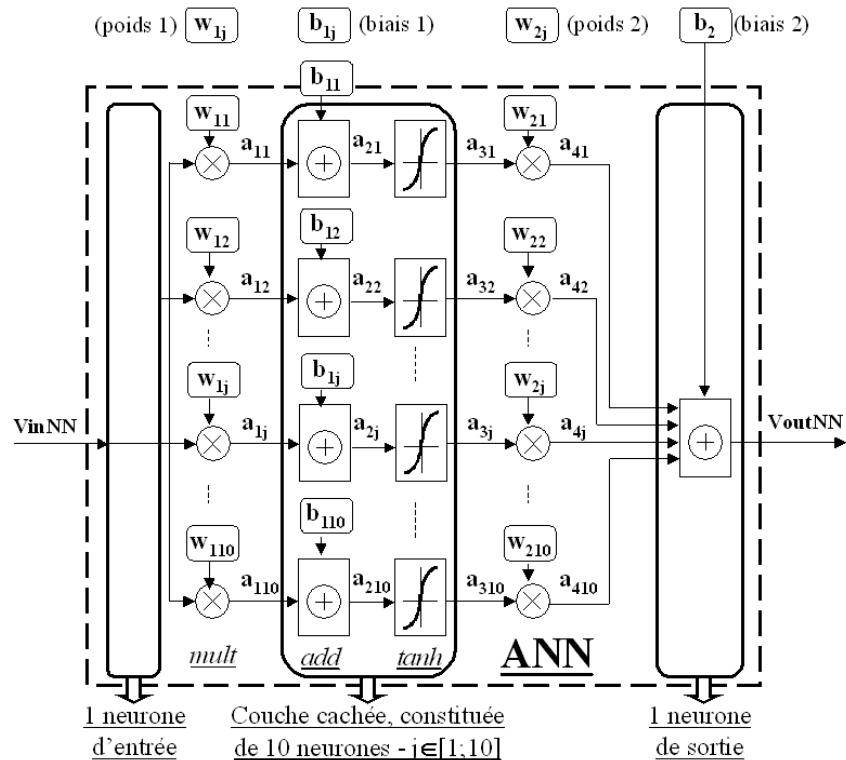


Figure C-d : Structure des réseaux de neurones (NN1 ou NN2).

Le taux de transmission est de 100Mbp/s, soit une bande passante totale de 25MHz pour une modulation 16-MAQ. Il a été choisi d'implémenter le réseau de manière analogique-numérique (*mixed-mode*), de manière à profiter des avantages des deux types d'intégration ([Tap-Béteille, 2003]). Les fonctions de base du perceptron multi-couches (multiplieur, additionneur, tangente hyperbolique) sont réalisées à l'aide de circuits analogiques, tandis que la mise à jour des poids et biais est effectuée à l'aide d'un système numérique (FPGA). La communication entre les deux types d'intégration est assurée par des convertisseurs analogiques-numériques et numériques-analogiques.

Cette méthode permet d'optimiser la puissance consommée, la bande passante et la taille du réseau de neurones tout en facilitant la mise à jour des poids et biais. Le système numérique ne doit pas être nécessairement rapide, puisque cette mise à jour est effectuée toutes les secondes, de manière à compenser les lentes variations de la caractéristique non linéaire de l'amplificateur dues aux changements de température, au vieillissement, etc.

C.c. Description des fonctions implémentées et testées

La technologie CMOS 0.6 μ m compatible avec une tension d'alimentation 5V a été choisie pour réaliser un premier démonstrateur de manière à avoir une bande passante élevée et une dynamique importante. Une technologie plus fine, telle que la technologie CMOS 0.35 μ m, permet a priori une bande passante plus importante. Cependant, la tension d'alimentation plus faible (3.3V) pénalise la dynamique des signaux d'entrée.

La capacité de la structure globale choisie à linéariser la caractéristique de l'amplificateur haute puissance a été prouvée par l'intermédiaire des simulations comportementales tenant compte des bruits intrinsèques à la structure, des non-homogénéités et des tensions de décalage (*offset*) relatives à l'intégration ([Langlet, 2002], [Roviras, 2004]). Les différentes fonctions de base du perceptron multi-couche, c'est-à-dire le multiplieur, l'additionneur et la fonction d'activation non linéaire (tangente hyperbolique) ont ensuite été implémentées et qualifiées ([Tap-Béteille, 2003]).

Le multiplieur et l'additionneur ont été réalisés dans la même cellule (figure C-e). Le rôle du multiplieur est de contrôler par un poids w_1 l'amplitude du signal différentiel d'entrée Δv_1 correspondant à la tension d'entrée V_{inNN} du réseau de neurones. La cellule utilisée pour effectuer cette multiplication est une cellule de Gilbert ([Gray, 2001]). Le rôle de l'additionneur est de rajouter au produit des signaux Δv_1 et w_1 un terme relatif proportionnel

au biais b_1 . En régime "petit signal", au premier ordre, le courant différentiel de sortie Δi_{out_add} de la cellule "multiplieur-additionneur" est proportionnel à la somme de la tension différentielle d'entrée Δv_1 multipliée par le poids w_1 et d'un autre courant différentiel proportionnel au biais b_1 :

$$\Delta i_{out_add} = i_{out_add1} - i_{out_add2} = K_{mult1} w_1 \Delta v_1 + K_{add1} b_1 \quad (C-4)$$

K_{mult1} et K_{add1} sont les coefficients de proportionnalité relatifs au multiplieur et à l'additionneur. Le courant différentiel de sortie est transféré à la cellule suivante (tangente hyperbolique) à l'aide de simples miroirs de courant [Gatet, 2006].

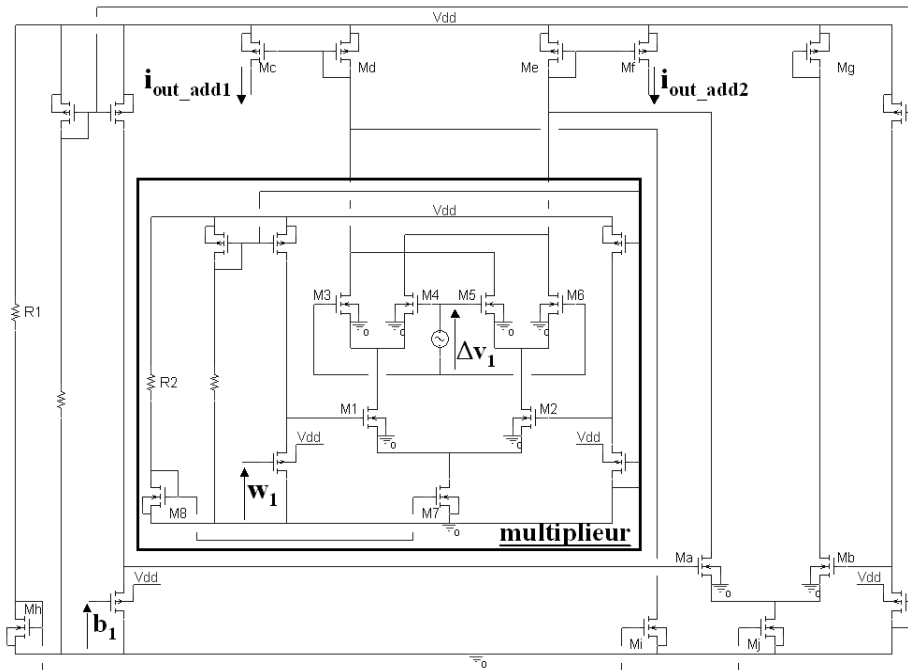


Figure C-e : Structure du "multiplieur-additionneur".

Le rôle de la fonction d'activation est de "compresser" le signal d'entrée différentiel ΔV_{in_th} proportionnel au courant différentiel de sortie Δi_{out_add} du "multiplieur-additionneur". Pour réaliser une distorsion de type sigmoïdale, la caractéristique de transfert au premier ordre $\Delta I_{out_th} \cong I_0 \tanh(K_{th} \Delta V_{in_th})$ de la paire différentielle est utilisée (figure C-f) [Gatet, 2006]. Ainsi, le courant différentiel de sortie $\Delta I_{out_th} = I_{out_th1} - I_{out_th2}$ présente une amplitude plus ou moins écrêtée en fonction de l'amplitude de la tension différentielle d'entrée ΔV_{in_th} .

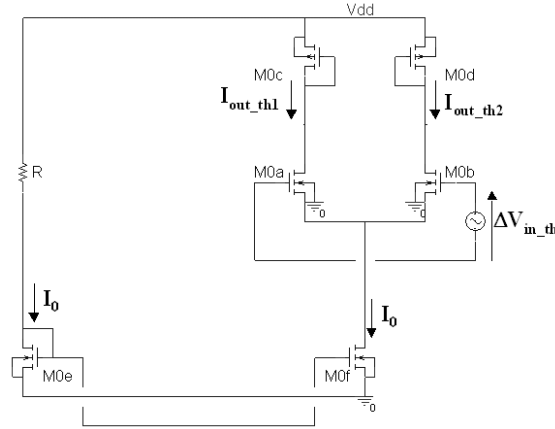


Figure C-f : Structure de la cellule "tangente hyperbolique".

D'après l'équation (C-4), l'expression du courant différentiel de sortie du multiplieur qui suit la fonction tangente hyperbolique est donné par :

$$\Delta i_{out_mult2} = K_{mult2} w_2 \tanh(K_{th}(K_{mult1} w_1 V_{inNN} + K_{add1} b_1)) \quad (C-5)$$

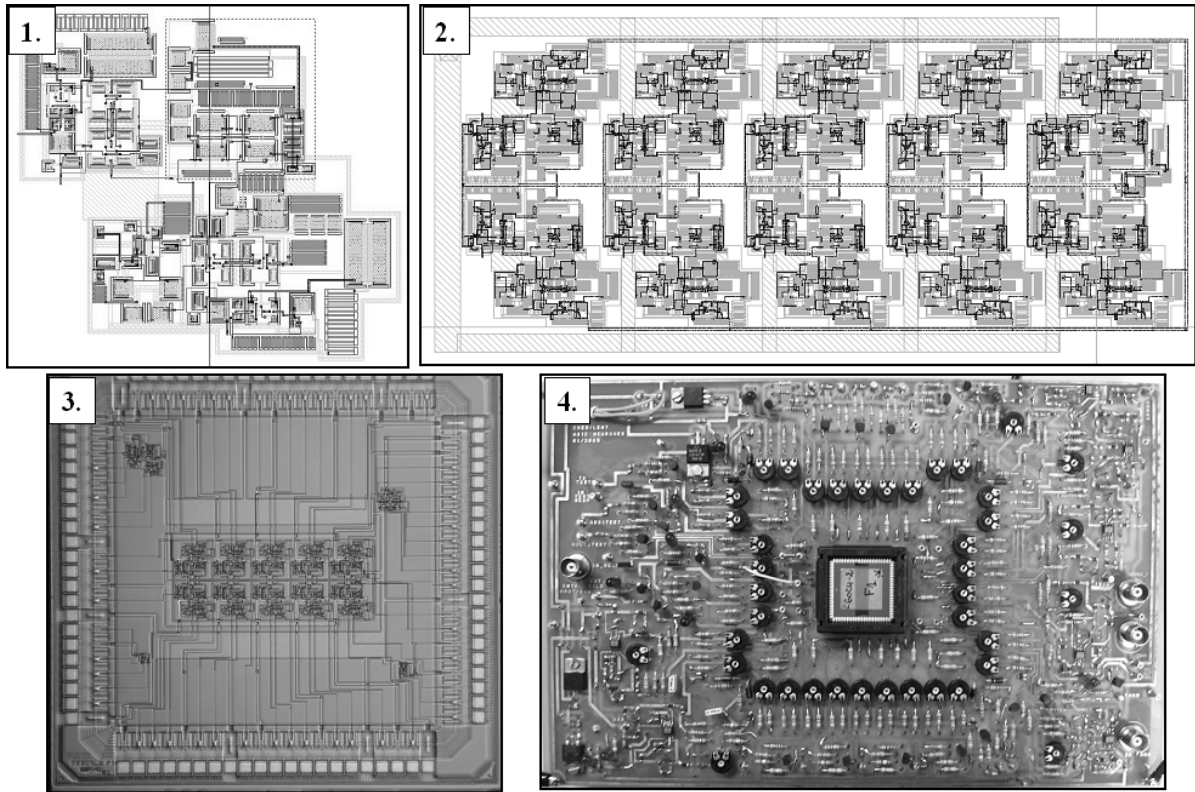
avec K_{mult2} le coefficient de proportionnalité introduit par le deuxième multiplieur.

La tension de sortie V_{outNN} du réseau de neurones est proportionnelle à la somme des courants différentiels de sortie des dix multiplieurs situés entre la couche cachée et le neurone de sortie ($j \in [1;10]$) et d'un autre courant différentiel proportionnel au biais b_2 (figure C-d). Ainsi, en affectant à chaque paramètre issu du neurone j l'indice j ($j \in [1;10]$), la tension de sortie V_{outNN} du réseau de neurones est donnée par :

$$V_{outNN} = K_{NN} [K_{add2} b_2 + \sum_{j=1}^{10} (K_{mult2} w_{2j} \tanh(K_{th}(K_{mult1} w_{1j} V_{inNN} + K_{add2} b_2))] \quad (C-6)$$

avec K_{add2} le coefficient de proportionnalité relatif à l'additionneur situé sur la couche de sortie et K_{NN} le coefficient de proportionnalité relatif à la conversion "courant différentiel – tension" en sortie du réseau de neurone.

Les figures C-g.1 et g.2 montrent respectivement les layouts d'un neurone et du réseau de neurones complet. La figure C-g.3 est une photographie de l'ASIC. La figure C-g.4 représente le PCB réalisé pour tester l'ASIC. L'ensemble des poids et biais est contrôlé en dehors de l'ASIC, par l'intermédiaire de potentiomètres ([Gatet, 2006], [Gatet, 2007]).



*Figure C-g : 1. Layout d'un neurone (dimensions: $330\mu\text{m} \times 300\mu\text{m}$, soit 0.1mm^2 – Taille d'un multiplieur: $70\mu\text{m} \times 70\mu\text{m}$) ;
 2. Layout du réseau de neurones ;
 3. Photographie de l'ASIC ;
 4. Photographie du PCB, avec l'ASIC 84 broches situé au milieu.*

La figure C-h représente un exemple de caractéristiques de transfert mesurées au niveau du buffer de sortie d'un "multiplieur-additionneur" de manière à observer l'influence du poids w_{1j} . La tension $V_{\text{out_add}1j}$ observée en sortie du buffer est proportionnelle au courant différentiel de sortie $\Delta i_{\text{out_add}1j}$ de la cellule. En faisant varier V_{inNN} et w_{1j} pour un biais b_{1j} nul, on détermine l'équation du troisième ordre se rapprochant le plus de la tension observée au niveau du buffer de sortie du multiplieur-additionneur :

$$V_{\text{out_mult}1j} = w_{1j}(3,54V_{\text{inNN}} - 3,11V_{\text{inNN}}^3) \quad (\text{C-7})$$

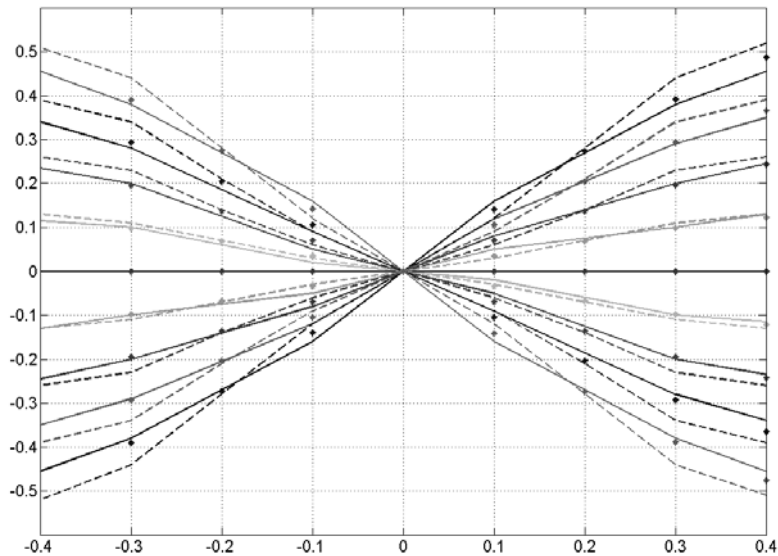


Figure C-h : Comparaison entre simulation (pointillés), test (trait plein) et équation du 3^{ème} ordre (points) au niveau du buffer de sortie d'un multiplieur-additionneur, en fonction de la tension d'entrée V_{inNN} , pour un biais nul et pour un poids variant de $-0.4V$ à $0.4V$ par pas de $0.1V$.

La figure C-i montre le dépassement de la zone de linéarité de la fonction "tangente hyperbolique". La limite est atteinte pour une tension différentielle d'entrée ΔV_{in_thj} voisine de $400mV_{c\grave{a}c}$.

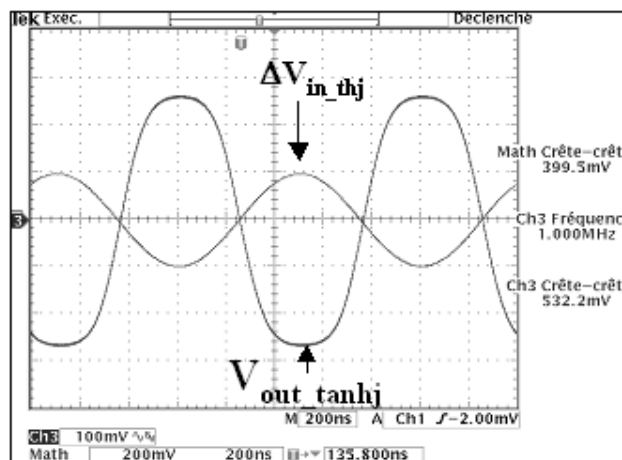


Figure C-i : Observation d'une compression du signal de sortie attribuée à la fonction tangente hyperbolique (amplitude du signal d'entrée de la fonction tangente hyperbolique : $\Delta V_{in_thj} = 400mV_{c\grave{a}c}$).

Enfin, la figure C-j montre quatre oscillogrammes effectués à partir de diverses tensions issues du réseau de neurones. Sur chaque oscillogramme, les quatre tensions représentées de haut en bas sont respectivement la tension d'entrée V_{inNN} , deux tensions V_{out_thj1} et V_{out_thj2} issues de deux neurones parmi les dix, et la tension de sortie V_{outNN} .

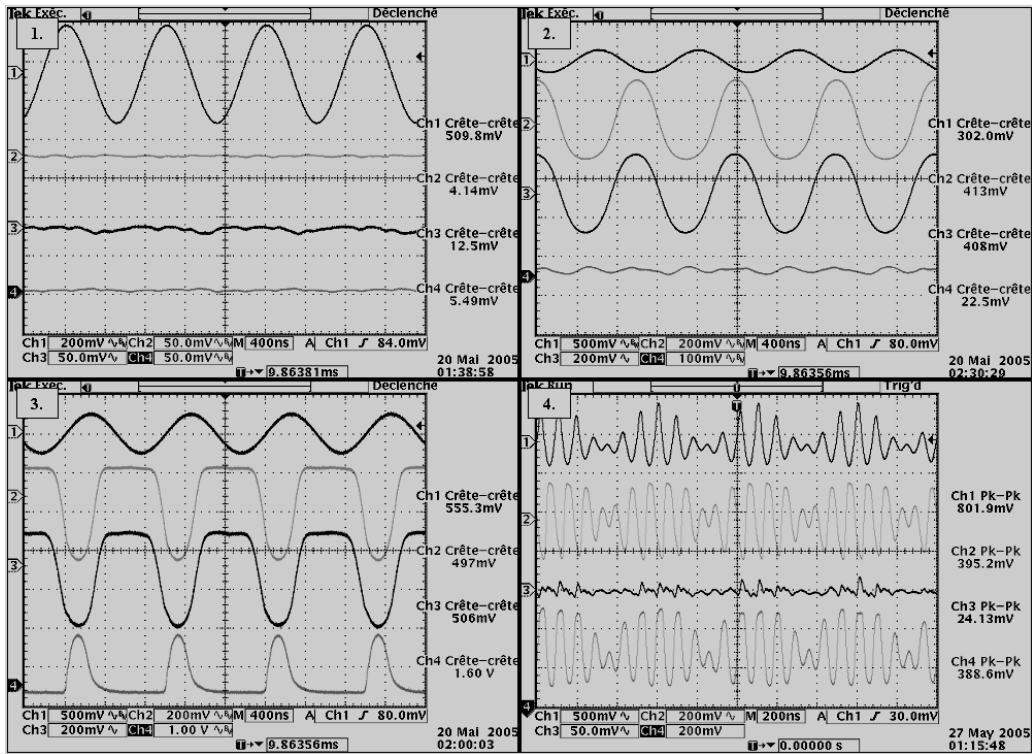


Figure C-j : Exemples d'oscillogrammes réalisés de manière à valider fonctionnellement le réseau de neurones.

1. Conséquence de la mise à 0 de l'ensemble des poids w_{ij} pour une tension d'entrée V_{inNN} non nulle.
2. Conséquence de la mise à 0 de l'ensemble des poids w_{2j} pour une tension d'entrée V_{inNN} et des poids w_{1j} non nuls.
3. Conséquence de la mise à 0.2V de l'ensemble des poids et biais w_{1j} , b_{1j} et w_{2j} pour une tension d'entrée V_{inNN} non nulle.
4. Observation d'un signal d'entrée V_{inNN} modulé (1MHz et 10MHz) issu de la fonction "module au carré" et observation du signal de sortie V_{outNN} correspondant pour l'ensemble des poids w_{1j} nuls, sauf w_{1j1} , pris égal à 0.1V.

Ainsi, à l'issue de ces différents tests, le réseau de neurones analogique a été qualifié fonctionnellement. Cette validation fonctionnelle va permettre de passer à la réalisation de la mise à jour périodique des poids et biais à l'aide de circuits numériques, de manière à corriger les erreurs de linéarité dues aux variations de température et au vieillissement des composants.

Annexe D

Conception de l'ASIC

Le neurone a été implanté dans un boîtier de type JLCC44, sur un substrat contenant plusieurs autres circuits. Les masses des différents circuits ont été soigneusement maillées et reliées à 6 plots. Chaque plot d'entrée-sortie présente des protections contre les décharges électrostatiques (*ESD protections*) par l'intermédiaire de diodes CMOS. Tous les plots d'entrées présentent une résistance série de 500Ω de manière à limiter une variation trop importante de la tension de grille en cas de décharge.

La figure D-a présente l'attribution des broches allouées au neurone élémentaire pour chacune des entrées/sorties. Le neurone élémentaire utilise 17 broches de l'ASIC hors masses. Une des 17 broches est allouée à l'alimentation +3.3V, sept broches aux deux entrées V_{inNN_1} et V_{inNN_2} , aux quatre poids et biais w_{11} , w_{12} , b_1 et w_2 , et à la sortie du neurone a_4 , deux broches à la polarisation en courant du neurone et de la fonction tangente hyperbolique et sept broches à des tests intermédiaires sur le neurone de manière à valider chaque cellule intermédiaire (Tableau D- α).

Seules 10 broches sont nécessaires au bon fonctionnement de l'ASIC, hors masses : une alimentation, six broches pour les deux entrées et les quatre poids et biais, deux broches pour les polarisations des courants et une broche pour la sortie du neurone.

La figure D-b présente le layout de l'ASIC contenant, outre le neurone, un régulateur de tension (à gauche), une photodiode à avalanche (au centre) et une structure double-échantillonnage corrélé de type Clamp utilisée pour traiter des signaux CCD jusqu'à 10MHz (en bas à droite).

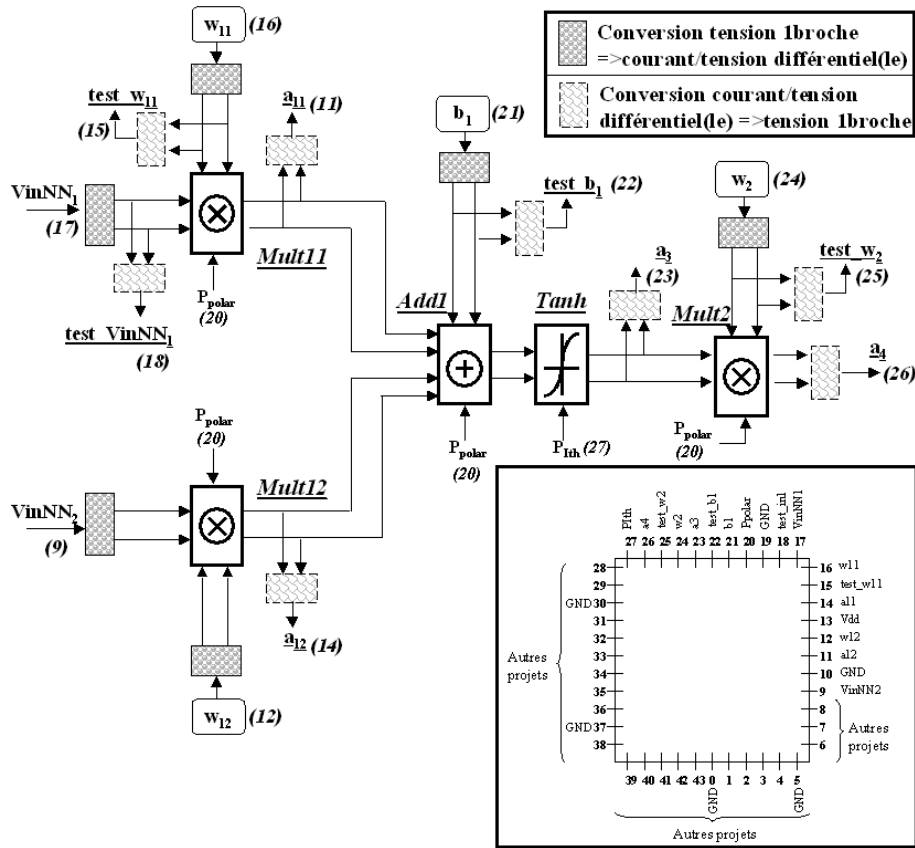


Figure D-a : Attribution des broches pour chaque entrée/sortie.

Tableau D-α : Attribution des plots d'entrée/sortie de l'ASIC.

Type d'entrée/sortie	Nom de la broche	n°	Description
Masses (6)	GND	0,5,10,19,30,37	Masses communes à l'ensemble des projets contenus dans l'ASIC
Alimentation (1)	Vdd	13	Alimentation spécifique au neurone élémentaire
Entrées (6+2)	VinNN ₁	17	Première entrée du neurone
	VinNN ₂	9	Deuxième entrée du neurone
	w ₁₁	16	Poids w ₁₁ référencé à 0V
	w ₁₂	12	Poids w ₁₂ référencé à 0V
	b ₁	21	Biais b ₁ référencé à 0V
	w ₂	24	Poids w ₂ référencé à 0V
	Ppolar	20	Polarisation en courant du neurone (I ₀)
	Pith	27	Polarisation en courant de la tangente hyperbolique (I _{th})
Sorties (1+7)	a4	14	Sortie du multiplieur Mult2 / Sortie du neurone
	a11	14	Sortie du multiplieur Mult11
	a12	11	Sortie du multiplieur Mult12
	a3	23	Sortie de la tangente hyperbolique
	test_in1	18	Test de la conversion de la tension VinNN1 en courant différentiel
	test_w11	15	Test de la conversion du poids w11 en tension différentielle
	test_b1	22	Test de la conversion du poids b1 en tension différentielle
	test_w2	25	Test de la conversion du poids w2 en tension différentielle

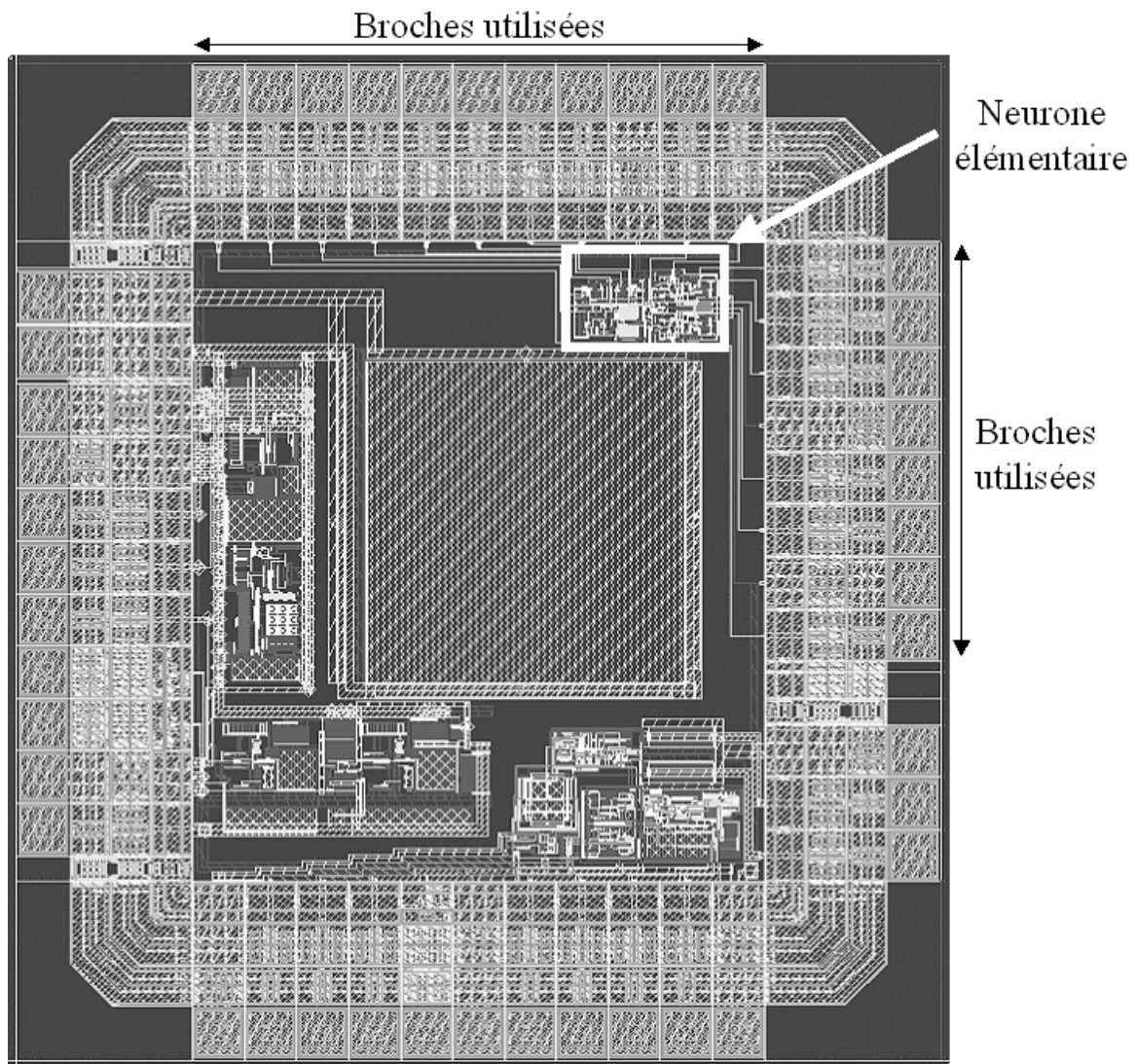


Figure D-b : Layout de l'ASIC complet. Le neurone élémentaire utilise 17 broches et occupe 1/20 de la surface silicium.

REFERENCES DES ANNEXES

- [Bouchired, 1998] : BOUCHIRED S., IBNKAHLA M., ROVIRAS D., CASTANIE F., Neural Networks : A Tool for Satellite UMTS Channel Equalization, *DSP 98* (ESA).
- [Bouchired, 1999] : BOUCHIRED S., IBNKAHLA M., ROVIRAS D., CASTANIE F., Equalization of satellite mobile channels with neural network techniques, *Space Communications 15*, IOS Press, p. 209-220, 1998-99.
- [Broyden, 1970] : BROYDEN C.G., *The convergence of a class of double-rank minimization algorithms 2 : the new algorithm*, Journal of the institute of Mathematics and its applications, 6, p. 222-231.
- [Dreyfus, 2004] : DREYFUS G., MARTINEZ J.M., SAMUELIDES M., GORDON M.B., BADRAN F., THIRIA S., HERAULT L., *Réseaux de neurones, méthodologie et applications, sous la direction de Gérard Dreyfus*, 2^{ème} édition, Eyrolles, avril 2004.
- [Gatet, 2006] : GATET L., TAP-BETEILLE H., LESCURE M., ROVIRAS D., MALLET A., Design and Test of a CMOS MLP Analog Neural Network for Fast On-Board Signal Processing, *13th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2006)*, Nice, France, Dec. 11-13, 2006, Pages: 922-925.
- [Gatet, 2007] : GATET L., TAP-BETEILLE H., LESCURE M., ROVIRAS D., MALLET A., Functional tests of a 0.6 μ m CMOS MLP Analog Neural Network for fast on-board signal processing, *Analog Integrated Circuits and Signal Processing, Springer Science & Business Media* (première lecture).
- [Gray, 2001] : GRAY P., HURST P., LEWIS S., MEYER R., *Analysis and design of analog integrated circuits, 4th ed.*, John Wiley and sons, 2001.
- [Haykin, 1994] : HAYKIN S., Neural Network : A Comprehensive Foundation, *Prentice Hall*, 1994.
- [He, 2006] : HE Z.Y., GE J.H., GENG S.J., WANG G., An improved look-up table predistortion technique for HPA with memory effects in OFDM systems, *IEEE Transactions on Broadcasting*, Volume 52, Issue 1, March 2006, Page(s):87 – 91.
- [Langlet (a), 2001] : LANGLET F., ABDULKADER H., ROVIRAS D., MALLET A., CASTANIE F., Adaptive Predistorsion for Solid State Power Amplifier using Multi-Layer Perceptron, *Globcom 2001, Proceedings*, San Antonio, Texas, USA, November 2001.
- [Langlet (b), 2001] : LANGLET F., ABDULKADER H., ROVIRAS D., MALLET A., CASTANIE F., Comparison of Neural Network Adaptive Predistorsion Techniques for Satellite Down Links, *IJCNN 2001, Proceedings*, Washington DC, USA, July 2001.
- [Langlet, 2002] : LANGLET F., ROVIRAS D., MALLET A., CASTANIE F., Mixed analog/digital implementation of MLP NN for predistortion, *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on Volume 3*, Page(s):2825 – 2830 (12-17 May 2002).
- [Langlet, 2004] : LANGLET F., *Etude et Implantation de la Prédistorcion d'Amplificateurs à Bord de Satellites à base de Réseaux de Neurones*. Thèse de doctorat, Institut National Polytechnique de Toulouse, mai 2004.
- [Lazzarin, 1994] : LAZZARIN G., PUPOLIN S., SARTI A., Nonlinearity Compensation in Digital Radio Systems, *IEEE Transactions on Communications*, Vol. 42, No. 2/3/4, February/March/April 1994.

- [Levenberg, 1944] : LEVENBERG K., A method for the solution of certain non-linear problems in least squares, *Quarterly Journal of Applied Mathematics*, 2, p. 164-168.
- [Marquardt, 1963] : MARQUARDT D.W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society of Industrial and Applied Mathematics*, 11, p. 431-441.
- [Roviras, 2004] : ROVIRAS D., ABDULKADER H., TAP-BETEILLE H, CASTANIE F., LESCURE M., MALLET A., Multi-Layer Perceptron Neural Network Implementation and Integration in CMOS Technology, *International Conference on Information and Communications Technologies : From Theory to Application (ICCTA 94)*, Damascus, Syria, 19-23 April 2004.
- [Rumelhart, 1986] : RUMELHART D.E., HINTON G.E., WILLIAMS R.J., Learning internal representations by error backpropagation, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, p.318-362, MIT Press.
- [Saleh, 1981] : SALEH A., Frequency-independent and frequency-dependant non linear models of TWT amplifiers, *IEEE Trans. On Comm., Vol. COMM 29*, Nov.1981.
- [Tap-Béteille, 2003] : TAP-BETEILLE H., ROVIRAS D., LESCURE M., MALLET A., CASTANIE F., Integration in CMOS Technology of a High Power Amplifier Predistorter, *SPSC2003*, Catania, Italy, 24-26 September 2003.
- [Tummla, 1997] : TUMMLA M., DONOVAN M.T., WATKINS B.E., NORTH R., Volterra series based modeling and compensation of nonlinearities in high power amplifiers, *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997* , Volume 3, 21-24 April 1997, Page(s):2417 – 2420.
- [Werbos, 1974] : WERBOS P.J, *Beyond regression : new tools for prediction and analysis in the behavioural sciences*, Ph. D. thesis, Harvard University.

Liste des figures et tableaux

Figure I-a : Un neurone réalise une fonction non linéaire bornée $y = f(x_1, x_2, \dots, x_n)$ où les $\{x_i\}$ sont les entrées et y la sortie.

Figure I-b : Neurone de base du perceptron multi-couches.

Figure I-c : Neurone RBF à non-linéarité gaussienne.

Figure I-d : Représentation d'un réseau de neurones à n entrées, une couche de N_j neurones cachés et N_m neurones sur la couche de sortie.

Figure I-e : Méthode itérative de rétropropagation de l'erreur appliquée à un perceptron multi-couches de manière à figer ses poids et biais.

Figure I-f : Mise en œuvre de la capacité de classification du réseau de neurones par sa faculté d'approximation de fonctions non-linéaires.

Figure II-a : Représentation de la tête optique d'un télémètre à diode laser fonctionnant par temps de vol.

Figure II-b : Schéma synoptique d'un télémètre FMCW. Une ligne à retard facilite la mesure de la fréquence de battement.

Figure II-c : Schéma de principe du télémètre laser par déphasage avec méthode hétérodyne. Les filtres passe-bande accordés sur la fréquence intermédiaire permettent de réduire la bande passante équivalente de bruit.

Figure II-d : 1. Partie émettrice et réceptrice du télémètre laser par déphasage.
2. Tête optique du télémètre.

Figure II-e : Schéma de principe du télémètre laser par déphasage avec méthode hétérodyne et mélange optoélectronique au sein de la photodiode à avalanche (APD).

Figure II-f : Amplitude du signal de sortie de l'amplificateur transimpédance $V_{outTRANS}$ en fonction de la distance pour les 4 types de surfaces S_1 à S_4 à incidence normale.

1. Pour une valeur de distance fixée ($D=1m$), les 4 surfaces peuvent être facilement discriminées par l'intermédiaire de l'amplitude du signal $V_{outTRANS}$.

2. Lorsque la distance entre le télémètre et la cible n'est pas fixée, une même amplitude $V_{outTRANS}$ peut correspondre aux 4 types de surfaces testées, placées à des distances différentes.

Figure II-g : 1. Restriction de l'intervalle de distance à l'intervalle $[0.93m;1.05m]$, de manière à ce que chaque amplitude appartenant à l'intervalle V_{outS_i} ($i \in [1;4]$) soit bijective par rapport au type de surface S_i testée pour un angle d'incidence nul.

2. Restriction de l'intervalle de distance à l'intervalle $[0.97m;1.02m]$, de manière à ce que chaque amplitude appartenant à l'intervalle V_{outS_i} ($i \in [1;4]$) soit bijective par rapport au type de surface S_i testée pour un angle d'incidence variant de $-\pi/5$ à $+\pi/5$. Les droites grises et noires représentent l'amplitude du signal $V_{outTRANS}$ pour un angle d'incidence respectivement égal à 0 et à $\pm\pi/5$.

Figure III-a : Variations de l'amplitude du signal photoélectrique en fonction de la distance. La caractéristique idéale proportionnelle à l'inverse du carré de la distance est tracée en pointillés ; l'ensemble des points expérimentaux est représenté par des croix. Pour chaque valeur de distance, 5 mesures ont été réalisées dans des conditions expérimentales différentes.

Figure III-b : Fonction g que doit réaliser le réseau de neurones de manière à ce que sa sortie soit proportionnelle à la distance (L 'angle d'incidence θ_0 est considéré constant et K_0 est choisi égal à $0.1V.m^{-1}$).

Figure III-c : Système complet de levée d'indétermination sur la mesure de distance issue d'un télémètre laser par déphasage à l'aide d'un réseau de neurones.

Figure III-d : Structure du réseau de neurones réalisant la fonction g .

Figure III-e : Courbes représentant la répartition des valeurs des amplitudes d'entrée du réseau de neurones, fonction des mesures de distance choisies dans l'intervalle $[0.5m;3.5m]$.

Figure III-f : Approximation de la fonction g idéale (pointillés) à partir des 9 valeurs de distance (points) à l'aide d'un polynôme de degré 7 (trait plein). L'erreur commise est inférieure à 10mV.

Figure III-g : Méthode utilisée pour calculer les poids et biais à partir des valeurs de distance issues du jeu d'apprentissage.

Figure III-h : De 1. à 4.: Variation des poids et biais lors des $n_{ptse}=100$ dernières itérations. 5.: Valeurs $V_{inNN}(k)$ présentées en entrée du réseau de neurones lors des 50 dernières itérations, correspondant aux valeurs aléatoires de distance choisies dans l'ensemble d'apprentissage.

6.: Sorties idéale $V_{outNNid}(k)$ (pointillés) et réelle $V_{outNN}(k)$ lors des 50 dernières itérations. Les deux courbes sont quasiment superposées.

7. Erreur $e(k)$ (III-17) (trait plein) et erreur maximale e_{max} (pointillés).

Figure III-i : 1.: Ensemble des valeurs d'entrée V_{inNN} du réseau relatives aux 100 valeurs de distance appartenant au jeu de validation. Les 9 valeurs de l'ensemble d'apprentissage sont représentées par des points.

2.: Tensions de sortie idéale ($V_{outNNid}$) et réelle (V_{outNN}) du RN en fonction de V_{inNN} .

3.: Tensions de sortie idéale ($V_{outNNid}$) et réelle (V_{outNN}) du RN en fonction de D .

4.: Erreur (trait plein) ; l'erreur maximale $\pm e_{max}$ est représentée en pointillés.

5. Diagramme de dispersion.

Figure III-j : Découpage de l'espace des tensions d'entrée en quatre zones relatives à chacune des surfaces. Chaque zone est limitée par les valeurs des amplitudes des signaux pour des angles d'incidence respectivement égaux à 0 et $\pm\theta_{0max}$. Chaque frontière S_i-S_j est équidistante des zones relatives aux surfaces S_i et S_j .

Figure III-k : Tension d'entrée V_{inNN_1} du réseau de neurones en fonction de la distance en échelle log-log pour un angle d'incidence égal à $\pi/6$. Les courbes en trait plein sont relatives aux mesures expérimentales réalisées avec le télémètre laser par déphasage. Les droites grises pointillées correspondent aux courbes idéales (suivant la relation III-1) relatives à chacune des surfaces. Les droites noires pointillées représentent les frontières S_i-S_j entre les zones relatives à chacune des surfaces.

Figure III-l : Représentation du système de détection de surface mettant en œuvre un télémètre laser par déphasage et un réseau de neurones.

Figure III-m : Structure du réseau de neurones pour la discrimination de surfaces.

Figure III-n : Méthode de répartition des valeurs de distance appartenant au jeu d'entraînement, représentée en échelle log-log. Les différentes valeurs d'entraînement choisies pour calibrer le réseau sont mises en valeur par des points.

Figure III-o : Courbes représentant l'amplitude d'entrée V_{inNN_1} du réseau de neurones associée aux 4 types de surfaces en fonction de la distance dans l'intervalle $[0.5m;2m]$, pour cinq valeurs d'angles d'incidence comprises entre 0 et $\pi/5$. Les points représentent les valeurs d'entraînement.

Figure III-p : Méthode utilisée pour calculer les poids et biais à partir des valeurs de distance et d'angle d'incidence issues de leur jeu d'apprentissage respectifs.

Figure III-q : De 1. à 5.: Variation des poids et biais lors des $n_{ptse}=500$ dernières itérations. 6.: Sorties idéale ($V_{outNNid}(k)$ à l'itération k en pointillés) et réelle ($V_{outNN}(k)$ à l'itération k en trait plein) lors des 50 dernières itérations. Les deux courbes sont quasiment superposées.

7. Erreur $e(k)$ (III-17) (trait plein) et erreur maximale $\pm e_{max}$ (pointillés).

Figure III-r : 1. et 2.: Ensemble des valeurs d'entrée V_{inNN_1} et V_{inNN_2} du réseau relatives aux 60 valeurs de distance et aux 10 valeurs d'angle d'incidence appartenant à leur jeu de validation respectif. Les valeurs relatives à l'ensemble d'apprentissage sont représentées par des points.

3.: Tensions de sortie idéales relatives aux surfaces testées ($V_{outNNid}$ - trait gras) et réelles (V_{outNN} - pointillés) du réseau de neurones.

4.: Erreurs en valeur absolue (trait plein) et erreur maximale e_{max} .

5.: Erreurs binaires.

Figure IV-a : Schéma d'un transistor NMOS.

Figure IV-b : Modèle petit signal haute fréquence du transistor MOS en zone source de courant.

Figure IV-c : Evolution de la caractéristique $I_D(V_{DS})$ d'un transistor MOS en technologie $0.35\mu m$ pour différentes longueurs de canal, avec un rapport W/L constant de 10 et pour un V_{GS} constant égal à 1V.

Figure IV-d : Miroir de courant simple.

Figure IV-e : Schéma équivalent petit signal du miroir de courant simple.

Figure IV-f : Miroir de courant cascodé.

Figure IV-g : Schéma équivalent pour la détermination de la conductance de sortie du miroir cascode.

Figure IV-h : Circuit différentiel à sources couplées.

Figure IV-i : Comparaison entre la caractéristique de transfert "exacte" de la paire différentielle NMOS ($I_0=180\mu A$, $W/L=35\mu A/1\mu A$) et la fonction relative à l'équation (IV-21). Les deux courbes sont quasiment superposées.

Figure IV-j : Domaine de linéarité du transfert d'une paire différentielle NMOS, avec $W/L = 2\mu m/1\mu m$ et $I_0 = 100\mu A$.

Figure IV-k : Plage admissible de la tension de mode commun v_M avec l'hypothèse d'une très faible amplitude de la tension différentielle v_d .

Figure IV-l : Circuit différentiel à sources couplées associées à deux sources suiveuses. Les sources suiveuses M_{P11} et M_{P12} permettent d'obtenir une tension de mode commun admissible de 0V.

Figure IV-m : Schéma du multiplieur.

Figure IV-n : Schéma-bloc du multiplieur. Le poids w est converti en poids différentiel $\pm w/2$. Le courant différentiel d'entrée ΔI_{in} correspond au courant différentiel de sortie de la cellule précédente.

Figure IV-o : Représentation (en trait plein) de la variation du courant différentiel de sortie ΔI_{Dmult} en fonction de la tension v_{in} pour différentes valeurs de poids w . Les courbes en pointillés sont relatives à la multiplication du courant différentiel ΔI_{in} issu de la cellule précédente avec les différents poids, selon l'équation (IV-31).

Figure IV-p : Observation du courant différentiel de sortie de l'additionneur ΔI_{Dadd} en fonction du biais b . La zone de linéarité associée à une précision relative de 15% est de l'ordre de $1.5V_{c\grave{a}c}$, celle associée à une précision relative de 5% est de l'ordre de $1V_{c\grave{a}c}$.

Figure IV-q : Schéma bloc de la fonction tangente hyperbolique faisant intervenir les cellules qui la précède. Pour alléger la figure, les transferts différentiels en courant entre cellules n'ont été représentés que par une flèche.

Figure IV-r : Schéma de la tangente hyperbolique.

Figure IV-s : Décalage de la tension commune aux deux résistances d'un V_{GS} .

Figure IV-t : Variations des courants différentiels d'entrée et de sortie de la fonction tangente hyperbolique en fonction de la tension d'entrée v_{inNN_1} pour différentes valeurs de poids w_{11} . La tension d'entrée v_{inNN_2} et le biais b_1 ont été pris égaux à 0V.

Figure IV-u : Variations des courants différentiels d'entrée et de sortie de la fonction tangente hyperbolique en fonction de la tension d'entrée v_{inNN1} pour différentes valeurs de biais b_1 . La tension d'entrée v_{inNN2} et le poids w_{11} ont été pris respectivement égaux à 0V et 0.1V.

Figure IV-v : Structure d'un neurone élémentaire.

Figure IV-w : Schéma bloc du neurone élémentaire réalisé.

Figure IV-x : Layout du neurone élémentaire.

Taille du neurone hors plots d'entrée/sortie : $276\mu\text{m} \times 148\mu\text{m} = 0.04\text{mm}^2$.

Figure IV-y : PCB utilisé pour tester le neurone élémentaire et les autres circuits implantés dans l'ASIC.

Figure IV-z : Comparaison entre la simulation (pointillés), le test (trait plein) et l'équation de la caractéristique la plus proche (points) du premier multiplieur en fonction de V_{inNN1} pour différentes valeurs de poids $w_{11} \in [-0.5\text{V}; 0.5\text{V}]$ par pas de 0.1V.

Figure IV-aa : Comparaison entre la simulation (pointillés), le test (trait plein) et l'équation de la caractéristique la plus proche (points) de la fonction tangente hyperbolique en fonction de V_{inNN1} pour différentes valeurs de poids $w_{11} \in [-0.5\text{V}; 0.5\text{V}]$ par pas de 0.1V. La tension d'entrée V_{inNN2} , le poids w_{12} et le biais b_1 sont fixés à 0V.

Figure IV-ab : Comparaison entre la simulation (pointillés), le test (trait plein) et l'équation de la caractéristique la plus proche (points) du deuxième multiplieur en fonction de V_{inNN1} pour différentes valeurs de poids $w_2 \in [-0.5\text{V}; 0.5\text{V}]$ par pas de 0.1V. Le poids w_{11} est fixé à 0.1V. La tension d'entrée V_{inNN2} , le poids w_{12} et le biais b_1 sont fixés à 0V.

Figure IV-ac : Structure du réseau de neurones. En gras, la structure d'un neurone élémentaire déjà réalisé et intégré dans un ASIC.

Figure IV-ad : Représentation du réseau de neurones implanté sur PCB, avec un neurone intégré dans chaque ASIC.

Figure IV-ae : Schéma électrique du réseau constitué de 3 neurones intégrés sur la couche cachée. Le neurone de sortie, réalisant la somme des courants issus des 3 neurones avec un courant proportionnel au biais b_2 , est directement implanté sur le PCB.

Figure V-a : Courbes représentant la répartition des valeurs des amplitudes d'entrée du réseau de neurones, fonction des valeurs de distance dans l'intervalle [0.5m;3.2m].

Figure V-b : Description de la phase d'entraînement.

Figure V-c : De 1. à 4.: Variation des poids et biais lors des 100 dernières itérations.

5.: Valeurs $V_{inNN}(k)$ présentées en entrée du réseau de neurones lors des 50 dernières itérations, correspondant aux valeurs aléatoires de distance choisies dans l'ensemble d'apprentissage.

6.: Sorties idéale $V_{outNNid}(k)$ (pointillés) et réelle $V_{outNN}(k)$ (trait plein) lors des 50 dernières itérations.

7. Erreur $e(k)$ (trait plein) et erreur maximale e_{max} (pointillés).

Figure V-d : 1.: Ensemble des valeurs d'entrée V_{inNN1} du réseau. Les valeurs relatives à l'ensemble d'apprentissage sont représentées par des points.

2.: Tensions de sortie idéale $V_{outNNid}$ (pointillés) et réelle V_{outNN} (trait plein) du réseau de neurones.

3.: Erreur (trait plein) ; l'erreur maximale $\pm e_{max} = \pm 195\text{mV}$ est représentée en pointillés.

Figure V-e : Comparaison entre résultats expérimentaux, simulation et idéalité.

Figure V-f : Courbes représentant la répartition des valeurs des tensions d'entrée V_{inNN1} du réseau de neurones, fonction du type de surface, de l'angle d'incidence variant de $-\pi/6$ à $\pi/6$ et des mesures de distance choisies dans l'intervalle [0.5m;1.25m]. Les points correspondent aux tensions d'entrée associées aux trois ensembles d'entraînement.

Figure V-g : Description de la méthode utilisée pour calculer les poids et biais.

Figure V-h : De 1. à 5.: Variation des poids et biais lors des 200 dernières itérations.
6.: Sorties idéale $V_{outNNid}(k)$ (pointillés) et réelle $V_{outNN}(k)$ (trait plein) lors des 50 dernières itérations. Les deux courbes sont quasiment superposées.

7. Erreur $e(k)$ (trait plein) et erreur maximale $e_{max}=\pm 170mV$ (pointillés).

Figure V-i : 1. et 2.: Ensemble des valeurs d'entrée V_{inNN_1} et V_{inNN_2} du réseau relatives aux 35 valeurs de distance et aux 5 valeurs d'angle d'incidence appartenant à leur jeu de validation respectif. Les valeurs appartenant à l'ensemble d'apprentissage sont représentées par des points.

3.: Tensions de sortie idéale relative à la surface testée ($V_{outNNid}$ - trait gras) et réelle (résultant de l'équation (V-6)) du réseau de neurones.

4.: Erreur en valeur absolue (trait plein) et erreur maximale $e_{max}=0.17V$.

5.: Erreur binaire.

Figure V-j : Comparaison entre résultats expérimentaux, simulation et idéalité.

Figure V-k : Approximation de la tangente hyperbolique à l'ordre 3.

$$a_3 = a_2 - \frac{a_2^3}{3}.$$

Figure V-l : Schéma d'implantation numérique du neurone.

$$a_4 = w_2(a_2 - \frac{a_2^3}{3}) \text{ avec } a_2 = 4w_{11}V_{inNN_1} + 4w_{12}V_{inNN_2} + b_1.$$

Figure V-m : Caractéristique du multiplieurMult11 en fonction du signal V_{inNN_1} pour différentes valeurs de poids w_{11} .

Figure V-n : Caractéristique de la fonction tangente hyperbolique.

Figure V-o : Puissances consommées par la fonction tangente hyperbolique (pointillés) et par le neurone (trait plein) en fonction de la fréquence.

Figure V-p : Photo du PCB de test du neurone numérique.

Figure A-a : Calcul de la fonction de coût des moindres carrés pour un système d'une donnée ne dépendant que d'une variable.

Figure A-b : Calcul de la fonction de coût des moindres carrés pour un vecteur d'entrée indicé α (avec $\alpha \in [1;N]$) dépendant de n variables dans un système où l'on veut modéliser la variation de N_m données.

Figure A-c : Un signal que l'on voudrait modéliser.

Figure A-d : Nombre fini N de mesures $z_1(x_1)$. Dans cet exemple, les N points de mesure sont répartis uniformément sur l'intervalle de mesure.

Figure A-e : Approximation de la fonction de régression $g(x_1, W)$, réalisée par un réseau de neurones, à partir des points expérimentaux $z_1(x_1)$ de la figure A-d.

Figure B-a : Mise en évidence du phénomène de sur-apprentissage.

Figure C-a : Architecture du système de prédistortion embarqué dans le satellite.

Figure C-b : Représentation des non-linéarités AM/AM normalisée (C-b.1) et AM/PM (en radians) (C-b.2) en fonction de l'amplitude du signal d'entrée normalisée.

Figure C-c : Structure du système de prédistortion.

Figure C-d : Structure des réseaux de neurones (NN1 ou NN2).

Figure C-e : Structure du "multiplieur-additionneur".

Figure C-f : Structure de la cellule "tangente hyperbolique".

Figure C-g : 1. Layout d'un neurone (dimensions: $330\mu\text{m} \times 300\mu\text{m}$, soit 0.1mm^2 – Taille d'un multiplieur: $70\mu\text{m} \times 70\mu\text{m}$) ;

2. Layout du réseau de neurones ;

3. Photographie de l'ASIC ;

4. Photographie du PCB, avec l'ASIC 84 broches situé au milieu.

Figure C-h : Comparaison entre simulation (pointillés), test (trait plein) et équation du 3^{ème} ordre (points) au niveau du buffer de sortie d'un multiplieur-additionneur, en fonction de la tension d'entrée V_{inNN} , pour un biais nul et pour un poids variant de -0.4V à 0.4V par pas de 0.1V .

Figure C-i : Observation d'une compression du signal de sortie attribuée à la fonction tangente hyperbolique (amplitude du signal d'entrée de la fonction tangente hyperbolique : $\Delta V_{in_thj} = 400\text{mV}_{c\grave{a}c}$).

Figure C-j : Exemples d'oscillogrammes réalisés de manière à valider fonctionnellement le réseau de neurones.

1. Conséquence de la mise à 0 de l'ensemble des poids w_{1j} pour une tension d'entrée V_{inNN} non nulle.

2. Conséquence de la mise à 0 de l'ensemble des poids w_{2j} pour une tension d'entrée V_{inNN} et des poids w_{1j} non nuls.

3. Conséquence de la mise à 0.2V de l'ensemble des poids et biais w_{1j} , b_{1j} et w_{2j} pour une tension d'entrée V_{inNN} non nulle.

4. Observation d'un signal d'entrée V_{inNN} modulé (1MHz et 10MHz) issu de la fonction "module au carré" et observation du signal de sortie V_{outNN} correspondant pour l'ensemble des poids w_{1j} nuls, sauf w_{1j1} , pris égal à 0.1V .

Figure D-a : Attribution des broches pour chaque entrée/sortie.

Figure D-b : Layout de l'ASIC complet. Le neurone élémentaire utilise 17 broches et occupe $1/20$ de la surface silicium.

Tableau III- α : Les tensions de sortie idéales du réseau de neurones et les intervalles de sortie associés à chacun des types de surfaces sont représentées dans les deux premières colonnes. Les codes associés aux sorties respectives de chacun des détecteurs de seuil et relatifs au type de surface testée sont représentés dans les trois dernières colonnes.

Tableau IV- α : Variation de la conductance de sortie g_{ds} pour différentes longueurs de canal, avec un rapport W/L constant de 10 et pour $V_{GS} = 1\text{V}$ et $V_{DS} = 1\text{V}$.

Tableau IV- β : Consommation en courant des cellules et puissance dissipée.

Tableau V- α : Tensions de sortie analogiques idéales du réseau de neurones, intervalles de sortie et tensions de sortie idéales des différents détecteurs de seuil associés.

Tableau D- α : Attribution des broches d'entrées/sorties.

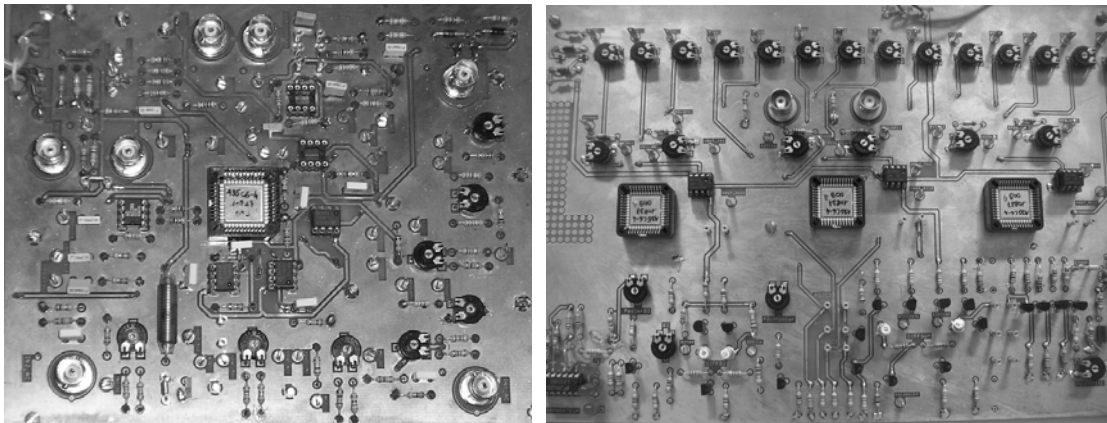
Résumé

L'objectif de cette thèse est de montrer l'intérêt d'adapter un réseau de neurones à un type de système optoélectronique de mesure de distance, la télémétrie laser par déphasage. Deux études ont été menées en parallèle. La première consiste à lever l'indétermination sur la mesure de distance déduite de la mesure de déphasage. La seconde étude permet la classification de différents types de surfaces à partir de deux signaux issus du télémètre.

Une première étude met en valeur la capacité du réseau de neurones à accroître la plage de mesure de distance sans modifier la résolution du télémètre. Elle a été réalisée à partir de mesures expérimentales afin de prouver le réel intérêt de la méthode comportementale développée.

Une deuxième étude ouvre une nouvelle perspective relative à l'utilisation d'un télémètre laser par déphasage, celle d'effectuer la classification de différents types de surfaces sur des plages de distances et d'angles d'incidence variables.

Pour valider expérimentalement ces deux études, les cellules de base d'un neurone de type perceptron multi-couches ont été simulées puis intégrées de manière analogique. Un démonstrateur du réseau de neurones global a été implanté à partir de neurones élémentaires intégrés mis en parallèle. Une étude de la conception des mêmes cellules en numérique a été réalisée en parallèle afin de justifier les avantages associés à chaque type d'intégration.



Photographies des circuits imprimés de test d'un neurone élémentaire et du réseau de neurones analogiques.

Les phases d'entraînement et de validation expérimentales du réseau intégré ont été effectuées extérieurement à l'ASIC pour les deux applications, par l'intermédiaire de l'équation de transfert déterminée après caractérisation expérimentale et qualification du réseau de neurones global.

Les résultats expérimentaux issus de la première étude montrent qu'il est possible d'obtenir à partir des signaux de sortie du télémètre et du réseau de neurones, une mesure de distance de précision ($50\mu\text{m}$) sur un intervalle de mesure 3 fois plus important que celui limité à la mesure du déphasage.

Concernant l'application de discrimination de surfaces, le réseau de neurones analogique implanté est capable de classer quatre types de cibles sur l'intervalle $[0.5\text{m} ; 1.25\text{m}]$ pour un angle d'incidence pouvant varier de $-\pi/6$ à $+\pi/6$.

Mots clés

- Réseaux de neurones
- Technologie CMOS basse tension
- Classification de surfaces
- Mesure de distance
- Circuits intégrés analogiques (ASIC)
- Télémétrie laser