

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITÉ FERHAT ABBAS -SETIF-1-
UFAS (ALGERIE)

MEMOIRE

Présenté à la Faculté de Technologie
Département d'Electronique

Pour Obtention du Diplôme de

MAGISTER

Option : Contrôle

Présenté par :

Mr: KORIKER Mustapha

THEME:

MODELISATION DES TRANSISTORS FET APPLIQUÉ AUX DISPOSITIFS RADIOFREQUENCES EN UTILISANT L'APPROCHE NEURONALE

Soutenu le 17/ 12 /2013 devant la commission d'examen :

Mr. A.BOULOUBA	Professeur à l'Université de Sétif	Président
Mr. A. BARTIL	Maitre de Conférences à l'Université de Sétif	Rapporteur
Mr. F.DJAHLI	Professeur à l'Université de Sétif	Examineur
Mr. N.BOUKEZZOULA	Maitre de Conférences à l'Université de Sétif	Examineur

Remerciements

Le travail présenté dans ce mémoire a été effectué au département d'électronique de l'université Ferhat Abbas-SEFIF.

Je remercie premièrement Dieu le tout puissant pour la volonté, la santé et la patience qu'il m'a donnée durant toutes ces années d'études.

Nous tenons également à exprimer nos vifs remerciements à notre encadreur Monsieur **Bartil Arres** pour m'avoir fait confiance en proposant ce thème d'une part et pour son suivi continu tout le long de la réalisation de ce mémoire d'autre part. Il n'a pas cessé de me donner conseils et remarques.

Mes sincères remerciements vont aussi à Messieurs les membres du jury d'avoir bien voulu juger ce travail :

Monsieur le Professeur A.Bouloufa qui m'a fait l'honneur de présider le jury.

Monsieur Le Professeur F.Djahli qui a accepté de juger ce travail.

Monsieur le Docteur N.Boukezzoula qui a répondu favorablement et pour l'intérêt qu'il porte à ce travail.

Mes derniers remerciements vont aussi à tous les enseignants du département d'électronique qui ont contribué à notre formation.

Enfin nous tenons à exprimer notre reconnaissance à tous nos amis et collègues pour leur soutien et disponibilité.

Dédicace

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Je dédie ce modeste travail

- A mes chers **parents**, A mes **frères**, **sœurs**, ma femme (**ben alia Souad**) A tous les membres de ma grande famille.
- A tous mes **professeurs** qui m'ont appris durant toutes mes années d'étude.
- A tous les **enseignants** qui m'ont aidé par leurs disponibilités.
- A tous mes amis Chouiter, **Mohamed**, **Abdel aloidoud**, Ahmed, tayeb, **Sohaybe**, **Ali**, Ismail, Belgacem, saber, **Baker**, **moktar**....
- A tous mes amis que je ne vais pas nommer car ils sont nombreux.
- A tous mes collègues sans exception & à toutes les promos 2013de l'électronique.

Je vous remercie tous...

Mustapha.Koriker

Abréviations et symboles

RNA	réseaux de neurones artificiels
ANN	Artificial Neural Network
EQM	Erreur Quadratique Moyenne
w_{ij}	Les poids
MLP	multi layer perceptron
PMC	perceptron multicouche
RBF	Radial Basis Function
RBFNN	Radial Basis Function Neural Network
μ	Le coefficient d'apprentissage
GPB	La rétro-propagation du gradient de l'erreur
CG	gradient conjugué
BFGS	Broyden, Fletcher, Goldfarb et Shanno
α	le terme d'inertie (momentum).
RF	radiofréquences
FET	transistors à effet de champ
MOSFET	transistor Métal-Oxyde-Semi-conducteur à effet de champ
G	grille
S	source
D	drain
BV_{ds}	Tenue en tension(V)
ϵ_{ox}	La permittivité électrique de l'oxyde
ϵ_{si}	La permittivité électrique du silicium
μ	Mobilité de porteurs dans le canal
μ_n	Mobilité des électrons
μ_p	Mobilité des trous
C	Capacité de grille par unité de longueur
C_{ox}	Capacité d'oxyde
C_{si}	Capacité du film de silicium
g_m	La transconductance
g_{DS}	L'admittance
G_V	Gain en tension

I_D	Courant de drain
$I_{DS_{sat}}$	Courant de drain (saturation)
I_{off}	Le courant à l'état bloqué
I_{on}	Le courant à l'état passant
L	Longueur du canal
L_g	Longueur de la grille
N_A	Concentration en dopants de type P
N_D	Concentration en dopants de type N
N_i	Densité de porteur intrinsèque
n MOSFET	Transistor à canal n
p MOSFET	Transistor à canal p
q Charge	élémentaire électrostatique
R_{on}	La résistance à l'état passant
Si	Silicium
SiO_2	Oxyde de silicium (isolant)
SOI	Silicon On Insulator
SON	Silicon On Nothing
T	Température
ϕ_{ms}	Le travail de sortie métal semi-conducteur
Q_{SS}/ C_{ox}	La tension correspondant à l'effet des charges à l'interface
ϕ_F	Le niveau de fermi pour $N_{a \max}$
T_{ox}	Épaisseur de l'oxyde de grille (SiO ₂)
T_{si}	Épaisseur du Silicium
V_{DD}	Tension d'alimentation de drain
V_{DB}	Tension Drain-Substrat
V_{DS}	Tension Drain-Source
V_{DSaT}	Tension Drain-Source de saturation
V_{FB}	La tension de bonde plate
V_{GS}	Tension Grille-Source
V_{SB}	Tension Source-Substrat
V_{th}	Tension de seuil
W	La largeur de canal
ZCE	Zone de charge d'espace

C_{ox}	Capacité d'oxyde
Si	Silicium
VLSI	Very Large Scale Integration
q	Charge élémentaire électrostatique
W	La largeur de canal
ψ_s	Le potentiel de surface
ψ_{sL}	Le potentiel de surface à la fin de drain
ψ_{s0}	Le potentiel de surface à la source
f	Une fonction pour faciliter le calcul de ψ_s
Q	Densité de charge
V	La différence de tension
K_p	Transconductance du MOSFET

Liste des figures

Figure I.1.Schéma fonctionnel d'un neurone biologique	6
Figure I.2.Modèle de base d'un neurone formel.....	7
Figure I.3.Fonctions Heaviside.....	8
Figure I.4.Fonctions signe.....	8
Figure I.5.Fonction linéaire.....	9
Figure I.6.Fonction linéaire a seuil.....	10
Figure I.7.Fonction sigmoïde.....	10
Figure I.8.Fonction gaussienne.....	11
Figure I.9.Fonction logique OU et le neurone correspondant.....	11
Figure I.10.Réseaux monocouche.....	12
Figure I.11.Réseau Multicouche type feedforward.....	13
Figure I.12.Réseau récurrent symétrique.....	14
Figure I.13.Architecture du réseau RBFNN.....	15
Figure II.1.Mode d'apprentissage des réseaux de neurones non supervisé.....	22
Figure II.2.Mode d'apprentissage des réseaux de neurones supervisé.....	23
Figure II.3.algorithme de rétro-propagation du gradient	26
Figure II.4.Courbe explicative du phénomène du minimum local.....	35
Figure II.5.Descente de gradient dans l'espace des poids	38
Figure III.1.Vue schématique du transistor MOS de type N.....	42
Figure III.2.Diagramme de bandes du transistor MOS en régime d'inversion Forte suivant l'axe y en un point quelconque du canal.....	44
Figure III.3.Comparaison entre la modélisation et la mesure du courant de drain en régime linéaire pour un transistor MOS avec un oxyde ultra mince $T_{ox} = 2.1nm$	49
Figure III.4.Comparaison entre la modélisation et la mesure du courant de drain en fonction de la tension de drain pour différentes valeurs de la tension de grille pour un transistor MOS avec un oxyde ultra mince $T_{ox} = 2.1nm$	49
Figure III.5.Le MOSFET à canal n sur substrat massif. a) schéma du transistor b) illustration du potentiel distribué dans le transistor.....	50
Figure III.6. a)Représentation schématique du MOS en mode d'avalanche (jonction) b) Illustration du MOS lors de l'application d'une tension $V_d \text{ à } V_{GS} \text{ nul}$	54
Figure III.7.Illustration du phénomène de perçage sur un MOSFET en coupe.....	54
Figure III.8.Coupe schématique d'un VDMOS de puissance.....	55
Figure III.9.Coupe schématique d'un UDMOS de puissance.....	57
Figure III.10.Coupe schématique LDEMOS de type N en configuration standard.....	59
Figure III.11.Coupe schématique d'un LDMOS LOCOS de type N en Configuration Standard.....	59
Figure III.12.Coupe schématique d'un LDMOS resurf de type N en Configuration standard.....	60
Figure III.13.(a) Représentation symbolique, (b) Structure en nid d'abeille.....	61
Figure III.14.Illustration des différents régimes de fonctionnement d'un transistor MOS latéral de puissance.....	62

Figure III.15. Profil du dopage le long de l'interface Si/SiO ₂ caractéristique d'un LDMOS.....	64
Figure III.16. Variation expérimentale de la Résistance à l'état passant d'un transistor LDMOS en fonction de la tension de grille-source.....	65
Figure III.17. Variation expérimentale de $R_{on} *(V_g - V_t)$ en fonction de la tension grille-source du composant pour plusieurs zones d'extension N	65
Figure III.18. Différentes composantes de la résistance à l'état passant du LDMOS.....	66
Figure III.19. Caractéristiques de sortie mesurées d'un LDMOS. Comparaison des courbes affectées (mesure DC) ou non affectées (mesure pulsée) par l'auto échauffement.....	69
Figure III.20. Principe de mesure en impulsion d'un transistor MOSFET.....	70
Figure III.21. Représentation des éléments capacitifs du transistor LDMOS.....	71
Figure III.22. Caractéristique $I_D = f(V_{DS}, V_{GS}, T_j)$	73
Figure III.23. Caractéristique de transfert $I_D = f(V_{GS})$	74
Figure IV.1. Caractéristiques I_d en fonction de V_{ds} pour $T_j = 150^\circ\text{C}$	79
Figure IV.2. Caractéristiques I_d en fonction de V_{ds} pour $T_j = 150^\circ\text{C}$	80
Figure IV.3. Caractéristiques I_d en fonction de V_{ds} pour $T_j = 25^\circ\text{C}$	81
Figure IV.4. Caractéristiques I_d en fonction de V_{ds} pour $T_j = 25^\circ\text{C}$	82
Figure IV.5. Organigramme pendant la phase d'apprentissage.....	83
Figure IV.6. Schémas des différentes couches de neurones de la structure du modèle neuronal.....	84
Figure IV.7. EQM en fonction du nombre d'itérations.....	87
Figure IV.8. Allure d'une Fonction sigmoïde.....	88
Figure IV.9. Fonction de transfert de type linéaire.....	89
Figure IV.10. Organigramme de la phase d'apprentissage	90
Figure IV.11. Organigramme de calcul du courant de drain I_D	91
Figure IV.12. Organigramme de la validation du modèle.....	92
Figure IV.13. Performance du modèle neuronal (RNA)	93
Figure IV.14. Performance du modèle RNA obtenu pour le régime ohmique.....	94
Figure IV.15. La performance du modèle RNA	94
Figure IV.16. Performance du modèle RNA obtenu pour régime ohmique.....	95
Figure IV.17. La performance du modèle RNA	95
Figure IV.18. La performance du modèle RNA obtenu pour régime ohmique.....	96
Figure IV.19. La performance du modèle RNA obtenu	96
Figure IV.20. Modèle RNA pour une augmentation de température T_j de 50°C a 150°C	97
Figure IV.21. Modèle RNA pour une diminution de température T_j de 150°C a 50°C	98
Figure IV.22. Le modèle RNA pour une augmentation rapide de la tension V_{GS} de 3.3V a 4V.....	99
Figure IV.23. Le modèle RNA pour une diminution rapide de la tension V_{GS} de 4V a 3.3V.....	99
Figure IV.24. Structure du modèle neuronal de la conductance $(\frac{\partial I_D}{\partial V_{DS}})$	101
Figure IV.25. Caractéristique $g_{DS} = f(V_{DS})$ pour $V_{GS}=2.5\text{V}$	102
Figure IV.26. Caractéristique $g_{DS} = f(V_{DS})$ pour $V_{GS}=3.5\text{V}$	102
Figure IV.27. Caractéristique $g_{DS} = f(V_{DS})$ pour $V_{GS}=4.5\text{V}$	103
Figure IV.28. Implantation du modèle RNA sur Simulink matlab.....	104

Figure IV.29. Schéma équivalent du modèle RNA sur Simulink matlab.....	104
Figure IV.30. Schéma interne de boîte O1.....	105
Figure IV.31. Caractéristique $I_D = f(V_{DS})$ du modèle neuronal (RNA)	106

Liste des tableaux

Tableau IV.1. Variation de l'erreur du test en fonction des différentes architecture.....	86
--	-----------

TABLE DES MATIERES

Introduction Générale	2
-----------------------------	---

Chapitre I : Etude bibliographique sur les réseaux de neurones

I.1. Introduction	5
I.2. Neurone biologique.....	5
I.3. Neurone formel.....	6
I.4. Fonction d'activation.....	8
I.4.1. Fonction a seuil binaire	8
I.4.1.1. Fonction Heaviside.....	8
I.4.1.2. Fonction Signe.....	8
I.4.2. Fonction linéaire.....	9
I.4.3. Fonction linéaire à seuil ou multi-seuils.....	9
I.4.4. Fonction sigmoïde.....	10
I.4.5. Fonction Gaussienne.....	11
I.5. Choix de la fonction d'activation.....	11
I.6. Structure d'un réseau neuronal artificiel	12
I.6.1. Réseau neuronal multicouche (Feedforward).....	12
I.6.2. Réseau récurrent.....	13
I.6.3. Réseau cellulaire.....	14
I.7. Réseau à fonction de base radiale RBFNN (Radial Basis Function Neural Network)	14
I.8. Architecture adéquate d'un réseau.....	16
I.9. Avantages et Inconvénients des réseaux de neurones.....	18
I.10. Propriétés principales des réseaux de neurones.....	18
I.10.1. Le parallélisme.....	18
I.10.2. La capacité d'adaptation.....	19
I.10.3. La mémoire distribuée.....	19
I.10.4. La capacité de généralisation.....	19
I.11. Conclusion.....	19

Chapitre II: Les Méthodes d'apprentissages

II.1. Introduction.....	21
II.2. L'apprentissage.....	21
II.2.1. Apprentissage non supervisé (sans contrainte)	22
II.2.2. Apprentissage supervisé.....	22
II.2.3. Apprentissage auto-supervisé.....	23
II.3. Les algorithmes d'apprentissage.....	23
II.3.1. Règle de Gross berg.....	24
II.3.2. Règle de Hebb.....	24
II.3.3. Règle de Delta.....	25
II.3.4. La rétro-propagation du gradient de l'erreur.....	26
II.3.4.1. Principe d'algorithme de la rétro propagation du gradient.....	27
II.3.4.2. Algorithme	27
II.3.5. Méthode du gradient conjugué.....	29
II.3.6. Algorithme de BFGS.....	30

II.4. Considérations pratiques liées à l'application du GBP.....	31
II.5. Considérations pratiques.....	32
II.6. Techniques d'accélération de la rétro-propagation.....	32
II.7. Accélération de l'algorithme avec le momentum.....	33
II.8. Le Minimum local.....	34
II.9. Choix du critère à minimiser.....	35
II.10. Avantages et inconvénients de l'optimisation par réseaux de neurones.....	36
II.10.1. Avantages.....	36
II.10.2. Inconvénients.....	36
II.11. Influence Les facteurs d'apprentissage et de lissage.....	36
II.12. Conclusion.....	39

Chapitre III: Etude Bibliographique des Transistors MOSFET de Puissance Appliqués aux Dispositifs Radiofréquences

III.1. Historique.....	41
III.2. Introduction.....	41
III.3. Transistor MOS.....	41
III.3.1. Structure MOS.....	41
III.3.2. Principe et régimes de fonctionnement.....	43
III.3.2.1. Régime linéaire.....	45
III.3.2.2. Régime de saturation.....	47
III.3.3. Fonctionnement des MOSFET en statique.....	50
III.3.4. Fonctionnement des MOSFET en dynamique.....	51
III.3.4.1. Capacités parasites.....	51
III.4. Comportement du transistor MOSFET pour des applications RF de puissance.....	53
III.4.1. Phénomène d'avalanche.....	53
III.4.2. Perçage.....	54
III.5. Transistors MOS de puissance pour des applications RF de puissance.....	55
III.5.1. Transistor VDMOS.....	55
III.5.2. Transistor LDMOS.....	58
III.5.2.1. Le transistor LDMOS LOCOS.....	59
III.5.2.2. Le LDMOS Resurf (Reduced surface field).....	60
III.5.3. Cas particulier du DMOSFET.....	61
III.6. Principe de fonctionnement.....	62
III.7. Régime de fonctionnement statique.....	63
III.7.1. Tension de seuil V_t	63
III.7.2. La résistance à l'état passant (R_{on}).....	64
III.7.3. Régimes de saturation et quasi-saturation.....	66
III.7.3.1. Régime de saturation.....	66
III.7.3.2. Régime de quasi-saturation.....	67
III.7.3.3. Tenue en tension (BV_{ds}).....	68
III.7.4. L'auto-échauffement.....	68
III.8. Le régime de fonctionnement en dynamique.....	70
III.8.1. LDMOS en régime dynamique.....	71
III.9. Avantages des transistors DMOS.....	72
III.10. Influence la température de jonction (TJ) sur le comportement du transistor.....	72

III.10.1.Modification de la caractéristique statique.....	73
III.10.1.1.Caractéristique $I_D = f(V_{DS}, V_{GS})$	73
III.10.1.2.Caractéristique de transfert $I_D = f(V_{GS})$	73
III.11.Conclusion.....	75

Chapitre IV: MODELISATION DES TRANSISTORS FET APPLIQUES AUX DISPOSITIFS RADIOFREQUENCES EN UTILISANT L'APPROCHE NEURONALE

IV.1.Introduction.....	77
IV.2. Modélisation du transistor MOSFET de puissance.....	78
IV.2.1. Choix de la base de données.....	78
IV.2.2. Apprentissage du réseau de neurone.....	82
IV.2.3. Le Choix nombre d'itération.....	87
IV.2.4. Structure du modèle neuronal du courant de drain et normalisation de la base d'apprentissage.....	88
IV.3. Organigramme des programmes élaborés.....	89
IV.3.1. Organigramme de la phase d'apprentissage.....	89
IV.3.2. Programme de calcul du courant I_D du MOSFET IRF7413.....	91
IV.3.3. Programme de validation.....	92
IV.4. Mesure la performance du modèle obtenu.....	93
IV.5. Variation de la température T_j du MOSFET.....	97
IV.6. Effet de la variation rapide de la tension V_{GS}	98
IV.7. Modélisation neuronale de la conductance de sortie du transistor MOSFET de puissance.....	100
IV.8. Modèle neuronal (RNA) du MOSFET de puissance sur simulink Matlab.....	103
IV.9. Conclusion.....	106
Conclusion générale.....	108
Bibliographie.....	110

Introduction générale

Introduction Générale

L'utilisation de simulateurs pour la conception des amplificateurs de puissance est conditionnée par l'existence de modèles non linéaires précis et fiable. Par conséquent un grand intérêt est porté au développement de modèles non linéaires des composants radiofréquences.

Plusieurs techniques de modélisation des transistors MOSFET de puissance sont utilisées actuellement ; Ils diffèrent les uns des autres par le modèle non linéaire qu'ils offrent.

Les modèles physique, le modèle électrique et le modèle de type boîte noire sont limités et présente certains inconvénients : nombre de variable, espace mémoire, temps de calcul...) ; Ainsi pour lever ces limitations on a pensé à une autre technique utilisant l'approche neuronale.

Le but de notre travail est d'utiliser le modèle neuronale pour la modélisation du courant de drain d'un transistor MOSFET de puissance dédié aux applications radiofréquences et montrer sa capacité; Le courant de drain représente le non linéarité la plus importante de ces transistors.

Le travail de ce mémoire est réparti en quatre chapitres :

Dans le premier chapitre on introduit les principales architectures de réseaux de neurones artificiels. Après une présentation de certains modèles, nous avons rappelés les principales caractéristiques et propriétés des réseaux de neurones artificiels ainsi que leur principe de fonctionnement.

Le second chapitre est consacré aux différents types d'apprentissages ; En effet il existe plusieurs méthodes d'apprentissage ou algorithmes connexionnistes. On a défini l'apprentissage non supervisé, l'apprentissage auto-supervisé et l'apprentissage supervisé. Notre travail repose sur la dernière méthode d'apprentissage.

Le troisième chapitre est consacré à une étude bibliographique sur les différentes architectures des transistors MOSFET de puissance afin de mieux cerner les paramètres qui influent sur le réseau de caractéristiques des transistors de puissance.

Dans le quatrième chapitre on montre la capacité des réseaux de neurones artificiels pour la modélisation des transistors MOS de puissance. Cette étude est basée sur la modélisation de la caractéristique courante- tension d'un transistor MOSFET de puissance.

Chapitre I

Etude bibliographique sur les réseaux de neurones

I.1. Introduction

Le but primordial de ce chapitre est l'étude bibliographique sur les réseaux de neurones. L'avantage principal d'un réseau de neurones réside dans sa capacité d'apprentissage des informations. En effet un réseau de neurones peut apprendre, généraliser et modéliser des fonctions non linéaires à plusieurs dimensions.

Tout au long de ce chapitre nous allons chercher à éclairer les concepts généraux des Réseaux de neurones et détailler d'avantage les notions auxquelles nous avons fait appel pour élaborer notre travail.

Dans le but de comprendre le fonctionnement d'un neurone formel on introduit tout d'abord le fonctionnement d'un neurone biologique

I.2. Neurone biologique

Les neurones vivants forment la partie la plus importante du système nerveux central. Ces neurones sont reliés entre eux par des liaisons appelées axones. Ces axones conduisent les signaux électriques de la sortie d'un neurone vers l'entrée (synapse) d'un autre neurone. Les neurones font une sommation des signaux reçus en entrée et en fonction du résultat obtenu vont fournir un courant en sortie. La figure I.1 illustre le fonctionnement d'un neurone biologique.

La structure d'un neurone biologique se compose de trois parties principales [1]:

- **Somma**: Cellule d'activité nerveuse au centre du neurone.
- **Axone**: Il est attaché au somma électriquement actif, ce dernier conduit l'impulsion généré par le neurone.
- **Dendrites** : Elles sont électriquement passives, elles reçoivent les impulsions d'autres neurones.

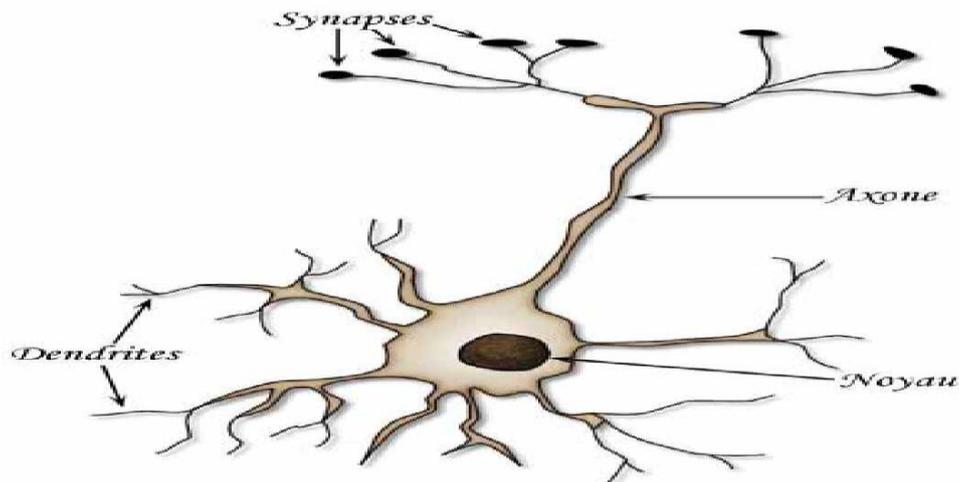


Figure I.1. Schéma fonctionnel d'un neurone biologique [1]

En fonction des signaux qu'il reçoit par l'intermédiaire des dendrites, le neurone décide d'envoyer ou non aux autres neurones un signal se propageant le long de son axone et des axones terminaux.

Si le seuil d'excitabilité du neurone est atteint il y a émission de potentiel d'action dans l'axone. Les synapses permettent la communication entre les dendrites et l'axone. La synapse permettant la connexion entre deux neurones peut être excitatrice ou inhibitrice de l'activité du neurone récepteur.

I.3. Neurone formel

Le neurone formel est un modèle mathématique simplifié du neurone biologique. Généralement on peut considérer un neurone formel comme un automate relié à d'autres automates proches par des liaisons affectées de poids réels. Les signaux d'entrées du neurone formel sont issus des neurones voisins et il fournit à sa sortie un signal dépendant des poids réels et l'unité de sommation

La première modélisation d'un neurone découle des résultats des travaux significatifs de Mac Culloch et Pitts (1943) [2].

La figure I.2 ci-dessous illustre le fonctionnement d'un neurone formel

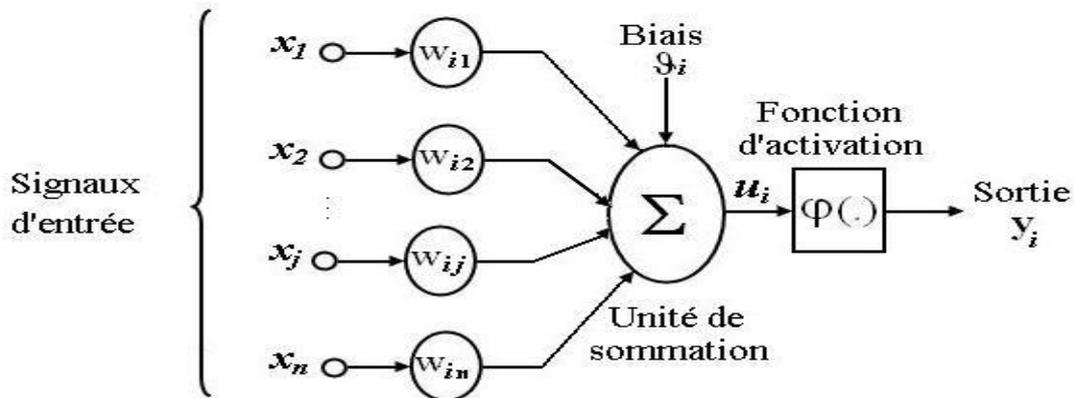


Figure I.2. Modèle de base d'un neurone formel

Le neurone formel est caractérisé par les paramètres définis ci-dessous [2] :

- Les entrées du neurone formel x_i ; $i=1,2,\dots, n$ ou n représente le nombre de neurones de la couche amont.
- Des paramètres de pondération w_{ij} : Poids de la connexion dirigée du neurone j vers le neurone i .
- L'unité de sommation des sorties des neurones de la couche amont qui sont connectées au neurone i joue un rôle primordial car elle représente l'activation du neurone.
- La fonction de transfert ou fonction d'activation a plusieurs formes : fonction à seuil ou multi seuils (Sigmoïde, Heaviside...)
- Une sortie y_i du neurone formel est une fonction pondérée de la somme U_i

La sortie u_k du neurone formel est donnée par la relation (I.1) :

$$U_i = \sum_{j=0}^n W_{ij} * x_i - \vartheta_i \tag{I.1}$$

A partir de cette valeur, une fonction d'activation calcule la valeur de sortie j du neurone. C'est cette valeur qui sera transmise aux autres neurones :

$$y_i = \varphi(u_j) \tag{I.2}$$

I.4. Fonctions d'activation

Cette fonction permet de définir l'état interne du neurone en fonction de son entrée totale et elle représente l'état d'activation du neurone. On peut citer les principales formes de la fonction d'activation [3] :

I.4.1. Fonction à seuil binaire

I.4.1.1. Fonction Heaviside: Elle est définie par :

$$h(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \tag{I.3}$$

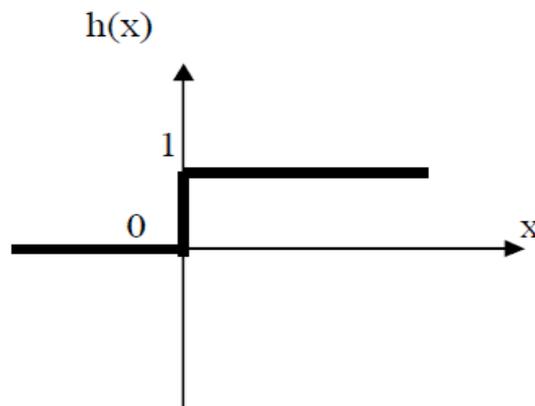


Figure I.3. Fonctions Heaviside

I.4.1.2. Fonction Signe: Elle est définie par

$$h(x) = \begin{cases} +1 & \text{si } x \geq 0 \\ -1 & \text{sinon} \end{cases} \tag{I.4}$$

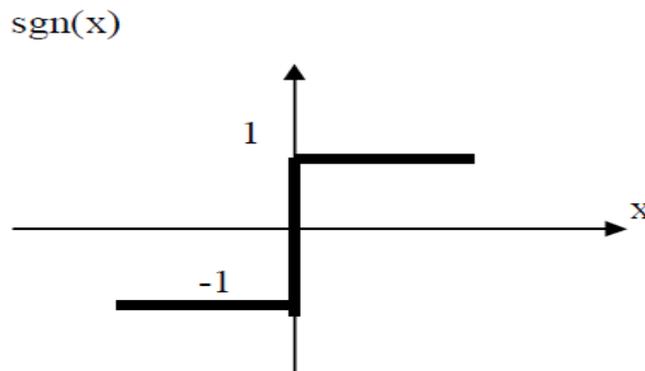


Figure I.4. Fonctions signe

Le seuil introduit une non-linéarité dans le comportement du neurone, cependant il limite la gamme des réponses possibles à deux valeurs.

I.4.2. Fonction linéaire

C'est l'une des fonctions d'activations les plus simples, sa fonction est définie par :

$$F(x)=x \tag{I.5}$$

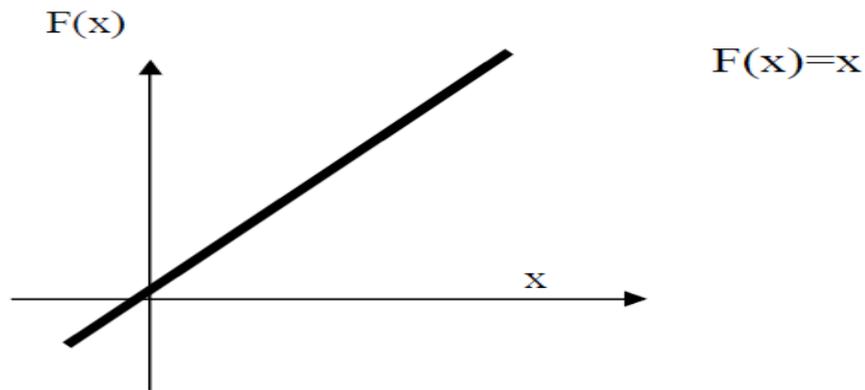


Figure I.5. Fonction linéaire

I.4.3. Fonction linéaire à seuil ou multi-seuils

On peut la définir comme suit :

$$f(x) = \begin{cases} x & x \in [u, v] \\ v & x > v \\ u & x < u \end{cases} \tag{I.6}$$

Cette fonction représente un compromis entre la fonction linéaire et la fonction seuil : entre ses deux barres de saturation, elle confère au neurone une gamme de réponses possibles. En modulant la pente de la linéarité, on affecte la plage de réponse du neurone. Ceci est illustré par la figure I.6.

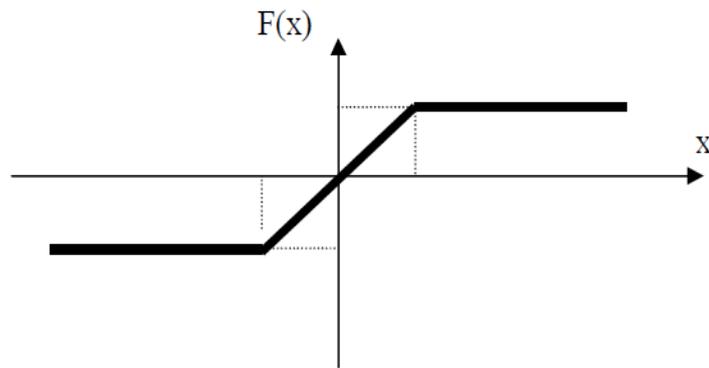


Figure I.6. Fonction linéaire a seuil

I.4.4. Fonction sigmoïde

Elle est l'équivalent continu de la fonction linéaire. Etant continu cette fonction est dérivable, d'autant plus que sa dérivée est simple à calculer [32].

La figure I.7 illustre la fonction sigmoïde :

On a :

$$f(x) = \frac{1}{1+e^{-x}} \tag{I.7}$$

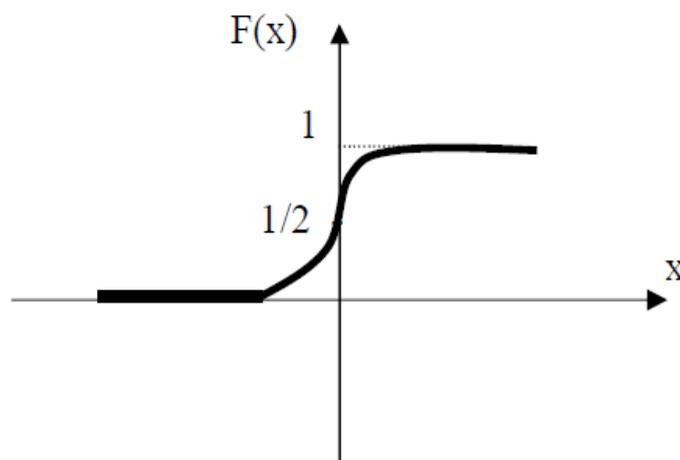


Figure I.7. Fonction sigmoïde

I.4.5. Fonction Gaussienne

Elle est définie par :

$$s = \exp(-a^2/\sigma) \tag{I.8}$$

Et elle est illustrée par la figure I.8

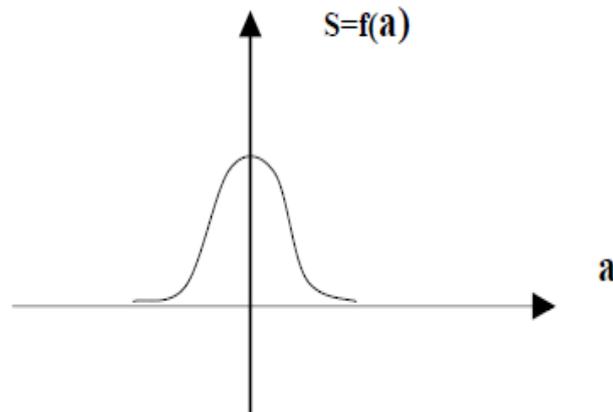


Figure I.8. Fonction gaussienne

I.5. Choix de la fonction d'activation

Le choix de la fonction d'activation dépend de l'application souhaitée. S'il faut avoir des sorties binaires on est orientée habituellement vers la première fonction. Une entrée spéciale est pratiquement toujours introduite pour chaque neurone. Cette entrée, normalement appelée biais (bias en anglais), sert pour déplacer le pas de la fonction d'activation sur l'axe S. La valeur de cette entrée est toujours 1 et le déplacement dépend alors seulement du poids de cette entrée spéciale. Déjà avec un neurone utilisant la fonction d'activation à seuil et une entrée de biais (x_0) il est possible de calculer les fonctions logiques ET et OU comme le montre la figure I.9 [4].

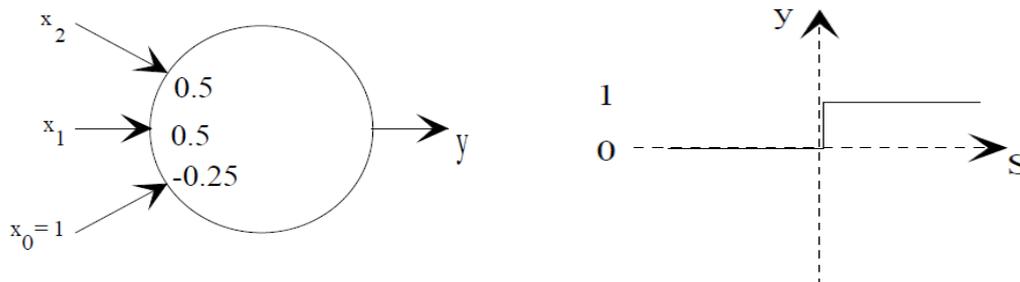


Figure I.9. Fonction logique OU et le neurone correspondant

Si $X_1 = 0$ et $X_2 = 0$, $S = -0.25$ et $f(S) = 0$.

Si $X_1 = 0$ et $X_2 = 1$, $S = 0.25$ et $f(S) = 1$.

Si $X_1 = 1$ et $X_2 = 0$, $S = 0.25$ et $f(S) = 1$.

Si $X_1 = 1$ et $X_2 = 1$, $S = 0.75$ et $f(S) = 1$.

I.6. Structure d'un réseau neuronal artificiel

La structure des connexions entre les différents neurones détermine la topologie du réseau.

En général on peut distinguer trois groupes de réseaux de neurones selon leur topologie :

*Réseau neuronal multicouche

*Réseau récurrent

*Réseau cellulaire

I.6.1. Réseau neuronal multicouche (Feedforward)

La structure de ce réseau peut être simple composé d'une seule couche. La structure d'un réseau monocouche est telle que des neurones organisés en entrée soient entièrement connectés à d'autres neurones organisés en sortie par une couche modifiable de poids. La figure I.10 illustre la structure monocouche [5].

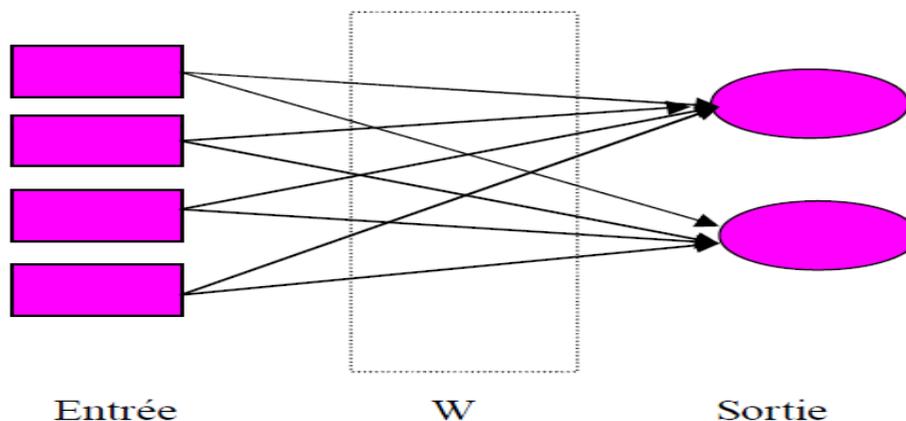


Figure I.10. Réseaux monocouche

Le réseau neuronal multicouche est constitué de plusieurs couches à savoir :

- Une couche d'entrée
- Plusieurs couches cachées (entre les couches d'entrée et de sortie)
- Une couche de sortie

Les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche, et les connexions ne se font qu'avec les neurones de couches avales. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurone d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones d'entrée, couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelées couches cachées [5].

La figure I.11 illustre la structure du perceptron multicouche MLP (multi layer perceptron) et ses traits caractéristiques.

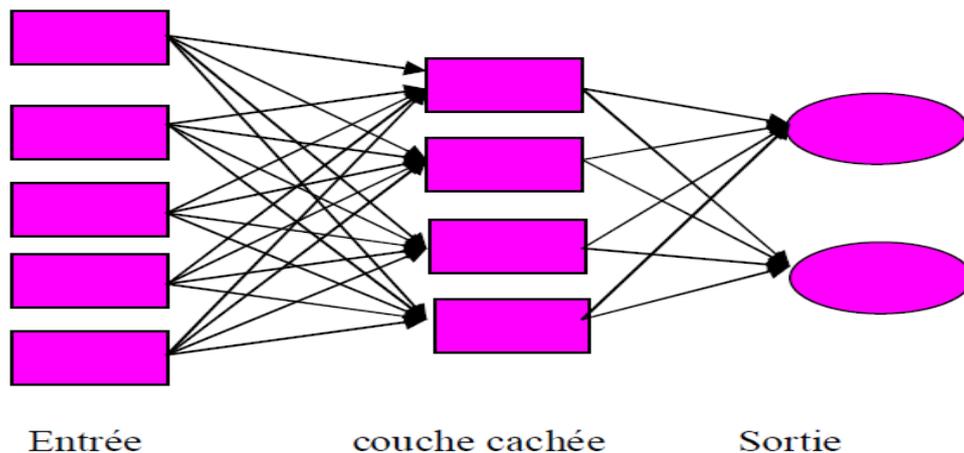


Figure I.11. Réseau Multicouche type feedforward

I.6.2. Réseau récurrent

La structure d'un réseau récurrent est telle que des connexions apparaissent entre la sortie et les neurones des couches amont (il a feedback entre la sortie, les couches cachées et la couche d'entrée). C'est un réseau non linéaire dynamique.

Les réseaux récurrents constituent la deuxième grande catégorie de réseaux de neurones, avec les réseaux type perceptron. Trois classes importantes de réseaux récurrents sont présentées dans ce paragraphe : les réseaux récurrents à couches (le réseau de Elman et le réseau de Jordan), les réseaux à compétition (réseaux ART et réseaux de Kohonen, « Self-organisation mapping ») et les réseaux à connexions symétriques (réseau de Hopfield).

Les réseaux de Hopfield sont des réseaux de neurones totalement interconnectés, récursifs. Il n'y a plus de notions de couche comme dans le perceptron. Les réseaux de neurones récurrents et auto-récurrents permettent d'obtenir des résultats intéressants comparativement aux réseaux à couche. La figure I.12 illustre la structure d'un réseau récurrent symétrique :

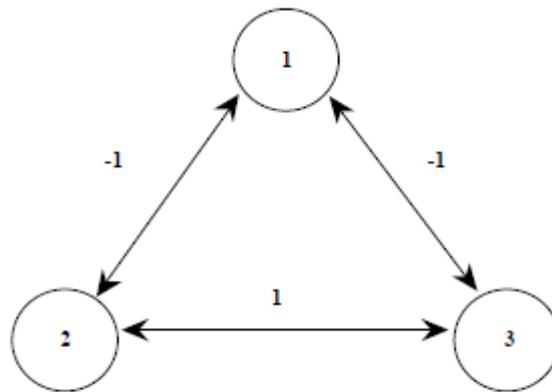


Figure I.12. Réseau récurrent symétrique

I.6.3. Réseau cellulaire

La structure d'un réseau cellulaire est particulière, chaque neurone est relié à tous les neurones qui l'entourent.

Les connexions entre neurones peuvent être dans les deux sens (bidirectionnels).

I.7. Réseau à fonction de base radiale RBFNN (Radial Basis Function Neural Network)

Le réseau à fonction de base radiale RBF (Radial Basis Function) est basé sur une architecture qui s'organise en deux couches seulement, une couche cachée et une couche de sortie comme l'illustre la figure I.12 ci-dessous [3][6].

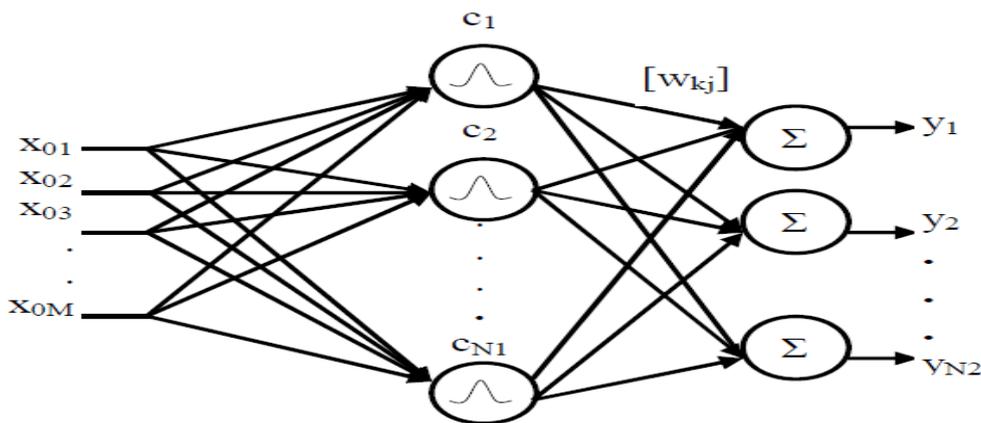


Figure I.13. Architecture du réseau RBFNN

La couche cachée, constituée des noyaux (ou neurones) RBF effectue une transformation non linéaire de l'espace d'entrée. La couche de sortie calcule une combinaison linéaire des sorties de la couche cachée. Chaque noyau élémentaire calcule la distance entre l'entrée et son centre qu'il passe ensuite dans une non linéarité concrétisée par une fonction d'activation $\phi(\cdot)$ qui est généralement de type gaussienne [3][7].

La valeur que prend la sortie du noyau gaussien est d'autant plus importante que l'entrée est plus proche de son centre et tend vers zéro, lorsque la distance entrée centre devient importante. La sortie du réseau RBF est donnée par :

$$y_i = \sum_{k=1}^n W_{kj} \phi_k(\|x - c_k\|) \tag{I.9}$$

$$\phi(\xi) = \exp\left(-\frac{\xi^2}{2N^2}\right) \tag{I.10}$$

$\|\cdot\|$ dénote la norme euclidienne, x le vecteur d'entrée, c_k est le centre associé au noyau k . N_1 le nombre de noyaux de la couche cachée et w_{kj} les poids associés à la couche de sortie. Le paramètre 'η' permet de contrôler la vitesse de décroissance de la fonction ϕ .

L'utilisation habituelle des RBF conserve une fonction d'activation linéaire en sortie mais l'utilisation d'une fonction non linéaire reste possible [7][8].

A l'instar du réseau MLP, le réseau RBF possède aussi la propriété d'approximation c'est à dire que le réseau est capable d'approximer n'importe quelle fonction continue avec une précision donnée pourvu qu'on fournisse un nombre de neurones suffisant.

L'apprentissage des réseaux RBF est composé d'une règle non supervisée pour l'adaptation des centres et une règle d'apprentissage supervisé pour l'adaptation des poids. Le choix de la position des centres et le nombre de neurones reste généralement arbitraire [3].

Une condition nécessaire et suffisante dans un réseau de type RBFNN est que la fonction d'activation ne soit pas un polynôme. La fonction radiale symétrique est généralement utilisée dans ce type de réseau. Elle est souvent gaussienne.

I.8. Architecture adéquate d'un réseau

L'architecture adéquate d'un réseau MLP ou RBFNN est conditionné par la réponse aux points suivants :

- Nombre de couches cachées qu'il faut utiliser(MLP)
- Nombre de neurones dans chaque couche
- connexion entre neurones
- valeurs initiales des poids de connexion (phase d'apprentissage).

La construction d'un réseau multicouche pour la modélisation ou l'identification des fonctions non linéaires dépend des points suivants :

- Choix de la fonction de transfert du neurone caché
- Choix du nombre de neurones cachés et du nombre de couches.
- Initialisation du réseau
- apprentissage du réseau

L'initialisation des paramètres du réseau MLP est généralement faite d'une manière aléatoire et distribués dans un intervalle bien défini.

Le nombre des neurones d'entrée correspond au nombre des variables du vecteur d'entrée utilisée pour résoudre le problème donné. Cependant il n'y a pas une suggestion systématique pour déterminer ce nombre. La sélection de ce paramètre doit être incluse dans le modèle de construction du processus. Idéalement on choisit un petit nombre pour les neurones d'entrée.

La couche cachée et les neurones cachés ont un rôle important pour l'application correcte des réseaux de neurones. Ce sont les neurones de la couche cachée qui prennent le modèle des données, effectuent une mise en correspondance non linéaire complexe entre les variables d'entrée et les variables de sortie. La plupart des scientifiques utilisent dans leurs réseaux de neurones seulement une couche cachée.

Cependant une couche cachée nécessite un nombre très grand des neurones cachés qui n'est pas désirable. Cela provoque un temps d'apprentissage et une capacité de généralisation du réseau très mauvais. En général pour déterminer le nombre des neurones cachés, on réalise des expériences et des essais et parmi eux on choisit le nombre qui provoque la plus petite erreur.

Cette erreur peut être mesurée et est appelée erreur moyenne. Elle est causée par les échantillons d'apprentissage. C'est l'erreur mesurée durant le processus d'apprentissage. Elle est calculée par l'expression suivante [9] :

$$E_{\text{apprentissage}} = \frac{1}{P_{\text{apprentissage}}} * \sum_{p=1}^{P_{\text{apprentissage}}} E^p \quad (\text{I.11})$$

Où $P_{\text{apprentissage}}$ est l'ensemble des échantillons d'apprentissage et E^p est la différence entre la sortie désirée et la sortie obtenue par des échantillons d'apprentissage :

$$E^p = \frac{1}{2} \sum_{o=1}^{N_o} (d_o^p - y_o^p)^2 \quad (\text{I.12})$$

La différence entre la sortie désirée et la sortie obtenue peut être minimisée si l'on a un grand nombre d'échantillons. On peut définir l'erreur moyenne d'essai par l'expression suivante :

$$E_{\text{essai}} = \frac{1}{P_{\text{essai}}} \sum_{p=1}^{P_{\text{test}}} E^p \quad (\text{I.13})$$

L'apprentissage des réseaux fera l'objet du chapitre deux.

Déterminer le nombre des neurones cachés, on réalise des expériences et des essais et parmi eux on choisit le nombre qui provoque la plus petite erreur.

Cette erreur peut être mesurée et est appelée erreur moyenne. Elle est causée par les échantillons d'apprentissage. C'est l'erreur mesurée durant le processus d'apprentissage. Elle est calculée par l'expression suivante :

$$E_{\text{apprentissage}} = \frac{1}{P_{\text{apprentissage}}} * \sum_{p=1}^{P_{\text{apprentissage}}} E^p \quad (\text{I.14})$$

I.9. Avantages et Inconvénients des réseaux de neurones

a- Parmi les avantages des réseaux de neurones on cite :

- Capacité de représenter n'importe quelle fonction, linéaire ou pas, simple ou complexe.
- Faculté d'apprentissage à partir d'exemples représentatifs, par la méthode de rétro propagation des erreurs. L'apprentissage (ou construction du modèle) est automatique,

b- Cependant ils présentent certains inconvénients :

- L'absence de méthode systématique permettant de définir la meilleure topologie du réseau et le nombre de neurones à placer dans la (ou les) couche(s) cachée(s),
- Le choix des valeurs initiales des poids du réseau et le réglage du pas d'apprentissage, qui jouent un rôle important dans la vitesse de convergence,
- La connaissance acquise par un réseau de neurone est codée par les valeurs des poids.

Pour justifier notre choix sur l'utilisation des réseaux de neurones pour la modélisation des fonctions non linéaires on met en évidence les principales propriétés dans le paragraphe qui suit.

I.10. Propriétés principales des réseaux de neurones

L'intérêt porté aujourd'hui aux réseaux de neurones tient sa justification dans les quelques propriétés intéressantes qu'ils possèdent et qui devraient permettre de dépasser les limitations de l'informatique traditionnelle, tant au niveau programmation qu'au niveau machine.

I.10.1. Le parallélisme

Cette notion se situe à la base de l'architecture des réseaux de neurones considérés comme ensemble d'entités élémentaires travaillant simultanément. Par l'étude du fonctionnement des

réseaux de neurones, on pourrait aboutir à des nouvelles techniques de formalisation de problème qui permettraient de les traiter en parallèle.

I.10.2. La capacité d'adaptation

Celle-ci se manifeste par la capacité d'apprentissage qui permet de tenir compte des nouvelles contraintes ou de nouvelles données du monde extérieur. Certains réseaux se caractérisent aussi par leur capacité d'auto organisation qui assure leur stabilité en tant que systèmes dynamiques capables de tenir compte de situations non encore connues.

I.10.3. La mémoire distribuée

Dans les réseaux de neurones, la mémoire correspond à une carte d'activation de neurones. Cette carte est en quelque sorte un codage du fait mémorisé ce qui attribue à ces réseaux l'avantage de résister aux bruits (pannes) car la perte d'un élément ne correspond pas à la perte d'un fait mémorisé.

I.10.4. La capacité de généralisation

Cette capacité est importante surtout dans le cas où la constitution de recueils d'expertise pour un système expert devient difficile (reconnaissance intuitive ou implicite). Les réseaux neuronaux peuvent apprendre à retrouver des règles à partir d'exemple.

I.11. Conclusion

Nous avons présenté dans ce chapitre les principes de base des réseaux de neurones. Nous avons décrit le neurone biologique et ceci dans le but d'étudier et de comprendre le fonctionnement d'un neurone formel.

Nous avons présenté les deux types de réseaux MLP perceptron multicouches et le réseau RBFNN et ce dans le but de modéliser le signal de composant radiofréquence.

Une comparaison entre les deux types de réseaux a été mise en évidence dans ce chapitre.

Chapitre II

Les Méthodes d'apprentissages

II.1. Introduction

L'apprentissage est réellement la propriété la plus intéressante des réseaux de neurones. Au début des années 60 Widrow a réalisé des réseaux capables d'apprentissage ; En effet le réseau perceptron était capable d'apprendre et réaliser certaines fonctions logiques.

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

L'apprentissage neuronal fait appel à des exemples de comportement et la stabilisation d'un tel réseau fait appel à l'algorithme de rétro propagation du gradient qu'on exposera dans la suite du chapitre.

Durant cette phase de fonctionnement, le réseau adapte sa structure (le plus souvent, les poids des connexions) afin de fournir sur ses neurones de sortie les valeurs désirées. Cet apprentissage nécessite des exemples désignés aussi sous l'appellation d'échantillons d'apprentissage associés à un algorithme d'apprentissage.

Dans ce chapitre on présentera tous les algorithmes ou méthodes d'apprentissage pour les réseaux connexionnistes.

II.2. L'apprentissage

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. En effet les paramètres du réseau neuronal subissent des modifications suivant des règles d'apprentissage jusqu'à ce que le réseau se stabilise, c'est-à-dire jusqu'à ce que la valeur souhaitée soit obtenue [9].

Une base de données regroupant toutes les informations est indispensable avant de procéder à l'apprentissage du réseau.

On distingue trois modes d'apprentissage :

- Apprentissage supervisé
- Apprentissage non supervisé
- Apprentissage auto-supervisé

- L'apprentissage est dit off-line si toutes les données sont dans une base d'exemples d'apprentissage et qui sont traités simultanément.
- L'apprentissage est dit on-line si les exemples sont présentés les uns après les autres au fur et à mesure de leurs disponibilités.

II.2.1. Apprentissage non supervisé (sans contrainte)

Dans l'apprentissage non supervisé, seules les valeurs d'entrées sont disponibles et le réseau est laissé libre de converger vers n'importe quel état final. La connaissance à priori de la sortie désirée n'est pas nécessaire, la procédure d'apprentissage est basée uniquement sur les valeurs d'entrées. Le réseau s'auto-organise de façon à optimiser une certaine fonction de coût. Dans ce mode on n'a pas besoin d'une base des sorties désirées et le réseau évolue tout seul jusqu'à la sortie désirée [28].

La figure II.1 illustre l'apprentissage non supervisé.

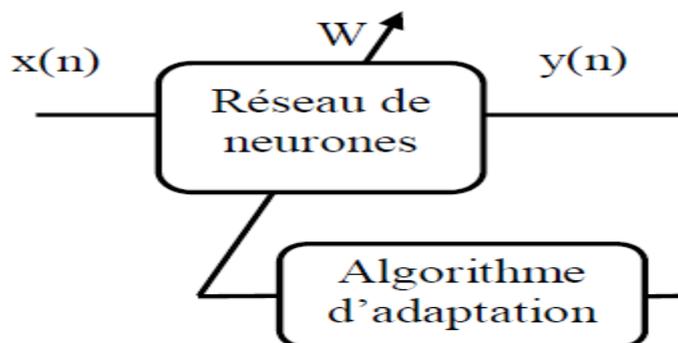


Figure II.1. Mode d'apprentissage des réseaux de neurones non supervisé [10]

II.2.2. Apprentissage supervisé

L'apprentissage est dit supervisé lorsque le réseau est forcé à converger vers un état final précis, ce qui nécessite la connaissance à priori de la réponse désirée $d(n)$. En effet dans ce mode d'apprentissage on fournira au réseau des exemples d'apprentissage qu'il va mémoriser et les comparer avec les valeurs obtenus à la sortie pendant cette phase.

La méthode la plus utilisée est la rétro propagation du gradient. Elle consiste à présenter des exemples au réseau, calculer sa sortie, ajuster les poids de façon à réduire l'écart entre cette sortie et la réponse désirée pour satisfaire un certain critère de performance [29].

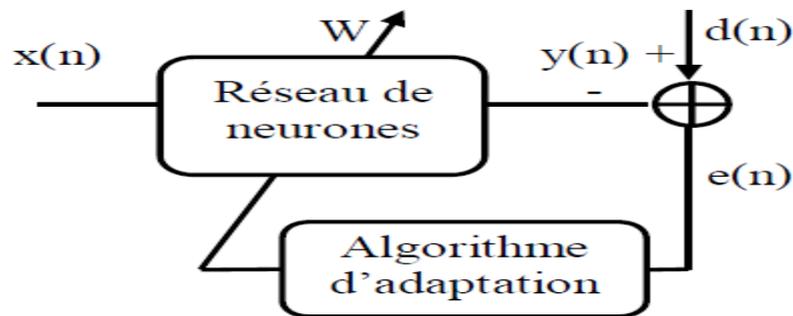


Figure II.2. Mode d'apprentissage des réseaux de neurones supervisé [10]

II.2.3. Apprentissage auto-supervisé

Le réseau de neurone évalue lui-même ses performances, sans aucune aide. Un objet est présenté à l'entrée du réseau de neurones, à qui on a indiqué la classe à laquelle appartient cet objet. Si le réseau ne le classe pas correctement, il mesure lui-même l'erreur qui le fait, et propage cette erreur vers l'entrée. Le réseau procède à autant d'itérations qu'il est nécessaire jusqu'à obtenir la réponse correcte.

Le signal d'erreur représente la différence entre la réponse actuelle du réseau et la sortie désirée à apprendre. En fonction de ce signal d'erreur il y a modification des paramètres du réseau de neurones.

II.3. Les algorithmes d'apprentissage

Dans les systèmes experts, les connaissances de l'expert ont une forme énumérée elles sont exprimées sous forme de règles. Dans le cas des réseaux de neurones, les connaissances ont une forme distribuée : elles sont codées dans les poids des connexions, la topologie du réseau, les fonctions de transfert de chaque neurone, le seuil de ces fonctions et la méthode d'apprentissage utilisée [11].

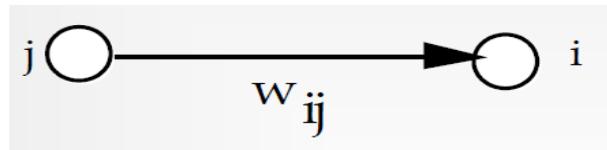
La base d'apprentissage est indispensable dans la phase d'apprentissage mais le grand problème qui doit être posé réside dans les règles d'apprentissage qui permettent aux réseaux de remplir les taches demandées.

On peut citer les règles d'apprentissage suivantes :

II.3.1. Règle de Gross berg

On augmente les poids qui entrent sur l'unité gagnante a_i (neurone receveur) s'ils sont trop faibles, pour les rapprocher du vecteur d'entrée a_j (neurone emetteur). C'est la règle d'apprentissage utilisée dans les cartes auto-organisatrices de Kohonen [33]

$$DW_{ij} = R * a_i * (a_j - W_{ij}) \tag{II.1}$$

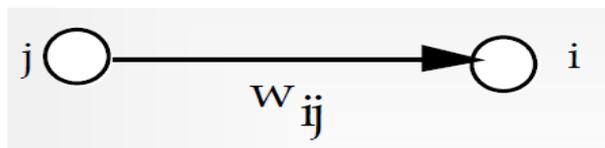


II.3.2. Règle de Hebb

Considérons deux neurones connectés entre eux (l'un émetteur et l'autre récepteur) et si les deux neurones sont actifs au même moment la connexion doit être renforcée. La connexion n'est pas modifiée si l'un ou les deux neurones ne sont pas actifs. Certains travaux ont montré que la connexion qui n'est modifiée est atténuée (règle de Singer et Rauscheker) [30].

	$a_i = -1$	$a_i = 1$
$a_j = -1$	$\Delta W_{ij} = R$	$\Delta W_{ij} = -R$
$a_j = 1$	$\Delta W_{ij} = -R$	$\Delta W_{ij} = R$

$$DW_{ij} = R * a_i * a_j \tag{II.2}$$



II.3.3. Règle de Delta

La règle d'apprentissage de Widrow-Hoff(DELTA) est une règle qui permet d'ajuster les poids d'un réseau de neurones pour diminuer à chaque étape l'erreur commise par le réseau de neurones (à condition que le facteur d'apprentissage soit bien choisi).

Un poids est modifié en utilisant la formule suivante [12][35]:

$$w_{k+1} = w_k + \alpha \delta_k x_k \quad (\text{II.3})$$

Où:

- w_k est le poids à l'instant k ,
- w_{k+1} le poids à l'instant $k+1$,
- α est le facteur d'apprentissage,
- δ_k caractérise la différence entre la sortie attendue et la sortie effective du neurone à l'instant k .
- x_k la valeur de l'entrée avec laquelle le poids w est associé à l'instant k .

Ainsi, si δ_k et x_k sont positifs tous les deux, alors le poids doit être augmenté. La grandeur du changement dépend avant tout de la grandeur de δ_k mais aussi de celle de x_k .

Le coefficient α sert à diminuer les changements pour éviter qu'ils deviennent trop grands, ce qui peut entraîner des oscillations du poids.

Deux versions améliorées de cet apprentissage existent, la version "par lots" et la version "par inertie" (momentum en anglais) [Hecht-Nielsen, 1990], dont l'une utilise plusieurs exemples pour calculer la moyenne des changements requis avant de modifier le poids et l'autre empêche que le changement du poids au moment k ne devienne beaucoup plus grand qu'au moment $k-1$.

II.3.4. La rétro-propagation du gradient de l'erreur

L'algorithme GPB est utilisé pour les réseaux multicouches de type feedforward et il est devenu célèbre grâce aux travaux de Le Cun et Parker.

Il n'y a pas de récursivité dans les connexions, et pas de connexions entre neurones de la même couche. Le principe de la rétro-propagation utilise le mode supervisé c'est à dire on présente au réseau un vecteur d'entrées, de procéder au calcul de la sortie par propagation à travers les couches, de la couche d'entrée vers la couche de sortie en passant par les couches cachées. Cette sortie obtenue est comparée à la sortie désirée, une erreur est alors obtenue. A partir de cette erreur, on calcule le gradient de l'erreur qui est à son tour propagé de la couche de sortie vers la couche d'entrée, d'où le terme de rétro-propagation.

Le principe de la retro propagation de l'erreur consiste à redistribuer l'erreur sur les différentes couches en partant de la couche de sortie ,en passant par les couches cachées et en remontant vers la couche d'entrée. Et ce en respectant la règle d'apprentissage de Hebb. Ceci permet la modification des poids du réseau et donc l'apprentissage. L'opération est réitérée pour chaque vecteur d'entrée et cela jusqu'à ce que le critère d'arrêt soit vérifié [11].

La figure II.3 illustre l'algorithme de retro propagation du gradient

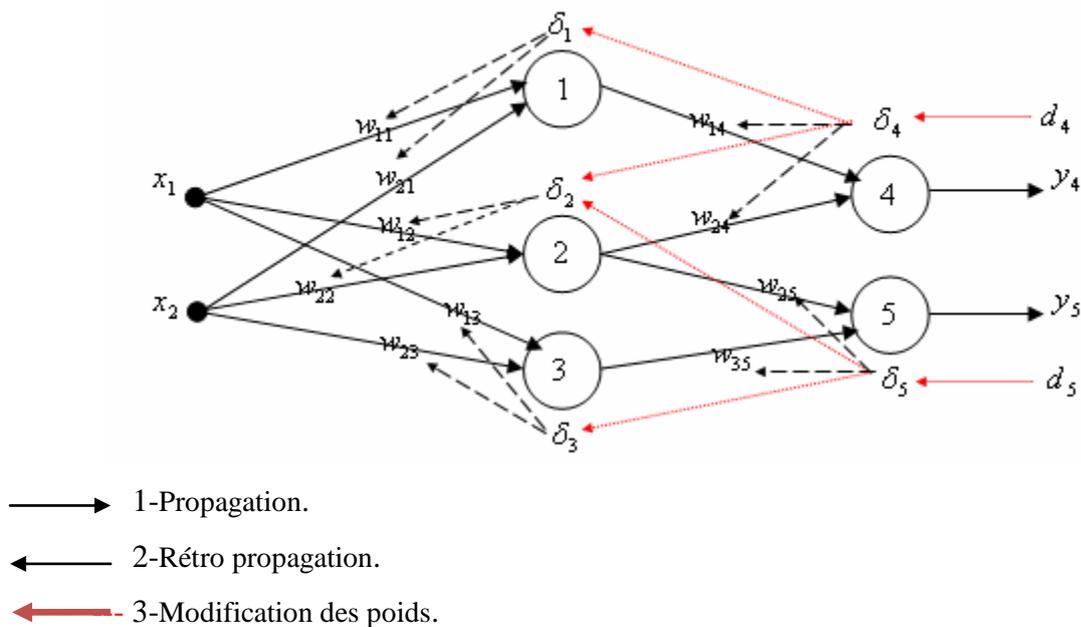


Figure II.3. Algorithme de rétro-propagation du gradient

II.3.4.1. Principe d'algorithme de la rétro propagation du gradient

La rétro-propagation du gradient de l'erreur est utilisée pour ajuster les poids et les biais du réseau afin de minimiser l'erreur quadratique entre la sortie du réseau et la sortie réelle. A chaque couple entrée/sortie, une erreur est calculée, le gradient, ou pente, de l'erreur est déterminé. Ensuite les poids et les biais sont modifiés en ligne sur le réseau. On réitère ces calculs jusqu'à l'obtention du critère d'arrêt [31].

II.3.4.2. Algorithme

L'algorithme de la rétro propagation du gradient de l'erreur se résume aux étapes suivantes :

1. Initialisation des poids w^q à des petites valeurs aléatoires. (Initialiser toutes les connexions et les seuils aléatoirement).

2. Présentation d'une entrée x_k et de la sortie désirée d_k . (Présenter un exemple et une sortie désirée)

3. Calcul de la sortie actuelle par propagation à travers les couches (calculer les sorties de chacune des couches) :

$$y_j^{[q]} = F\left(\sum_i w_{ji}^{[q]} \cdot y_i^{[q-1]}\right) \quad (\text{II.4})$$

Où F représente la fonction de transfert du neurone et $[q]$ la $q^{\text{ième}}$ couche du réseau.

4. Accumulation des erreurs en sortie :

$$\varepsilon = \sum_k (d_k - y_k^{[s]})^2 \quad (\text{II.5})$$

Où d_k est la sortie désirée associée au vecteur d'entrée x_k .

Où $y_k^{[s]}$ est la sortie obtenue sur la dernière couche au temps t .

ε est l'erreur cumulée pour k présentations de couples (x_k, d_k)

5. Rétro-propagation du gradient de l'erreur δ depuis la dernière couche vers la première couche :

- Pour chaque cellule de sortie :

$$\delta_i^{[s]} = - (d_i - y_i^{[s]}) \cdot F' (p_i^{[s]}) \quad (\text{II.6})$$

- Pour chaque cellule cachée :

$$\delta_i^{[q]} = - \sum_k \delta_k^{[q+1]} w_{ki} \cdot F' (p_i^{[q]}) \quad (\text{II.7})$$

6. Mise à jour des poids selon la règle (modifier les poids et les seuils récursivement):

$$\Delta w_{ij}^{[q]} = \mu \cdot (\delta_i^{[q]} \cdot x_j^{[q]}) \quad (\text{II.8})$$

La modification des poids et des biais est obtenue selon les équations suivantes [13]:

$$w_{ijk}(n+1) = w_{ijk}(n) + \mu \delta_{ik} x_{i-1j} \quad (\text{II.9})$$

$$b_{ik}(n+1) = b_{ik}(n) + \mu \delta_{ik} \quad (\text{II.10})$$

Où μ est le coefficient d'apprentissage compris dans l'intervalle [0,1]

Le choix du pas d'apprentissage μ influe beaucoup sur la rapidité de convergence, un pas trop petit ralentit l'apprentissage, un pas trop important provoque un risque d'instabilité.

7. Retour à 2 tant qu'il y a des couple ou exemples à présenter.

La méthode GBP possède beaucoup d'intérêt grâce à sa simplicité et sa capacité de résolution des problèmes complexes et variés. Cependant la phase d'apprentissage est l'inconvénient majeur de cet algorithme. Le choix des coefficients (coefficient d'apprentissage et pas du gradient) pose un problème majeur.

Ils existent différents modes d'apprentissage pour le GBP, on cite :

- Gradient total : On mémorise progressivement les gradients d'erreur obtenus après la présentation du vecteur d'entrée (exemple) ; La somme de ses gradients est utilisée pour modifier les paramètres. La convergence est faible dans ce mode d'apprentissage

- Gradient stochastique : Dans ce mode les paramètres sont modifiés exemple par exemple.
La convergence est plus rapide

II.3.5. Méthode du gradient conjugué(CG)

La méthode du gradient conjugué est une méthode différente des méthodes Newtoniennes.

C'est une méthode intermédiaire entre le GBP et les méthodes de Newton.

Cette méthode ou algorithme est caractérisé par le fait qu'à chaque itération d'apprentissage on modifie le coefficient d'apprentissage et le momentum. La direction de descente est générée à partir d'un modèle quadratique par itération suivant la formule [1]:

$$p_k = -g_{k+1} + \beta_k \cdot p_k \tag{II.11}$$

Qui doit satisfaire la condition du système conjugué suivante

$$p_i^T G p_j = 0 ; i \neq j \tag{II.12}$$

$$\beta_k = -\frac{g_{k+1}^T \cdot G_k p_k}{p_k^T G_k p_k} \tag{II.13}$$

On décline plusieurs reformulations pour β_k qui transforme le Hessien G_k à partir de la dérivée g_k :

- Formule de FETCHER-REEVES

$$\beta_k = \frac{g_{k+1}^T \cdot g_{k+1}}{g_k^T \cdot g_k} \tag{II.14}$$

- Formule de POLAK-RIBIERE :

$$\beta_k = \frac{(g_{k+1} - g_k)^T g_{k+1}}{g_k^T \cdot g_k} \tag{II.15}$$

Dans cet algorithme on suit les étapes suivantes :

- Initialiser tous les paramètres (itération 0)
- Choisir le pas du gradient minimisant la fonction d'optimisation souhaitée.
- Fin si la condition est vérifiée.

II.3.6. Algorithme de BFGS

L'algorithme de BFGS (du nom de ses inventeurs : Broyden, Fletcher, Goldfarb et Shanno) (Press et al. 1988) prend implicitement en compte les dérivées secondes et s'avère donc nettement plus performante que la méthode de rétro-propagation. Le nombre d'itérations est nettement plus faible et le temps de calcul réduit [34].

Le mode d'apprentissage utilisé dans cet algorithme est le mode d'apprentissage off-line.

L'algorithme de **BFGS** est une règle d'ajustement des paramètres qui a l'expression suivante :

$$\theta^K = \theta^{K-1} \pm \mu_K M_K \nabla J(\theta^{K-1}) \quad (\text{II.16})$$

Où M_K est une approximation, calculée itérativement, de l'inverse de la matrice Hessienne.

L'approximation de l'inverse du Hessien est modifiée à chaque itération suivant la règle suivante :

$$M_K = M_{K+1} + \left[1 + \left(\frac{\gamma_{K-1}^T M_{K-1} \gamma_{K-1}}{\delta_{K-1}^T \gamma_{K-1}} \right) \right] \frac{\delta_{K-1}^T \delta_{K-1}}{\delta_{K-1}^T \gamma_{K-1}} \pm \frac{\delta_{K-1} \gamma_{K-1}^T M_{K-1} + M_{K-1} \gamma_{K-1} \delta_{K-1}^T}{\delta_{K-1}^T \gamma_{K-1}} \quad (\text{II.17})$$

Avec:

$$\gamma_{K-1} = \nabla J(\theta^K) \pm \nabla J(\theta^{K-1}) \quad \text{Et} \quad \delta_{K-1} = \theta^K \pm \theta^{K-1} \quad (\text{II.18})$$

Nous prenons pour valeur initiale de M la matrice identité. Si, à une itération, la matrice calculée n'est pas définie positive, elle est réinitialisée à la matrice identité.

Une méthode "quasi newtonienne", n'est efficace que si elle est appliquée au Voisinage d'un minimum. D'autre part, la règle du gradient simple est efficace lorsqu'on est loin du minimum et sa convergence ralentit considérablement lorsque la norme du gradient diminue (c'est à dire lorsqu'on s'approche du minimum). Ces deux techniques sont donc complémentaires. De ce fait, l'optimisation s'effectue en deux étapes: utilisation de la règle du gradient simple pour approcher un minimum, et de l'algorithme de BFGS pour l'atteindre.

L'approximation initiale de l'inverse du hessien doit donc être une matrice symétrique et définie positive. Elle est souvent initialisée par la matrice diagonale unité. Le choix de la matrice hessienne initiale a un grand impact sur les performances de la méthode newtonienne.

Notons que le BFGS présente une convergence rapide pour les réseaux neuronaux de petites tailles et il est nettement meilleur que le GBP ; Par contre pour les réseaux neuronaux ou la taille est importante, le nombre d'informations devient élevée et le nombre de calcul pour chaque itération présente un problème.

Pour résoudre ce problème on a recours à l'algorithme LBFGS à mémoire limitée. L'avantage primordial de cet algorithme réside dans l'accès à la mémoire par l'utilisateur.

Notre choix est fixé sur l'algorithme GBP pour cela on présente dans le paragraphe qui suit les considérations pratiques liées à l'application de cet algorithme

1. Appliquer un vecteur d'entrée $x_p = (x_{p,0,1}, x_{p,0,2}, \dots, x_{p,0,n})^T$ aux nœuds d'entrées puis initialiser les poids du réseau ;
2. Exécuter l'échantillon d'apprentissage à travers le réseau ;
3. Calculer les termes d'erreur de signal de la couche de sortie et les couches cachées en utilisant (4) et (6) respectivement ;
4. Mise à jour les poids de la couche de sortie et couches cachées en utilisant(8), (9) et(10) respectivement ;
5. Répéter ce processus jusqu'à ce que l'erreur **EP** devienne acceptable (**aller à 2**).

II.4. Considérations pratiques liées à l'application du GBP

- Résumé de l'algorithme de rétro-propagation

1. Appliquer un vecteur d'entrée $x_p = (x_{p,0,1}, x_{p,0,2}, \dots, x_{p,0,n})^T$ aux nœuds d'entrées puis initialiser les poids du réseau ;
2. Exécuter l'échantillon d'apprentissage à travers le réseau ;

3. Calculer les termes d'erreur de signal de la couche de sortie et les couches cachées en utilisant (4) et (6) respectivement ;
4. Mise à jour les poids de la couche de sortie et couches cachées en utilisant(8), (9) et(10) respectivement ;
5. Répéter ce processus jusqu'à ce que l'erreur EP devienne acceptable (aller à 2).

II.5. Considérations pratiques

- Les poids du réseau doivent être initialisés à de petites valeurs aléatoires.
- La valeur du taux d'apprentissage μ a un effet significatif sur les performances du réseau, si ce taux est petit l'algorithme converge lentement, par contre s'il est grand L'algorithme risque de générer des oscillations.
- Généralement μ doit être compris entre 0 et 1 pour assurer la convergence de l'algorithme vers une solution optimale.
- Il n'existe pas de règles permettant de déterminer le nombre de couches cachées dans un réseau donné ni le nombre de neurones dans chacune d'elles.
- Théoriquement, l'algorithme ne doit se terminer dès que le minimum de l'erreur commise par le réseau sera atteint, correspondant à un gradient nul, ce qui n'est jamais rencontré en pratique. C'est pourquoi un seuil est fixé a priori afin d'arrêter l'apprentissage.[1]

II.6. Techniques d'accélération de la rétro-propagation

Si bien que l'algorithme de rétro-propagation soit l'algorithme le plus utilisé pour L'apprentissage supervisé des réseaux multicouche voit son implantation se heurter à plusieurs difficultés techniques. Il n'existe pas de méthodes permettant de :

- Trouver une architecture appropriée (nombre de couches cachées, nombre de neurones Cachés et connexions) ;
- Choisir une taille et une qualité adéquate des exemples d'entraînement, choisir des valeurs initiales satisfaisantes pour les poids, et des valeurs convenables pour les paramètres d'apprentissage (facteur d'apprentissage) permettant d'accélération la vitesse de convergence.

- Eviter les effets de l'entraînement (dégradation des performances due à un entraînement prolongé) et la convergence vers un minimum local.

Pour trouver une architecture appropriée (nombre de neurones cachés), Ash a proposé une approche d'addition interactive de neurones dans la quelles on ajoute des neurones aux couches cachées pendant l'apprentissage. Un neurone est ajouté chaque fois que l'erreur se stabilise à un niveau inacceptable.

Cette technique s'appelle la technique de création dynamique de neurones (DNC: Dynamic Node Creation).

Pour éviter le problème des oscillations, beaucoup d'auteurs modifient l'algorithme en lui ajoutent un moment α .

II.7. Accélération de l'algorithme avec le momentum

La convergence du réseau par rétro propagation est un problème crucial car il requiert de nombreuses itérations. Pour pallier à ce problème, un paramètre est souvent rajouté pour accélérer la convergence. Ce paramètre est appelé le momentum.

Le momentum est un moyen efficace pour accélérer l'apprentissage et aussi pour pouvoir sortir des minimums locaux.

La règle de mise à jour des poids devient alors :

$$w_{ijk}(n+1) = w_{ijk}(n) + \mu \delta_{ik} x_{i-1j} + \alpha (w_{ijk}(n) - w_{ijk}(n-1)) \quad (\text{II.19})$$

$$b_{ik}(n+1) = b_{ik}(n) + \mu \delta_{ik} \quad (\text{II.20})$$

α représente le terme d'inertie (momentum).

Il est même possible que l'algorithme rencontre un minimum local. Le terme d'inertie (momentum $0 < \alpha < 1.0$) permet de sortir du minimum local et de limiter les oscillations lors de l'apprentissage ; en prenant en compte les évolutions des étapes précédentes et donc de converger plus rapidement. La phase d'apprentissage est souvent arrêtée lorsque l'erreur calculée sur l'ensemble de la base d'apprentissage est inférieure à un seuil déterminé par l'utilisateur. Il est possible d'arrêter l'apprentissage en fixant une limite au nombre d'itérations.

II.8. Le Minimum local

Le fait que l'apprentissage utilise un principe de descente de gradient sur la surface d'erreur pour modifier les poids, il est donc possible de tomber dans des minima locaux, auquel cas le réseau n'apprendra jamais l'approximation optimale. Heureusement, cela ne semble pas être un problème dans la plupart des cas.

Le RNA est une méthode qui se base sur le calcul de la surface d'erreur. La forme obtenue d'après l' équation de l'erreur est sous forme d'une convexe et en cherchant à minimiser l'erreur, la solution tend vers le minimum.

Le problème qui se pose dans ce cas, c'est qu'il peut y avoir un ou plusieurs minimums locaux induisant ainsi notre réseau en erreur, vu que si on a deux réseaux avec les mêmes paramètres il se peut que l'apprentissage de l'un soit meilleur que celui du second. Des méthodes proposées afin d'éviter les minima locaux :

- Modifier le pas d'apprentissage du réseau pour pousser le réseau hors des minima locaux.
- Réduire les poids du réseau par une petite quantité à chaque pas d'apprentissage.

L'activation d'un neurone sature quand ses liens possèdent de trop grands poids synaptiques. Ce problème est difficile à résoudre lors de l' apprentissage, car les valeurs d'activation extrêmes se traduisent souvent au niveau de la rétro-propagation par de petites corrections. Le réseau se trouve alors dans un minimum local. Réduire systématiquement tous les poids par une petite valeur à chaque correction contribue à garder petits les poids synaptiques, et peut résoudre ce problème.

- Relancer l'apprentissage plusieurs fois en utilisant des poids initiaux différents, ce qui entraîne un temps de calcul plus élevé [21].

La figure II.4 illustre le problème du minimum local

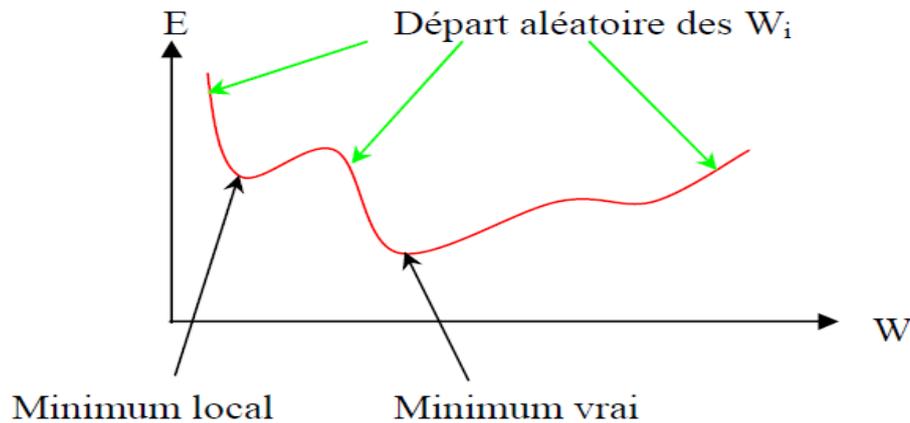


Figure II.4. Courbe explicative du phénomène du minimum local

II. 9. Choix du critère à minimiser

Dans le cas de la rétro-propagation de l'erreur, le critère à minimiser est une erreur Quadratique. L'application de l'algorithme du gradient nécessite la dérivabilité de la fonction de transfert. Le critère de minimisation d'erreur est le suivant [11]:

$$\varepsilon = \sum_k (d_k - y_k^{[s]})^2 \quad (\text{II.21})$$

L'algorithme présenté ici est de type «online», c'est à dire que l'on met à jour les poids pour chaque échantillon d'apprentissage présenté dans le réseau de neurones. Une autre méthode est dite en «batch», c'est-à-dire que l'on calcule d'abord les erreurs pour tous les échantillons sans mettre à jour les poids (on additionne les erreurs) et lorsque l'ensemble des données est passée une fois dans le réseau,

On applique la rétro propagation en utilisant l'erreur totale. Cette façon de faire est préférée pour des raisons de rapidité et de convergence.

II.10. Avantages et inconvénients de l'optimisation par réseaux de neurones

II.10.1. Avantages

- Ce fut un des premiers algorithmes développés pour l'apprentissage des réseaux de neurones multicouches de types feedforward. Il permet de pallier une carence de l'algorithme du perceptron qui est incapable de modifier les poids des couches cachées.
- L'implémentation informatique ne présente pas de difficultés.

II.10.2. Inconvénients

L'algorithme présente certains inconvénients, on cite :

- L'algorithme de rétro-propagation du gradient de l'erreur suit la descente du gradient de l'erreur : un minimum local peut rapidement bloquer la recherche des optima globaux;
- L'algorithme de rétro-propagation est gourmand en temps de calcul;
- Importance du choix du coefficient d'apprentissage, si le coefficient est trop grand la dynamique du réseau va osciller autour de l'optimum, s'il est trop petit, la convergence est lente.

II.11. Influence Des facteurs d'apprentissage et de lissage

Étant donné les équations (9) à (10), nous pouvons dire que la procédure d'apprentissage requiert une modification des poids des connexions

$$w_{ij}^h(t + 1); w_{kj}^0(t + 1); \theta_j^h(t + 1); \theta_k^0(t + 1) \tag{II.22}$$

À l'instant $t+1$, proportionnellement à

$$\frac{\partial CQ(t)}{\partial w_{ij}^h(t)} ; \frac{\partial CQ(t)}{\partial w_{kj}^0(t)} ; \frac{\partial CQ(t)}{\partial \theta_j^h(t)} \text{ et } \frac{\partial CQ(t)}{\partial \theta_k^0(t)} \tag{II.23}$$

Respectivement.

La constante de proportionnalité est le facteur d'apprentissage μ qui est aussi défini comme étant le paramètre d'intensité de l'apprentissage ou le pas du gradient. Si nous adoptons une

rigueur proche de celle des réseaux de neurones biologiques, et sachant que η permet de renforcer ou non la connexion qui relie deux neurones, nous pouvons dire que, en fonction de l'instant t , μ augmente si la connexion est prépondérante et diminue sinon.

Par conséquent, η n'est pas identique pour toutes les connexions et elle est fonction du niveau d'activation, à un instant t , entre deux neurones interconnectés. Ainsi une des difficultés de l'apprentissage neuronal avec descente de gradient est de trouver, à chaque instant t , le facteur d'apprentissage optimal qui permettra une meilleure convergence. Ce problème est le plus souvent résolu par l'implantation d'algorithmes tels que la méthode Levenberg-Marquardt qui utilise le jacobien ou la méthode de Newton qui utilise la matrice hessienne. Pour plus de détails, nous invitons le lecteur à lire Haykin(1994).

En pratique, une autre solution pour résoudre ce problème, est de choisir un seul facteur d'apprentissage mais assez grand de façon à obtenir une courbe de convergence moins lisse, offrant beaucoup de points anguleux (Figure II.4). Une courbe de convergence moins lisse crée le plus souvent des zones de minima locaux. Aussi, on conseille d'introduire dans les équations de modification des poids des connexions (données en début de cette section) un facteur de lissage noté α qui a pour rôle de rendre plus lisse la courbe de convergence et ce facteur de lissage tient compte de la variation des poids entre les instants $t-1$ et t . Ainsi, on a les nouvelles équations suivantes:

$$w_{ki}^0(t+1) = w_{ki}^0(t) + \eta(1 - \alpha) \frac{\partial CQ(t)}{\partial w_{kj}^0(t)} + \alpha \Delta(w_{kj}^0(t)) \quad (\text{II.24})$$

$$\theta_k^0(t+1) = \theta_k^0(t) + \eta(1 - \alpha) \frac{\partial CQ(t)}{\partial \theta_k^0(t)} + \alpha \Delta(\theta_k^0(t)) \quad (\text{II.25})$$

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \eta(1 - \alpha) \frac{\partial CQ(t)}{\partial w_{ji}^h(t)} + \alpha \Delta(w_{ji}^0(t)) \quad (\text{II.26})$$

$$\theta_j^h(t+1) = \theta_j^h(t) + \eta(1 - \alpha) \frac{\partial CQ(t)}{\partial \theta_j^h(t)} + \alpha \Delta(\theta_j^h(t)) \quad (\text{II.27})$$

Avec les variations des poids entre les instants $t-1$ et t données par:

$$\Delta(w_{kj}^0(t)) = w_{kj}^0(t) - w_{kj}^0(t-1) \quad (\text{II.28})$$

$$\Delta(\theta_k^0(t)) = \theta_k^0(t) - \theta_k^0(t-1) \quad (\text{II.29})$$

$$\Delta(w_{ij}^h(t)) = w_{ij}^h(t) - w_{ij}^h(t - 1) \tag{II.30}$$

$$\Delta(\theta_j^h(t)) = \theta_j^h(t) - \theta_j^h(t - 1) \tag{II.31}$$

La Figure II.5 illustre un cas simple de courbes de convergence avec deux Poids de connexions W1 et W2. C'est une figure en trois dimensions (CQ, W1 Et W2), dont CQ est fonction de W1 et W2. Une courbe de convergence est une courbe de CQ(t) en fonction de chaque poids

$$w_{ji}^h(t); w_{kj}^0(t); \theta_j^h(t) \text{ et } \theta_k^0(t) \tag{II.32}$$

Obtenu aux instants 0, 1, 2,..., t, t+1,..., c'est-à-dire qu'à chaque instant t on Calcule le coût et les poids. Nous pouvons remarquer que les trois courbes de convergence

a), b) et c) convergent vers le même point (CQ* , W1* , W2*) mais l'allure des courbes diffère.

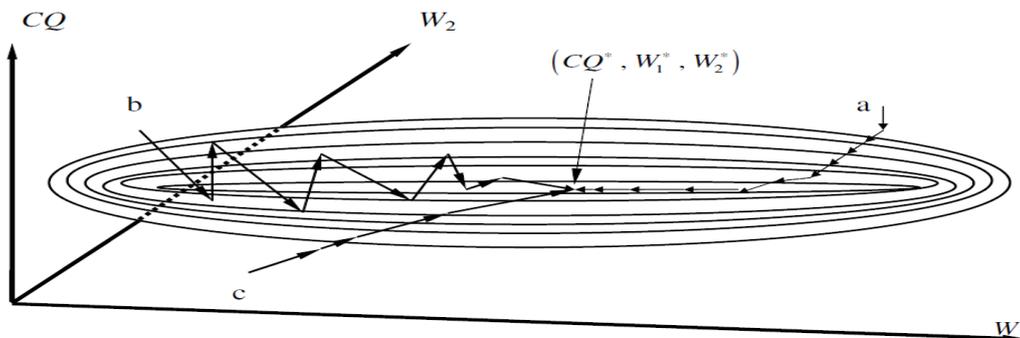


Figure II.5. Descente de gradient dans l'espace des poids:

- a) pour un petit facteur d'apprentissage μ
- b) pour un facteur d'apprentissage μ élevé
- c) Avec un facteur d'apprentissage μ élevé et un facteur de lissage α .

Sachant que chaque longueur de flèche représente un pas de descente de Gradient à un instant t, la courbe a) tend vers le point de convergence à petits pas, les courbes b) et c) tendent aussi vers le point de convergence mais à grands pas et à une allure moins lisse au niveau de la courbe b).

II.12. Conclusion

On a décrit dans ce chapitre les principaux algorithmes d'apprentissage. Le GBP est un algorithme très utilisé et représente un algorithme d'apprentissage pour les réseaux multicouches de type Feedforward ; L'application du GBP a donné des résultats acceptables mais son problème majeur réside dans sa vitesse de convergence très lente.

L'algorithme BFGS a donné de bons résultats dans les applications où la taille du réseau neuronal est réduite

En conclusion on peut dire qu'un réseau de neurones utilisant la GBP peut modéliser des fonctions non linéaires multidimensionnelles.

Chapitre III

Etude Bibliographique des Transistors MOSFET de Puissance Appliqués aux Dispositifs Radiofréquences

III.1. Historique

La théorie sur les transistors à effet de champ (Field Effect Transistor) a été conçue dans les années 1920 – 1930, soit 20 ans avant que le transistor bipolaire (Bipolar Junction Transistor) fut inventé. A cette époque J.E. Lilienfeld (USA) propose un modèle de transistor basé sur le contrôle du courant par l'application d'un champ électrique.

Par manque de matériaux semi-conducteurs appropriés, le développement de ce type de transistor fut très lent. William Shockely proposa un premier transistor JFET en 1952. Les premiers produits industriels firent définitivement leur apparition dans les années 1970.

III.2.Introduction

Le but primordial recherché dans ce chapitre est d'introduire les technologies les plus importantes du MOSFET et son application aux dispositifs radiofréquences(RF).

Le comportement électrique d'un MOSFET de puissance devient peu adapté devant un dispositif RF avec des niveaux de puissance élevés à hautes fréquences et à des niveaux de tension élevés.

Les principaux phénomènes limitant l'utilisation RF (radiofréquence) du MOSFET de puissance seront abordés ainsi que les particularités technologiques des MOSFET dédiés aux dispositifs RF (LDMOS, VDMOS).

Nous exposons en premier lieu le transistor MOS qui représente la base de différentes technologies MOSFET.

III.3.Transistor MOS

III.3.1.Structure MOS

Avant de présenter les équations indispensables permettant le calcul du courant de drain du transistor MOS, il est nécessaire de définir les différentes notations utilisées [14][15].

La figure III.1 illustre le graphe d'une vue schématique d'un transistor MOS type N.

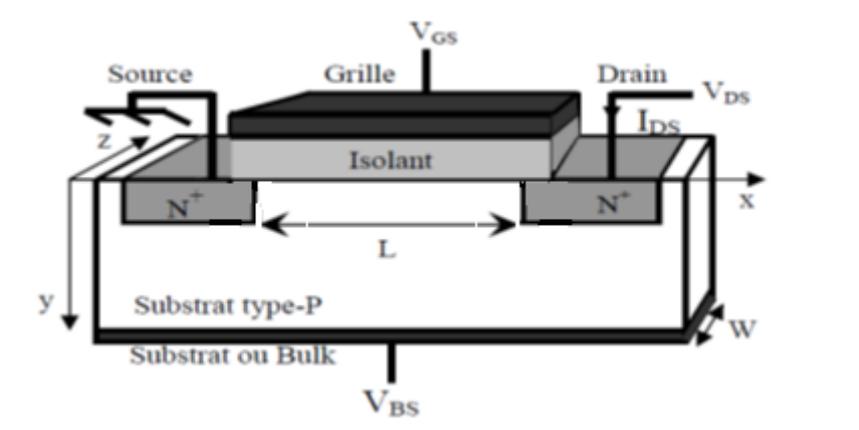


Figure III.1. Vue schématique d'un transistor MOS de type N

Le transistor MOS (ou MOSFET pour transistor Métal-Oxyde-Semi-conducteur à effet de champ) à canal N est un dispositif quadripolaire constitué d'une couche conductrice (métal ou poly silicium fortement dopé) appelée électrode de grille (G), de source (S), de drain (D) et de substrat (B) semi-conducteur sur lequel repose une fine couche d'oxyde isolant (SiO_2) d'épaisseur t_{ox} (Figure III.1).

La longueur du transistor, notée L , correspond à la longueur de sa grille (recouvre légèrement les régions de source et de drain) et sa largeur est notée W . La région entre les jonctions de source et de drain est appelée la région du canal et est définie par sa longueur L et sa largeur W .

On considère par la suite un transistor à canal surfacique, c'est-à-dire dans lequel la conduction est assurée par les porteurs minoritaires du substrat (électrons dans le cas d'un NMOSFET), à l'interface entre le diélectrique de grille et le substrat. Notons que le MOSFET possède deux électrodes supplémentaires par rapport à la capacité MOS, qui sont constituées de deux caissons dopés N^+ pour un NMOS (réservoirs à électrons).

Ainsi, de nombreuses propriétés du transistor MOS découlent de celles de la capacité MOS.

III.3.2.Principe et régimes de fonctionnement

Le principe de fonctionnement du transistor MOS (ou MOSFET) repose sur la modulation d'une densité de porteurs d'une zone semi-conductrice par un champ électrique qui lui est perpendiculaire. Ce champ électrique est appliqué par l'électrode de commande (la grille) à travers un isolant (diélectrique de grille).

Les porteurs créés sont des charges mobiles: électrons dans le cas d'un transistor NMOS, trous dans le cas d'un transistor PMOS. Lorsque la tension appliquée sur la grille V_{GS} est supérieure à une tension seuil appelée tension de seuil, notée V_{th} , ces charges mobiles constituent un canal de conduction entre la source et le drain.

Lorsqu'une différence de potentiel, V_{DS} , est appliquée entre la source et le drain, les porteurs affluant (côté source, de façon conventionnelle) sont collectés par le drain sous la forme d'un courant. Nous parlons de transistor NMOS lorsque le substrat est dopé avec des atomes de type accepteur et que les porteurs minoritaires sont les électrons. Dans ce cas, les régions de source et de drain sont fortement dopées avec des atomes de type donneur (une conduction assurée par des électrons).

Le transistor PMOS possède un substrat dopé avec des atomes de type donneur et les porteurs minoritaires sont les trous. Ainsi, de façon macroscopique, le transistor MOS se comporte comme un dispositif régulant un courant entre deux électrodes par une commande en tension [14].

Rappelons qu'il existe trois valeurs particulières de la tension V_{GS} :

- V_{FB} : tension V_{GS} à appliquer pour que $\Psi_s = 0$ au niveau de la source (aussi appelée tension de bandes plates).
- V_{mg} : tension V_{GS} à appliquer pour que $\Psi_s = \phi_{FF}$ au niveau de la source.
- V_{TH} : tension V_{GS} à appliquer pour que) $\Psi_s = 2\phi_F - \phi_c$ au niveau de la source.

Notons l'apparition de l'écart entre les quasi-niveaux de Fermi, ϕ_c , qui dépendent de la tension V_{DS} . En effet, les zones de drain et de source imposent un écart entre les quasi-niveaux de Fermi des électrons, E_{Fn} et des trous, E_{Fp} aux bornes du canal.

Cet écart, ϕ_c est égal à $(E_{Fp} - E_{Fn})/q$ et prend pour valeur à la source $\phi_c(0) = V_{sb}$ et au drain: $\phi_c(0) = V_{db} - V_{sb}$.

Le substrat étant de type P, le quasi-niveau de Fermi des trous, E_{Fp} est égal au niveau de Fermi dans le volume du semi-conducteur, E_F , et ne varie pas le long du canal, seul le niveau énergétique E_{Fn} varie. La Figure III.2 illustre le diagramme de bandes du transistor MOS [15].

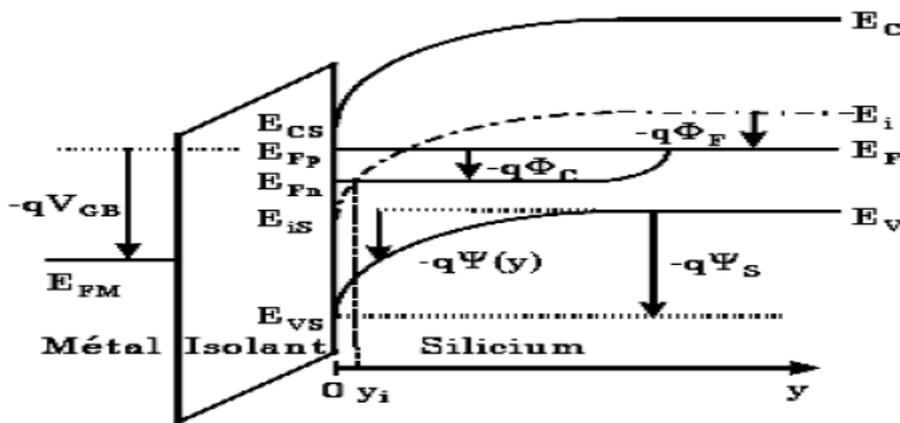


Figure III.2. Diagramme de bandes du transistor MOS en régime d'inversion forte suivant l'axe y en un point quelconque du canal

La courbure des bandes d'énergie du semi-conducteur est notée $\psi(y)$ et la courbure totale correspond au potentiel de surface, Ψ_s .

Le choix du sens des flèches a pour origine la tension que l'on applique entre la grille et le substrat. Cela revient à faire la différence entre les niveaux de Fermi du métal et du semi-conducteur.

Le potentiel de volume du semi-conducteur ϕ_F a pour expression [24]:

$$\phi_F = \frac{Kt}{q} \ln\left(\frac{N_A}{n_i}\right) = \frac{1}{\beta} \ln\left(\frac{N_A}{n_i}\right) = -\frac{1}{q}(E_F - E_i) \tag{III.1}$$

Les paramètres importants du transistor sont les suivants :

– La tension sous le seuil (V_{th}) est la tension de grille nécessaire à la formation d'un canal de conduction (couche d'inversion) entre la source et le drain. La tension de seuil typique d'un NanoMOSFET est comprise entre 0.2 et 0.4 V.

– La transconductance ou gain (g_m) est définie par [23]:

$$g_m = \left(\frac{\partial I_{DS}}{\partial V_{GS}} \right)_{V_{DS}=cte} \quad (III.2)$$

On peut remarquer que la transconductance augmente très rapidement et doit avoir une grande valeur lorsque la longueur de canal devient inférieure à 100 nm.

– La conductance du canal (g_{DS}) donnée par :

$$g_{DS} = \left(\frac{\partial I_{DS}}{\partial V_{DS}} \right)_{V_{GS}=cte} \quad (III.3)$$

– La pente sous le seuil : elle représente la tension de grille à appliquer (en régime sous le seuil) pour augmenter le courant de drain d'une décade [14]:

$$S = \frac{dV_{GS}}{d(\log I_{DS})} \quad (III.4)$$

Le but du paragraphe suivant est l'étude du transistor dans les deux régimes linéaires et de saturation qui sont indispensables à la modélisation

III.3.2.1. Régime linéaire

Le courant total près de l'interface Si-SiO₂ est la somme des courants de diffusion et de conduction des porteurs libres. En inverse, on considère que la concentration de majoritaires est nulle dans le canal et on peut écrire [16]:

$$\vec{j}(x, y) = q(u_n n \vec{e} + D_n \vec{\nabla} n) = \vec{j}_n \quad (III.5)$$

Où : u_n est la mobilité des électrons en champ faible dans la couche d'inversion, D_n le coefficient de diffusion des électrons et $\vec{\nabla} n$ gradient de concentration.

Si on se place dans l'approximation graduelle de Shockley, c'est à dire lorsque le transistor fonctionne en mode non saturé ou encore lorsque le canal n'est pas pincé, on considère que les lignes de champ sont parallèles à l'interface, l'équation (III.5) devient :

$$j_n = \underbrace{qu_n n \varepsilon_x}_{j_1} + \underbrace{qD_n \text{grad}_x n}_{j_2} \quad (\text{III.6})$$

ε_x étant le champ suivant l'axe x (Figure III.2), j_1 est la composante de dérive (due au champ) et j_2 représente le courant de diffusion.

Considérant le coefficient d'Einstein $D_n = u_n KT/q$ où u_n est la mobilité des électrons suppose constante, et sachant que ε_x dérive d'un potentiel scalaire, ce qui se traduit par :

$$\varepsilon_x = -\text{grad}(\phi_c) \quad (\text{III.7})$$

Cette équation peut être simplifiée comme suit

$$J_n = -qu_n n \frac{d\phi_c}{dx} \quad (\text{III.8})$$

Où les relations en potentiel sont données par :

$$\begin{aligned} \phi_c(x = 0) &= -V_{BS} \\ \phi_c(x = L_G) &= V_{DS} - V_{BS} \\ \phi_c(0 < x < L_G) &= V(x) - V_{BS} \end{aligned}$$

Le courant total s'obtient en intégrant la densité J_n sur toute l'épaisseur de la couche d'inversion y_i du transistor de largeur W_G :

$$I_{DS} = -W_G \int_{y_i}^0 J_n dy = W_G \int_{y_i}^0 qu_n n \frac{d\phi_c}{dx} dy = W_G u_n \frac{d\phi_c}{dx} \int_{y_i}^0 n dy \quad (\text{III.9})$$

De plus I_{DS} étant constant tout le long du canal, il vient donc :

$$\int_{L_G}^0 I_{DS} dx = I_{DS} = -W_G u_n \int_{-V_{BS}}^{V_{DS}-V_{BS}} Q_n d\phi_c \quad (\text{III.10})$$

Où
$$Q_n = \int_{y_i}^0 n dy$$

D'autre part l'influence de V_{BS} étant identique en tout point du canal sur les concentrations de porteurs libres, (III.9) devient :

$$I_{DS} = -\frac{u_n W}{L_G} \int_0^{V_{DS}} Q_n(V) dV = -\frac{u_n}{L_G} \int_0^{V_{DS}} (Q_{sc} - Q_D)(V) dV \quad (III.11)$$

Q_{sc} Peut être considéré comme la charge totale du semi-conducteur à laquelle on doit soustraire la charge Q_D de la zone désertée sous la zone d'inversion.

$$Q_{sc} = C_{ox}(V_{GS} - \phi_c - \phi_{ms} + \frac{Q_{ox}}{C_{ox}} - 2\phi_F) \quad (III.12)$$

$$Q_D = -\sqrt{2N_A \epsilon_{si} (\phi_c - \phi_{BS} + 2\phi_F)} \quad (III.13)$$

Pour obtenir l'expression de I_{DS} , il faut maintenant reporter (III.11) et (III.12) dans (III.10), le courant I_{DS} se réécrit alors sous la forme simplifiée :

$$I_{DS} = \frac{W_G}{L_G} u_n C_{ox} V_{DS} \left[V_{GS} - \frac{V_{FB} - 2\phi_F - \gamma(-V_{BS} + 2\phi_F)^{1/2}}{V_{TH}} - V_{DS} \frac{1}{2} \left(1 + \frac{\gamma}{2} (-V_{BS} + 2\phi_F)^{-1/2} \right) \right] \quad (III.14)$$

Ce qui permet d'exprimer le courant de drain d'un transistor MOS en régime linéaire sous sa forme synthétique :

$$I_{DS} = \frac{W_G}{L_G} u_n C_{ox} \left[V_{GS} - V_{TH} - \frac{\alpha}{2} V_{DS} \right] V_{DS} \quad (III.15)$$

III.3.2.2. Régime de saturation

Lorsque la tension de drain augmente, la zone de charge d'espace (ZCE) s'étend et réduit la charge d'inversion à la pointe du drain. Le transistor rentre en mode de saturation lorsque V_{DS} est suffisamment grand pour que $Q_{inv}(x = L_{eff})$ soit quasiment nulle.

Ceci se produit lorsque la tension V_D est égale à V_{DSaT} . On a alors :

$$V_{DSaT} = \frac{V_{GS} - V_T}{\alpha} \quad (III.16)$$

Il existe une autre approche pour déterminer la tension de drain de saturation. Il suffit de considérer que pour $V_{DS} > V_{DSat}$ le courant de drain ne varie pratiquement plus (cas idéal) ce qui revient à résoudre :

$$G_d = \left. \frac{\partial I_{DS}}{\partial V_{DS}} \right|_{V_{DSsat}} \quad (III.17)$$

Ceci permet de retrouver l'expression de I_{DSsat} couramment utilisée :

$$I_{DSsat} = \frac{W_G}{L_G} n C_{ox} \frac{(V_{GS} - V_T)^2}{2\alpha} \quad (III.18)$$

Les Figures III.3a, III.3b, III.4a et III.4b, illustrent clairement les caractéristiques $I_{DS}(V_{GS})$ et $I_{DS}(V_{DS})$ Calculées (avec (III.14) et (III.17)) et mesurées, pour un transistor à canal long ($W_G/L_G = 10/10\mu m$) et un transistor à canal court ($W_G/L_G = 10/0.13\mu m$).

On peut y observer une divergence entre le modèle et la mesure lorsque les tensions de grille et de drain augmentent. Sur les petites géométries, l'écart s'accroît, ce qui met en évidence la nécessité d'affiner les équations en incluant les effets liés à la diminution des dimensions des transistors MOSFET.

Dans la partie suivante, nous allons présenter ces effets, en partant de leurs origines physiques, pour arriver à une quantification mathématique permettant d'inclure leurs impacts sur le comportement électrique des dispositifs.

Nous allons en particulier décrire la réduction de la mobilité des porteurs dans le canal sous l'influence du champ électrique, la modulation de la longueur du canal liée à l'extension de la zone de charge d'espace de la jonction drain-substrat.

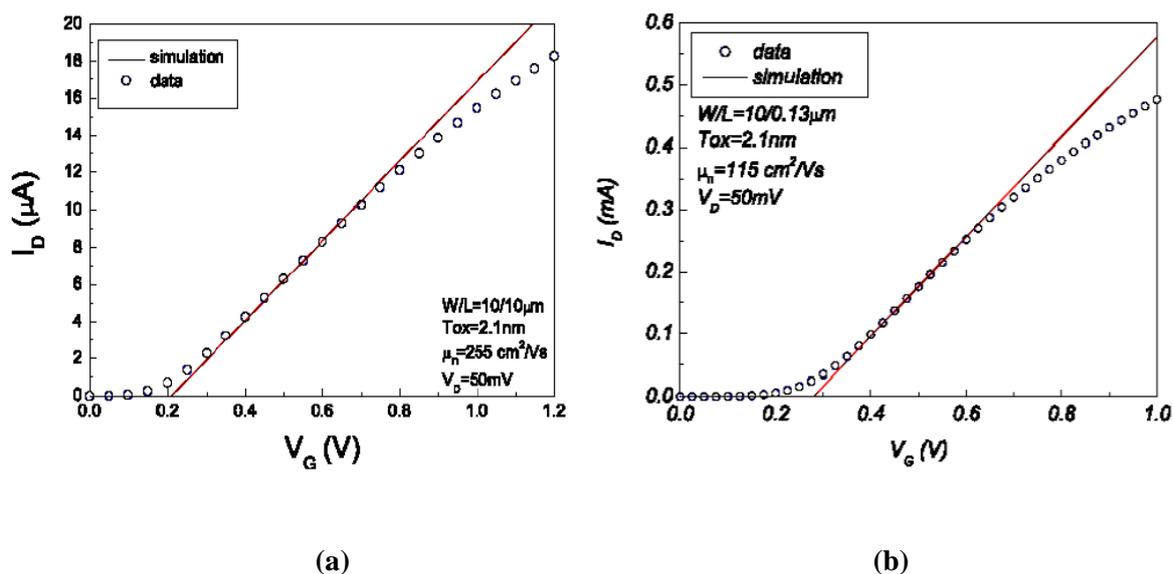


Figure III.3. Comparaison entre la modélisation et la mesure du courant de drain en régime linéaire pour un transistor MOS avec un oxyde ultra mince $T_{ox} = 2.1\text{nm}$
 (a) MOS à canal N long ($W_{eff}/L_{eff} = 10/10\mu\text{m}$) et
 (b) MOS à canal N court ($W_{eff}/L_{eff} = 10/0.13\mu\text{m}$)[27]

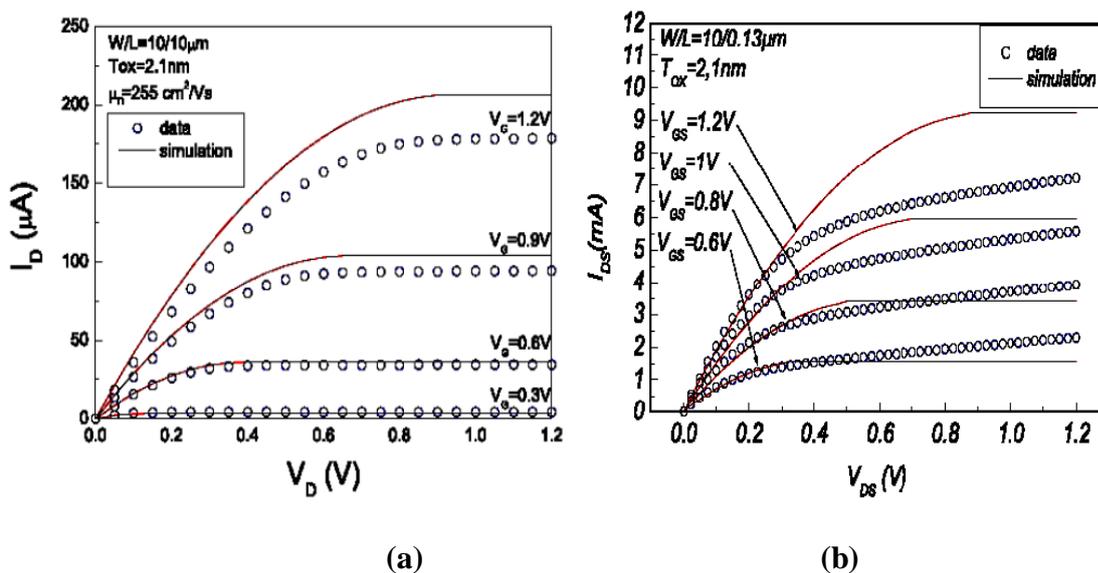


Figure III.4 Comparaison entre la modélisation et la mesure du courant de drain en fonction de la tension de drain pour différentes valeurs de la tension de grille pour un transistor MOS avec un oxyde ultra mince $T_{ox} = 2.1\text{nm}$
 (a) MOS à canal N long ($W_{eff}/L_{eff} = 10/10\mu\text{m}$) et
 (b) MOS à canal N court ($W_{eff}/L_{eff} = 10/0.13\mu\text{m}$)[27]

III.3.3. Fonctionnement du MOSFET en statique

La grille, l'oxyde de grille, la source, le drain, la zone active dont le canal et le substrat composent l'architecture classique d'un MOSFET sur substrat massif, ce qui illustré par la figure III.5-a.

Les potentiels appliqués sur la grille, V_g , et sur le substrat, V_{bDSaT} , modulent le type et la quantité des porteurs présents dans la zone active, voir figure 3.b.

En première approximation, la structure "grille/oxyde, grille/substrat" se comporte comme une capacité. La différence de potentiel entre l'interface "oxyde de grille/substrat" et la zone neutre du substrat s'appelle le potentiel de surface Ψ_s .

Ce potentiel varie suivant les conditions de polarisation de la grille et du substrat mais aussi en fonction du champ électrique longitudinal au canal (potentiels de source et de drain). Le potentiel, sous le canal, varie jusqu'à un seuil fixé dans la partie neutre du substrat.

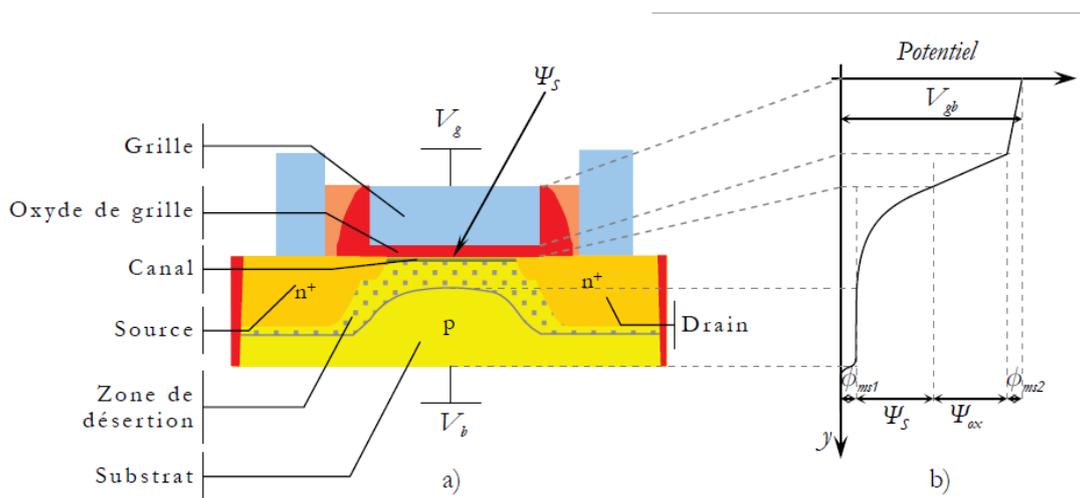


Figure III.5. Le MOSFET à canal n sur substrat massif. a) schéma du transistor;

b) illustration du potentiel distribué dans le transistor

La différence de potentiel est ($V_{gb}=V_g-V_b$), le potentiel Ψ_{OX} correspond à la différence de potentiel entre les deux interfaces de l'oxyde de grille.

Les différences des travaux de sortie au niveau de la grille et du substrat sont représentées par ϕ_{ms1} et ϕ_{ms2} on a :

$$\phi_{ms} = f\phi_{ms1} + f\phi_{ms2}$$

Dans la suite de ce paragraphe, l'étude de cette structure de type Métal Isolant Semi-conducteur ou MIS est approfondie.

III.3.4. Fonctionnement des MOSFET en dynamique

III.3.4.1. Capacités parasites

Si l'on considère la Figure III.5, on peut voir que la grille forme des capacités avec les différentes zones du MOSFET : entre grille et métallisation de source (C_{oxm}), entre grille et zone N^+ de source (C_{oxN^+}), entre grille et zone P du canal (C_{oxp}) et entre grille et drain (C_{GD}) [26].

De la même manière, lorsqu'une zone de charge d'espace se développe suivant la jonction P-N de la diode intrinsèque, une capacité C_{DS} apparaît, cette fois entre le drain et la source du transistor. C_{oxm} , C_{oxp} et C_{oxN^+} peuvent être considérées comme des capacités constantes si l'on suppose que les zones de charge d'espace qui peuvent apparaître dans les régions N^+ et P ont une épaisseur faible. Leur expression est donnée par :

$$C_{OXI} = A_i \frac{\epsilon_{ox}}{e_{ox}} \tag{III.19}$$

Où A_i représente la surface effective des deux armatures, ϵ_{ox} la permittivité de l'oxyde et e_{ox} l'épaisseur de la couche d'oxyde (supposée constante tout autour de la grille).

La capacité C_{GD} a un comportement un peu plus complexe : lorsque le transistor est bloqué, une zone de charge d'espace se développe dans la région v sous l'effet d'une polarisation drain-source positive, jusqu'à l'occuper toute entière [17].

En l'absence de cette polarisation, C_{GD} s'exprime par l'équation (III.20). Elle va montrer une forte sensibilité à V_{GD} . On peut la modéliser par deux capacités en série : l'une, variable, causée par l'extension de la zone de charge d'espace (C_{GDv}) ; l'autre, constante, formée par la couche d'oxyde (C_{oxD}).

$$C_{GD} = \frac{C_{OXD} * C_{GDJ}}{C_{OXD} + C_{GDJ}} \quad (III.20)$$

Selon [1], l'épaisseur de la zone de charge d'espace qui forme C_{GDJ} est donnée par :

$$e_{GDJ} = \sqrt{\frac{2\epsilon_{OX}V_{GD}}{q^n}} \quad (III.21)$$

Soit, en remplaçant e_{OX} et ϵ_{OX} par e_{GDJ} il vient::

$$C_{GDJ} = A_{GD} \sqrt{\frac{q\epsilon_{OX}N}{2V_{GD}}} \quad (III.22)$$

Elle est valable pour les tensions V_{GD} positives, en dessous, C_{GD} vaut C_{OXD} . La capacité C_{DS} est également causée par une zone de charge d'espace, qui apparaît lorsque la jonction P+-N- est polarisée en inverse.

Une formulation plus poussée de cette capacité sera proposée dans la section sur la modélisation de la diode intrinsèque. Dans ce qui nous intéresse, nous définirons trois capacités :

– C_{GS} Formée des trois capacités constantes C_{OXM} , C_{OXP} et C_{OXN} ; C_{GD} et C_{DS} . D'autre part, pour des raisons de facilité de mesure, on décrit parfois ces capacités sous la forme de C_{ISS} (capacité d'entrée, drain et source court-circuités), C_{RSS} (capacité de contre-réaction) et C_{OSS} (capacité de sortie, grille et source court circuitées) définies, en petits signaux, par [18]:

$$C_{ISS} = C_{GD} + C_{GS} \quad (III.23)$$

$$C_{RSS} = C_{GD} \quad (III.24)$$

$$C_{OSS} = C_{GD} + C_{DS} \quad (III.25)$$

Pour faire commuter un MOSFET, il faut donc charger ou décharger C_{ISS} , de manière à se placer sur le point de fonctionnement désiré de la caractéristique statique. Ces capacités parasites peuvent avoir un autre effet durant les commutations: C_{GS} et C_{GD} forment un diviseur capacitif qui peut venir charger la grille du transistor lors des commutations.

L'application d'un gradient de tension entre drain et source peut, via le diviseur de tension capacitif, venir modifier V_{GS} et faire commuter le MOSFET de façon incontrôlée [17].

III.4. Comportement du transistor MOSFET pour des applications RF de puissance

Les équations données précédemment correspondent au fonctionnement idéal du transistor MOS qui sont basées sur de nombreuses approximations. Les MOSFET dédiés aux applications de puissance ont leurs performances limitées par les phénomènes d'avalanche (jonction substrat-drain) et de perçage source-drain. Le paragraphe suivant expose ces différents phénomènes spécifiques aux MOS de puissance [20][19].

III.4.1. Phénomène d'avalanche

L'avalanche dans un transistor MOSFET a lieu dans la diode formée entre le drain et le substrat illustrée par le graphe de la figure III.6.a.

Elle se produit dans la zone désertée du substrat car celle-ci est plus importante que celle du drain en raison d'un dopage moindre avec un champ électrique maximal localisé en surface de l'interface P- N côté drain représentée dans la figure III.6.b.

C'est dans cette zone et à mesure que la tension appliquée V_{ds} augmente que le phénomène de multiplication des porteurs apparaît jusqu'à ce que se déclenche l'avalanche. La tension de drain correspondant à l'avalanche est appelée tension d'avalanche BV (breakdown voltage).

Plus le canal du transistor diminue plus le champ électrique sur le bord du canal, côté drain, sera très élevé pour des tensions relativement faibles. Par conséquent, le champ électrique qui déclenche l'avalanche est atteint pour de faible valeur de V_{ds} phénomènes spécifiques aux MOS de puissance [20].

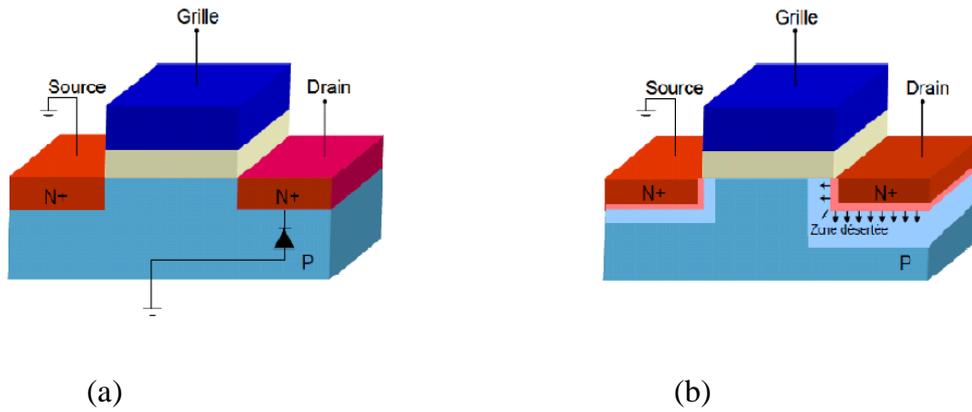


Figure III.6.a) Représentation schématique du MOS en mode d'avalanche (jonction)

b) Illustration du MOS lors de l'application d'une tension V_d à V_{GS} nul

III.4.2.Perçage

Un autre phénomène qui peut apparaître dans les transistors à faible longueur de canal. A mesure que la tension de drain augmente, la zone de charge d'espace cote drain ne cesse d'augmenter jusqu'à rencontre la zone désertée cote source ; Ce phénomène est illustré est dans la figure III.7.

Des valeurs élevées de V_{ds} peuvent même entraîner une jonction entre les deux zones : c'est le phénomène du perçage. Le comportement du transistor est complémentaire dégradé et si la longueur de grille est très faible, une injection de porteurs peut avoir lieu entre la source et le drain.

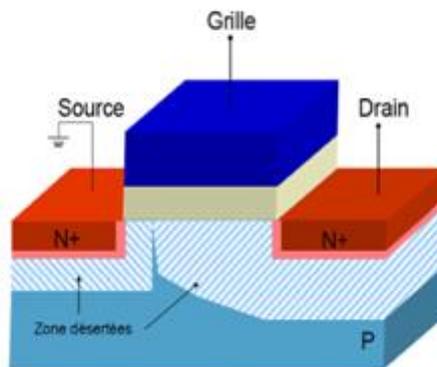


Figure III.7. Illustration du phénomène de perçage sur un MOSFET en coupe

Les phénomènes d'avalanche et de perçage sont des facteurs limitant la tension maximale de drain du dispositif. Cependant les transistors MOSFET de puissance nécessitent à la fois de faibles longueurs de canal et des valeurs de tension de drain élevées.

Il existe des solutions d'ordre technologique permettant d'éviter l'apparition prématurée de ces phénomènes. Ces solutions sont exposées dans le paragraphe suivant [20].

III.5. Transistors MOS de puissance pour des applications RF de puissance

Des structures ont été élaborées afin de pouvoir réaliser des transistors MOS de puissance tout en remédiant à l'apparition des phénomènes d'avalanche et de perçage. Deux grandes familles de transistor MOS de puissance sont référencées : les MOSFET de puissance verticale (VDMOS) et les MOSFET de puissance latérale (LDMOS).

III.5.1. Transistor VDMOS

Les premiers transistors de puissance développés au début des années 70 étaient des VMOS illustré par la figure III.8.

Ces transistors étaient élaborés à partir de sillons en forme de V graves par attaque anisotrope du silicium. Elle consiste à attaquer chimiquement et suivant les plans, une couche N- épitaxiée sur un substrat N+. La zone active est formée le long d'un sillon qui fait un angle de 54° avec la surface [19].

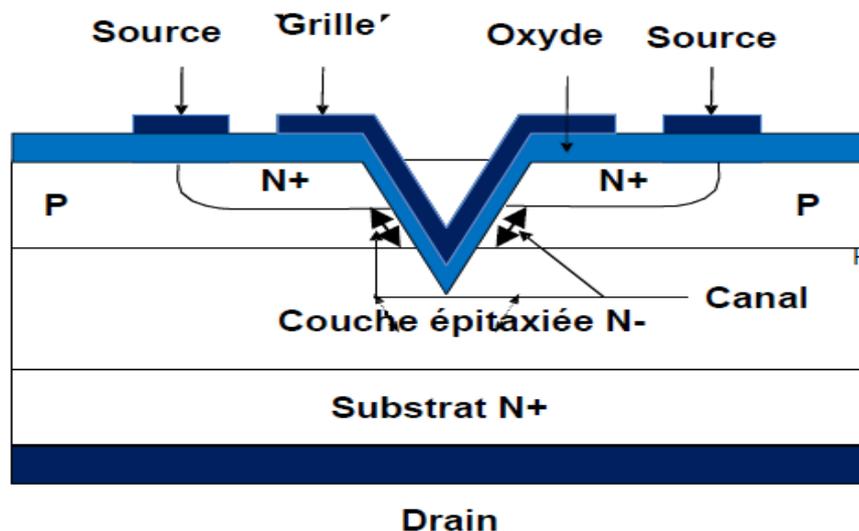


Figure III.8. Coupe schématique d'un VDMOS de puissance

Par rapport au transistor MOS plan classique, le progrès essentiel réside dans l'utilisation du volume du semi-conducteur pour former l'électrode de drain (plan inférieur du substrat).

Cette disposition permet une plus grande densité d'intégration puisque seulement deux électrodes, électrodes de grille et de source sont localisées sur la face supérieure de la puce.

À l'état passant, la tension grille-source V_{GS} est supérieure à la tension de seuil V_{TH} et un canal d'inversion de type N est ainsi formé en surface de la zone du canal permettant le passage du courant entre le drain et la source. Ce courant passe, en grande partie, dans le volume de la structure à travers la zone de drift N-. Une partie de ce courant passe en surface dans le canal et dépend de la mobilité des électrons dans la couche inversée [19].

La résistance du canal d'inversion est une limitation supplémentaire dans le cas des transistors VDMOS basse tension. La zone de drift N- assure au transistor VDMOS la faculté de bloquer la tension à l'état bloqué. Pour une structure VDMOS bien optimisée, la tension de claquage dépend du dopage et de l'épaisseur de la zone de drift. En effet, pour bloquer des tensions très élevées, la zone de drift doit être faiblement dopée et suffisamment large pour permettre à la zone de charge d'espace de s'étendre. Par conséquent, cette zone large et faiblement dopée devient, à l'état passant, une résistance très grande que l'on ne peut réduire que par augmentation de la surface active du composant ou par utilisation de nouveaux concepts comme le concept de la super jonction par exemple. L'utilisation de matériaux de forte énergie d'ionisation peut aussi être une très bonne solution à ce problème.

Il existe donc un compromis entre la résistance à l'état passant et la tension de claquage des transistors VDMOS de puissance. On peut remarquer aussi que l'électrode de grille joue ici le rôle d'une plaque de champ et réduit le champ électrique à la surface de la jonction "P-body/N- drift". Ceci n'est pas sans conséquence car cette électrode de grille étalée sur toute la surface intercellulaire entre les diffusions P source introduit une capacité parasite MOS entre la grille et le drain. Cette capacité, connue sous le nom de capacité Miller, cause une contre réaction entre la sortie et l'entrée du composant et réduit considérablement la fréquence de transition du transistor. Pour remédier à ce problème, on peut par exemple éliminer une partie du grill au-dessus de la couche épitaxie N-. Dans ce cas, une augmentation significative de la fréquence de transition du transistor peut être obtenue. Toutefois, cette élimination d'une partie de la grille provoque une réduction de la tension de claquage du

dispositif, à cause du fort champ électrique à la fin de métallisation de grille, et une augmentation de la résistance à l'état passant car la longueur de la zone accumulée est réduite.

D'autre part, la résistance équivalente du transistor VDMOS est composée de plusieurs résistances en série : résistance du canal, résistance accumulée à la surface de la zone intercellulaire entre les diffusions P source adjacentes, résistance JFET (Junction Field Effect Transistor) de la région entre les diffusions P source adjacentes, résistance de la zone volumique de drift N-, résistance de substrat et résistances de contact. L'effet résistif de la zone de JFET entre les diffusions P source est d'autant plus grand que le dopage de la zone épitaxie est faible, cas des transistors prévus pour fonctionner en hautes tensions, ou si la distance entre les diffusions P source est très faible. C'est ce dernier phénomène, augmentation de la résistance JFET en diminuant la distance intercellulaire, qui limite la densité d'intégration des composants MOS de puissance.

Le transistor UMOS a été proposé vers la fin de l'année 1970. Sa structure est illustrée par la figure III.10 ; Elle reprend celle du VMOS mais la forme du sillon est différente. Le front de pénétration de l'attaque anisotrope est arrêté prématurément, créant ainsi un fond plat dans le sillon qui permet de diminuer le champ électrique. Les mêmes difficultés technologiques ayant été rencontrées pour ce composant, il n'eut pas de développement industriel significatif [15].

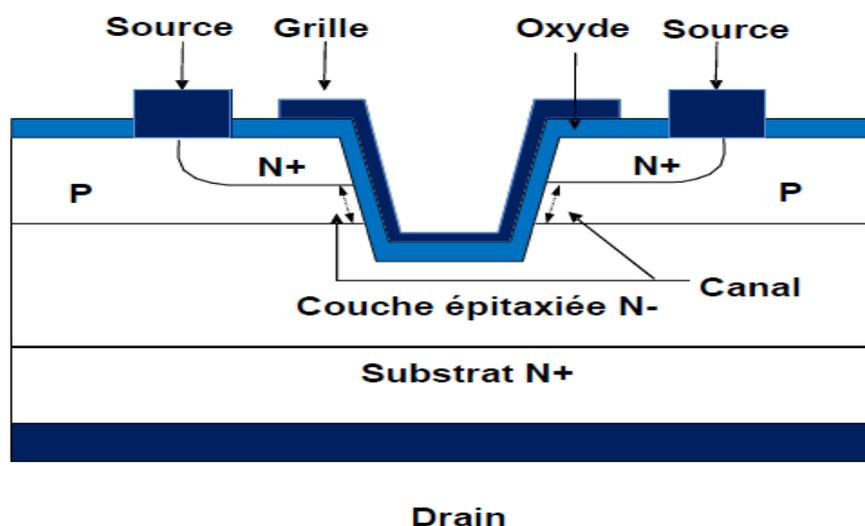


Figure III.9. Coupe schématique d'un UDMOS de puissance

III.5.2. Transistor LDMOS

Ces structures de puissances sont réalisées comme pour le VDMOS par les procédés de double-diffusion et d'auto-alignement sur grille en poly-silicium. Le transistor LDMOS possède l'avantage par rapport au transistor à configuration verticale VDMOS d'avoir intrinsèquement ses trois électrodes coplanaires, ce qui le rend favorable à l'intégration.

Il est également compatible avec les technologies BiCMOS (contraction de Bipolar-CMOS) avancées, ce qui fait de lui un candidat idéal pour assurer la partie puissance (c'est-à-dire interrupteur) dans un circuit intégré. Dans ces technologies les dispositifs de puissance sont des structures latérales LDMOS, c'est-à-dire que le courant traverse la structure par le drain situé à la surface de la puce, traverse le canal, puis la source elle aussi située à la surface de la puce comme le représente la figure III.10.

La tension de claquage est limitée dans cette structure à des tensions de l'ordre de 250 Volts. En effet, le claquage dans cette structure se produit généralement à la fin de la métallisation de grille (à cause de la faible épaisseur de l'oxyde de grille) ou à la Junction cylindrique PN-. Il dépend principalement du dopage de la zone de drift et de la distance L_d entre la fin de la métallisation de grille et le début de l'ouverture de la diffusion de drain.

La couche enterrée N+ permet de limiter l'extension de la zone de charge d'espace dans la région N- et d'éviter le perçage du substrat P. Elle peut permettre aussi d'annuler le gain du transistor parasite PN-P entre le substrat et la diffusion P source du transistor MOS de puissance. Le courant électrique passe en surface de la structure et dépend de la distance L_d ; cette distance est appelée aussi longueur de la zone de drift.

Plus cette distance est faible et plus la résistance à l'état passant du composant est faible. Malheureusement, la tension de claquage diminue quand L_d diminue. Il existe une distance de drift optimale assurant une tension de claquage maximale pour un dopage de drift donné. Ceci limite donc la densité d'intégration car le choix de la distance de drift, qui se trouve en surface de la structure, sera imposé par le compromis entre la tension de claquage et la résistance passante spécifique. De plus, la présence des trois électrodes sur la face supérieure de la puce ne facilite pas la densification des cellules mises en parallèle pour former le transistor LDMOS de puissance. La source et le drain se présentent le plus souvent sous forme de bandes parallèles inter digitées [25].

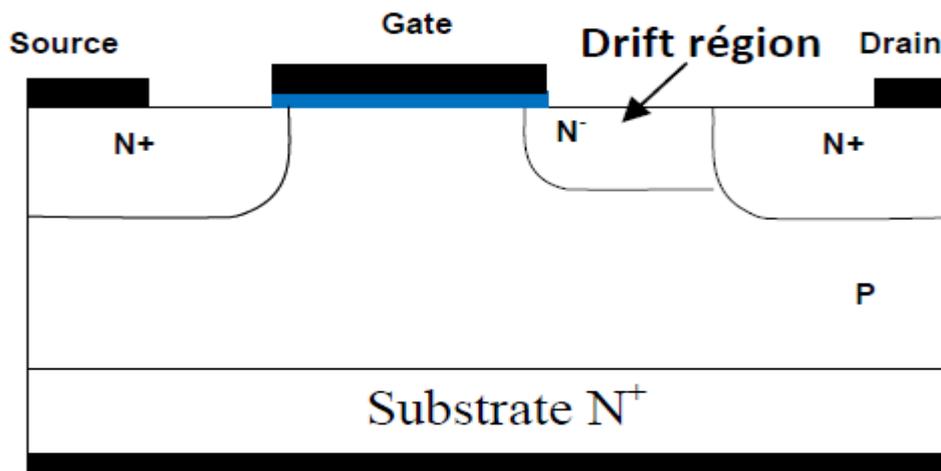


Figure III.10. Coupe schématique d'un LDEMOS de type N en configuration standard

De nombreuses variantes de la structure du transistor LDMOS ont été proposées pour améliorer sa tension de claquage ; on peut citer par exemple [15] :

III.5.2.1. Le transistor LDMOS LOCOS : (LOCAlOxidation in Silicon)

Dans cette structure illustrée par la figure III.11, le champ électrique à la fin de La métallisation de grille est fortement réduit grâce à une oxydation locale du silicium. Le problème majeur de cette structure est la dégradation de la surface de passage du courant à l'état passant et, par conséquent, une augmentation significative de la résistance passante spécifique comparée à celle des structures LDMOS classiques [19].

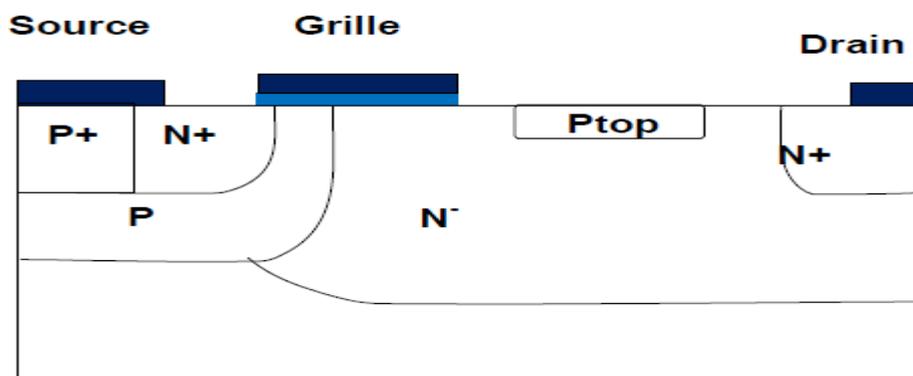


Figure III.11. Coupe schématique d'un LDMOS LOCOS de type N en configuration Standard

III.5.2.2. Le LDMOS Resurf (Reduced surface field)

Cette structure est obtenue en remplaçant l'épitaxie N- sur substrat N+ par une couche épitaxie N- sur un substrat P-. Le substrat P- doit être relié à la source, à travers la diffusion P source, pour que la diode plane ainsi formée soit polarisée en inverse et soutienne, dans certaines conditions, la tension drain source. Si la structure Resurf est bien conçue, la zone N- doit être complètement délatée avant que la zone de surface n'ait un point où le champ atteigne le champ critique de claquage. Si, en plus, la distance entre la source et la diffusion N+ de drain est telle que le perçage de la zone N+ drain soit effectif, le claquage peut s'opérer à la jonction plane N-P permettant ainsi d'améliorer nettement la tension de claquage par rapport à la structure LDMOS classique. Un autre avantage de cette structure est que la résistance à l'état passant reste identique à celle de la structure LDMOS classique alors que la tension de claquage est nettement améliorée. Cette technique a permis aux circuits intégrés de puissance monolithiques de dépasser des tensions de claquage de 250 Volts [15].

Cette structure est illustrée par la figure III.12ci dessous.

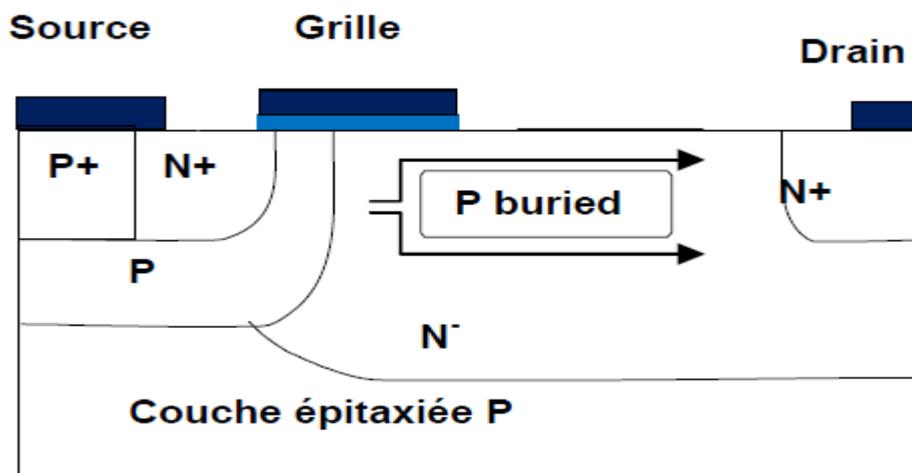


Figure III.12. Coupe schématique d'un LDMOS resurf de type N en configuration standard

III.5.3. Cas particulier du DMOSFET

La structure la plus répandue pour les MOSFET de puissance utilisés en commutation est celle correspond au DMOSFET. Nous nous limiterons donc uniquement à l'étude de cette structure.

Un MOSFET de puissance est formé d'un grand nombre de cellules élémentaires mises en parallèle.

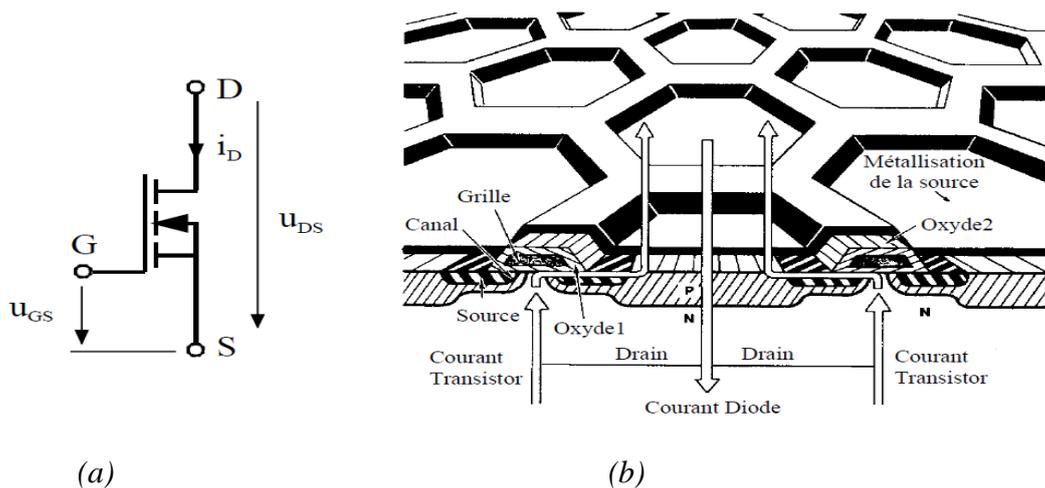


Figure III.13.(a) Représentation symbolique, (b) Structure en nid d'abeille

Le symbole usuel du MOSFET de puissance est représenté à la Figure III.13 (a). Figure III.13 (b) représente une vue en coupe de deux cellules voisines:

- l'embase métallique sur laquelle est posée la pastille de silicium constitue le contact de drain D.
- La zone en contact avec l'embase est une zone du type N appelée substrat. La partie N_0 faiblement dopée sert à assurer la tenue en tension à l'état bloqué. La partie $N+$ évite que l'épaisseur totale ne soit excessive.
- A la surface du substrat ont été diffusés des îlots de type P et dans ces îlots ont été diffusés des îlots de type $N+$.
- Une première couche d'oxyde isole la métallisation de grille G de la surface de la pastille entre les îlots $N+$ et le substrat.
- Une deuxième couche d'oxyde isole la métallisation de grille de la métallisation de source S qui relie entre elles les diverses cellules élémentaires.

III.6.Principe de fonctionnement

Dans un transistor LDMOS à enrichissement à canal N, polarisé en direct, le fonctionnement est régi par la contribution de trois zones ; Ceci est illustré par la figure III.14 ci-dessous:

- la zone active de canal qui assure la conduction du courant électrique du drain ; le canal d'inversion est formé à partir des porteurs minoritaires (électrons) induits en surface de la zone P (Pbody), sous l'effet d'une polarisation positive entre la grille et la source.
- La zone d'accès qui se forme sous la grille dans la zone superficielle dopée N, fonctionnant en régime accumulé. Sa présence est également due à une polarisation positive sur la grille.
- La zone d'extension dopée N, qui permet l'acheminement des électrons de la source vers le drain. Cette zone a pour rôle de permettre la tenue en tension du dispositif.

Les caractéristiques de sortie courant-tension d'un LDMOS sont aussi illustrées par la figure III.14.

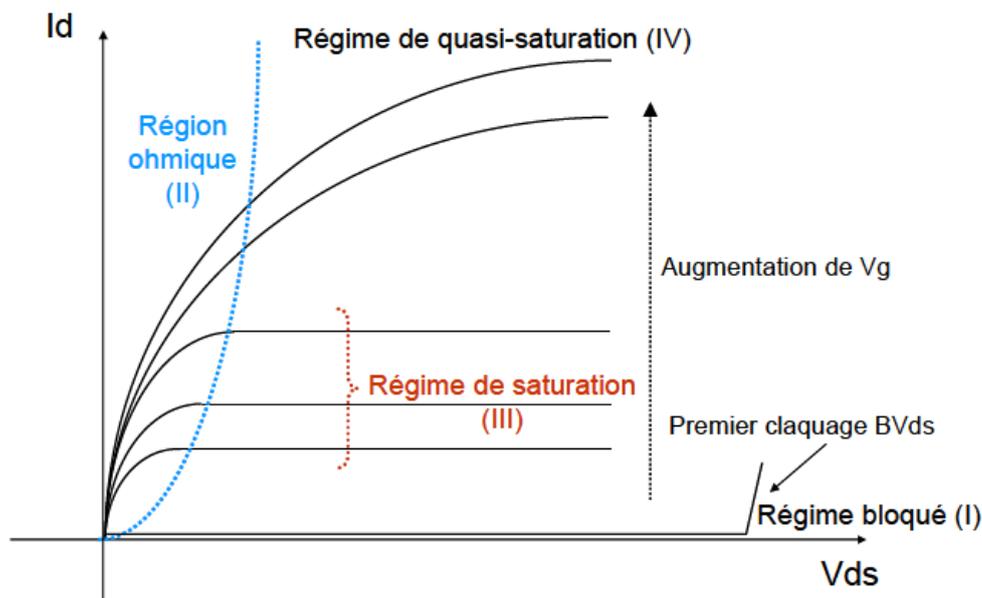


Figure III.14. Illustration des différents régimes de fonctionnement d'un transistor MOS latéral de puissance [19]

On distingue quatre régimes de fonctionnement. Trois de ces régimes sont classiques et communs au MOS standard et correspondent aux états suivants :

1. L'état bloqué où la tension de grille V_g est inférieure à V_t , tension de seuil du transistor. La tension de claquage source-drain est atteinte à mesure que la tension de drain appliquée V_d croît et correspond à $V_{ds} \sim BV_{ds}$.
2. Le régime ohmique ou linéaire où $V_{ds} < V_{gs} - V_t$
3. Le régime saturé où $V_{ds} \gg V_{gs} - V_t$
4. Le quatrième régime est connu sous le nom de régime de « quasi-saturation » et peut se produire à fort niveau de courant. Ces différents régimes de fonctionnement font l'objet d'une étude plus approfondie dans la suite de ce paragraphe.

III.7. Régime de fonctionnement statique

III.7.1. Tension de seuil V_t

La tension de seuil pour les LDMOS est définie par la tension de grille à partir de laquelle le transistor commence à conduire (formation du canal d'inversion entre la source et le drain). Les transistors LDMOS de type N+-P-N--N+ sont constitués d'un canal de type P avec un profil de dopage non-uniforme le long de l'interface Si/SiO₂.

Ce profil de dopage est illustré par la Figure III.15. La zone où la concentration en impuretés est la plus élevée impose la tension de seuil, qui correspond à la valeur maximale de dopage N_{amax} . L'expression de la tension de seuil est la suivante [10]:

$$V_T = \phi_{ms} - \frac{Q_{SS}}{C_{OX}} + 2\phi_F + 2\sqrt{(qN_{amax}\epsilon_s\phi_F)} / C_{OX} \quad (III.26)$$

Où ϕ_{ms} est le travail de sortie métal semi-conducteur, Q_{SS}/C_{OX} la tension correspondant à l'effet des charges à l'interface et ϕ_F le niveau de fermi pour N_{amax}

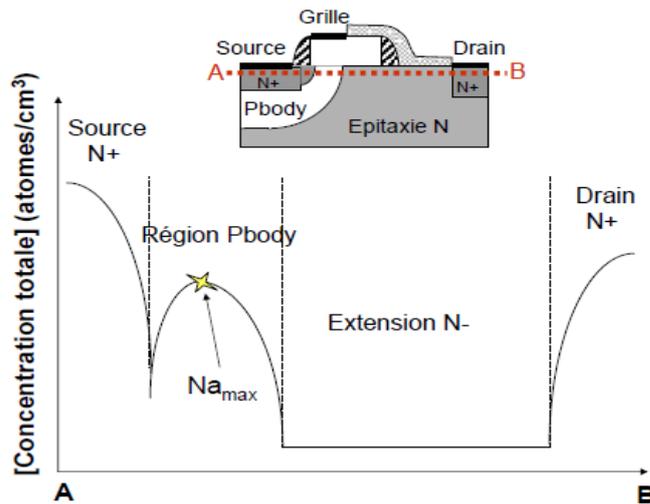


Figure III.15. Profil du dopage le long de l'interface Si/SiO₂ caractéristique d'un LDMOS

III.7.2. La résistance à l'état passant (R_{on})

La résistance à l'état passant est un des paramètres spécifiques des composants de puissance.

Elle est définie comme étant le rapport de la tension de drain sur le courant de drain lorsque cette même tension est proche de zéro avec une tension de grille de référence (2.5 V dans le cas de la technologie 0.25 μm) et elle donnée est par l'équation suivante [19] :

$$R_{ON} = \left(\frac{V_D}{I_D} \right)_{V_{d \rightarrow 0}} \quad \text{(III.27)}$$

Elle indique le niveau de performance d'un composant ; De la valeur de cette résistance dépendent les pertes en conduction. Plus sa valeur est élevée, plus son utilisation dans une application de puissance commutée reste critique mais acceptable pour des applications de puissance en régime linéaire. Pour des transistors avec une extension de drain, la résistance à l'état passant diminue lorsque la tension de grille V_g appliquée augmente comme la montre la figure III.15.

Cette diminution suit une tendance hyperbolique tant que le produit $R_{on} * (V_g - V_t)$ reste une fonction linéaire de la tension de grille (figure III.16). Ces propriétés sont identiques à celles d'un transistor MOS avec une configuration N+-P-N+. Néanmoins dans le cas des structures N+-P-N--N+ il apparaît que R_{on} est aussi dépendant des propriétés de la zone d'extension N-.

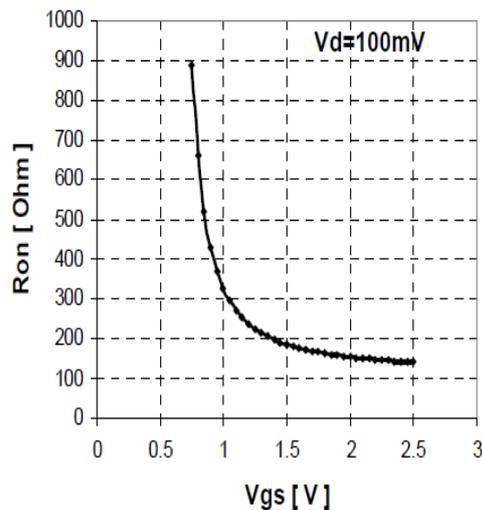


Figure III.16. Variation expérimentale de la résistance à l'état passant d'un transistor LDMOS en fonction de la tension de grille-source

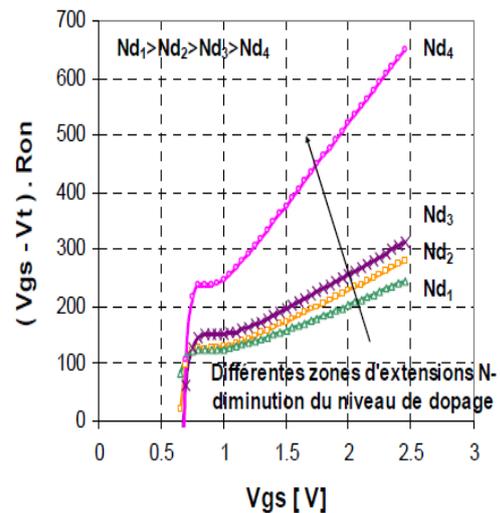


Figure III.17. Variation expérimentale de $R_{on} * (V_g - V_t)$ en fonction de la tension grille-source du composant pour plusieurs zones d'extension N

La résistance à l'état passant, R_{on} peut être considérée comme la somme de trois résistances série comme le montre la figure III.16 et s'exprime suivant l'équation suivante :

$$R_{ON} = R_{canal} + R_a + R_d \tag{III.28}$$

La résistance du canal d'inversion canal R correspond à la résistance à l'état passant d'un transistor MOS N+-P-N+ qui posséderait la même géométrie et les conditions de dopage de la zone Pbody sans extension de drain. Cette résistance dépend également de la tension de grille.

La résistance de drift R_d est la résistance d'extension de drain de type N-. Elle est fonction de la résistivité de la zone et indépendante de la tension de polarisation de grille V_g .

La résistance de la couche d'accumulation R_a est la résistance de la couche d'accès. Qui, sous polarisation grille-drain positive, devient accumulée. Cette résistance dépend également des données géométriques et de la polarisation de grille.

De manière plus générale, les résistances R_{canal} et R_a dépendent surtout de la tension de grille tandis que la résistance de la zone de drift R_d peut devenir linéairement dépendante de la tension de drain (à forte tension V_d).

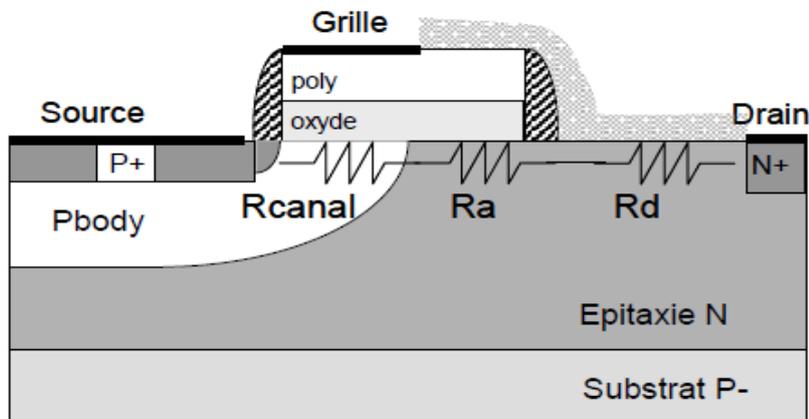


Figure III.18. Différentes composantes de la résistance à l'état passant du LDMOS

De plus les résistances qui ont le plus de poids dans la contribution à la résistance à l'état passant sont les résistances de drift et de canal. Pour des tensions de grilles élevées la résistance de drift domine toutes les autres composantes et est difficile à diminuer sans dégrader la tenue en tension du composant.

III.7.3. Régimes de saturation et quasi-saturation

III.7.3.1. Régime de saturation

Lorsque la tension de drain V_d augmente et est supérieure à $(V_g - V_t)$, la caractéristique courant-tension de drain sature quelque soit le type de transistor MOS étudié (régime de saturation illustré sur la Figure III.14).

Le courant de drain I_{ds} devient pratiquement indépendant de la tension de drain V_{ds} , une fois que celle-ci a dépassé la valeur particulière $V_{d_{sat}}$ dite tension de pincement. Le mécanisme physique responsable du passage de la région ohmique à la région saturée est le phénomène de saturation de la vitesse des porteurs dans le canal combiné à la diminution de la densité des porteurs, dans la zone près du drain, où un champ électrique élevé règne [4].

III.7.3.2. Régime de quasi-saturation

Le régime de quasi-saturation ne se manifeste pas en général dans les transistors MOS classiques. Il se produit dans les structures avec une extension de drain par un tassement des caractéristiques de sorties à fort V_g . Il se traduit par l'insensibilité du courant de drain à l'augmentation de la polarisation grille-source V_{gs} surtout vers les valeurs élevées de cette polarisation, ce qui a pour conséquence d'augmenter la conductance de sortie aux fortes polarisations de drain.

Le phénomène physique de quasi-saturation dans les MOS verticaux a été pour la première fois rencontré et interprété par Caquot et al. Par la suite de nombreux travaux ont permis de mieux interpréter ce phénomène dans les transistors de puissance.

Le transistor VDMOS haute tension (150V-500V) est couramment schématisé comme deux parties en série: le MOS intrinsèque et la zone d'extension de drain. La quasi-saturation combine à la fois l'insensibilité du courant vis-à-vis de la tension de grille et la dépendance quasi-linéaire du courant de drain en fonction de la tension de drain appliquée. Cette dépendance linéaire a été largement attribuée au comportement du MOS intrinsèque en régime quasi-linéaire.

Une première théorie développant l'insensibilité du courant de drain à la tension de grille a été proposée par Darwish et confirmée par Evans. Elle prend en compte la relation vitesse-champ électrique et la saturation de la vitesse des porteurs dans la zone de drift. La mobilité des porteurs évolue de façon hyperbolique par rapport au champ électrique et permet une modélisation fiable du quasi saturation pour les caractéristiques de sortie. Une autre approche prend en compte la formation d'une zone de déplétion entourant le canal ainsi que sa modulation avec la tension de drain V_d . La résistance de la zone N- est pincée par l'augmentation des zones de charges d'espace. La chute de tension présente entre le drain et le canal empêche d'atteindre la tension de pincement ($V_g - V_t$), ainsi la saturation n'est pas atteinte. Ce phénomène est similaire au pincement du JFET.

De plus les mécanismes d'activation de la quasi-saturation diffèrent suivant la configuration verticale ou latérale du composant. Récemment, un modèle analytique de la résistance de drift

déduit du phénomène de quasi-saturation a été proposé sur les LDMOS faible tension (<80V) par Montcoq. Seule l'hypothèse de la saturation de la vitesse des porteurs a été considérée.

Cette approche a été également confirmée par Anghel pour les LDMOS 20V où il apparaît que la formation d'une zone dilatée dans le chemin du courant aboutit à un comportement complètement différent de celui de la quasi-saturation.

III.7.3.3. Tenue en tension (BV_{ds})

Pour déterminer la tension de claquage drain-source, on considère le transistor à l'état bloqué, c'est à dire que le courant de drain est théoriquement nul. Le transistor LDMOS est considéré comme une diode P-N--N+ en régime bloqué.

Le claquage a lieu normalement là où le rayon de courbure de la jonction est minimal dans le cas de structures planaires (typiquement MOS standard ou LDMOS). Les paramètres du premier ordre déclenchant la tension de claquage sont le niveau de dopage, le rayon de courbure de la diffusion (l'effet de multiplication des porteurs dans les zones de canal qui sont les régions actives de la structure et également l'éventuelle conséquence de cette multiplication sur l'apparition de caractéristiques à résistance de sortie négative appelées caractéristique de second claquage sont négligés). Le mécanisme physique responsable du claquage est l'effet d'avalanche. Dès que le champ électrique est de l'ordre de 10^5 V/cm, l'accélération acquise par quelques porteurs, essentiellement d'origine thermique, qui transportent le courant inverse, est suffisante pour leur permettre de générer des paires électron-trou par ionisation par impact. Ces paires sont à leur tour accélérées, et peuvent créer d'autres paires, c'est l'effet d'avalanche. La caractéristique $I(V)$ de la tenue en tension est illustrée sur la Figure III.15.

III.7.4.L'auto-échauffement

Un autre aspect caractéristique observable sur les MOS de puissance est le phénomène d'auto échauffement. Cet effet principalement thermique est identifié à forts niveaux de courant ou de tension. Il se manifeste par la diminution du niveau de courant de sortie à mesure que la tension de drain appliquée augmente (Figure III.19).

Comme dans la plupart des dispositifs hauts tension la dissipation locale de puissance engendrent leur « échauffement ». C'est cette augmentation de température locale dans le

composant qui entraîne la réduction de la mobilité et par la même occasion la diminution de la vitesse des porteurs. Par conséquent à mesure que la température augmente le courant de sortie décroît. De plus cet effet qui se présente à la fois dans les régimes de saturation et de quasi-saturation est seulement dépendant du niveau de puissance continue DC d'entrée. Ce phénomène a fait l'objet de nombreuses études dans les années 90 dans de nombreux domaines (MOSFET, MESFET, BJT, SiC composants, Transistor GaAsAlGaAs) et plus particulièrement sur les technologies SOI car elles présentent un auto échauffement beaucoup plus important que dans les technologies standards silicium.

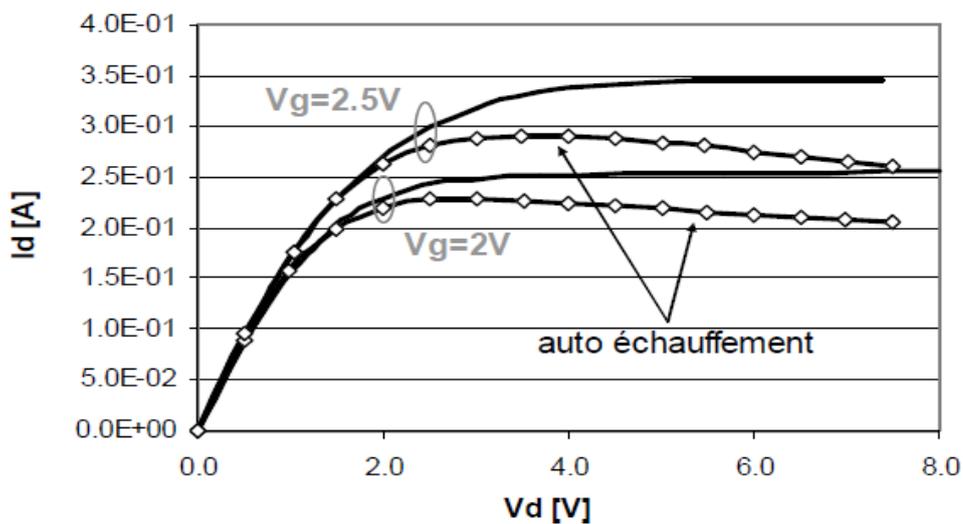


Figure III.19. Caractéristiques de sortie mesurées d'un LDMOS. Comparaison des courbes affectées (mesure DC) ou non affectées (mesure pulsée) par l'auto échauffement

La caractérisation de l'auto échauffement se fait par l'intermédiaire d'une « Caractérisation en impulsion » qui permet le contrôle de l'état thermique du transistor.

Ce procédé permet entre autre de réaliser des mesures $I(V)$ de transistors dans des conditions quasi-équithermiques. Il a fait son apparition au cours des années 80 ; les travaux de Paggiont décrit les caractéristiques non linéaires [22] du courant de drain en appliquant des impulsions uniquement sur la commande de grille. Par la suite, de nombreux chercheurs ont amélioré cette technique en proposant des bancs de mesures impulsionnels plus complexes et automatisés permettant l'acquisition des caractéristiques des transistors par excitation simultanée des deux accès du composant (grille et drain).

La Figure III.20 décrit le principe de caractérisation impulsionnelle d'un FET. Il consiste à décrire les caractéristiques d'entrée et de sortie du composant par de brèves impulsions issues d'un point de polarisation choisi. Pour un point de repos donné (V_{g0}, V_{d0}) il correspond un courant de repos I_{d0} . Une impulsion simultanée sur la grille et sur le drain d'amplitudes différentes fixe les tensions de polarisation (V_{gi}, V_{di}) pour lesquelles le composant délivre un courant de sortie I_{di} . La caractéristique complète de sortie du composant est alors obtenue en faisant varier les amplitudes des polarisations (V_{gi}, V_{di}).

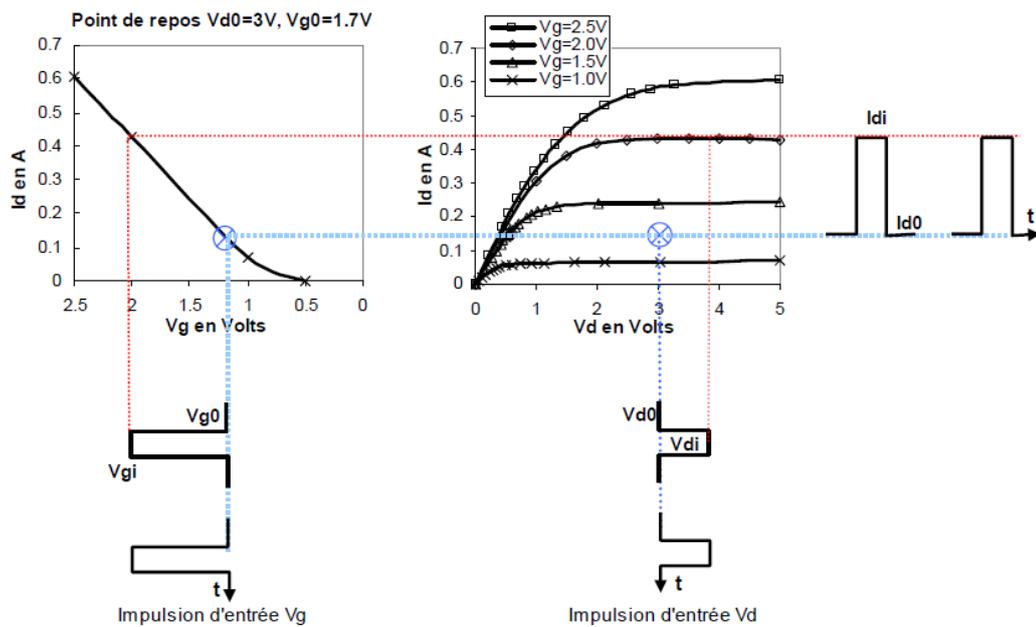


Figure III.20.Principe de mesure en impulsion d'un transistor MOSFET

III.8.Le régime de fonctionnement en dynamique

En dehors des figures de mérite statiques (R_{on} , BV_{ds} , $I_{sa}t$ etc...) d'un composant il est aussi intéressant d'avoir connaissance de ses caractéristiques dynamiques petit signal et grand signal. Ce paragraphe expose les composantes dynamiques du LDMOS. Les principales notions de mesures dynamiques petit-signal et grand-signal sont introduites ainsi que les figures de mérite qui en découlent.

III.8.1.LDMOS en régime dynamique

Les éléments capacitifs sont extraits afin de pouvoir représenter le comportement du transistor en dynamique. Il s'agit des capacités inter électrodes:

- la capacité drain-source C_{ds} correspondant à la jonction drain-source polarisée en inverse.
- la capacité grille-source C_{gs} relative à l'oxyde de grille. Elle est la mise en parallèle de la capacité localisée entre la zone du canal Pbody et la grille et la capacité entre la métallisation de la source et du poly silicium de grille.
- la capacité C_{gd} qui est l'association de deux capacités : une capacité d'oxyde et une capacité associée à la charge d'espace de la zone de déplétion N dans la zone d'extension. Il existe de plus une capacité C_{gd} parasite principalement liée aux métallisations.

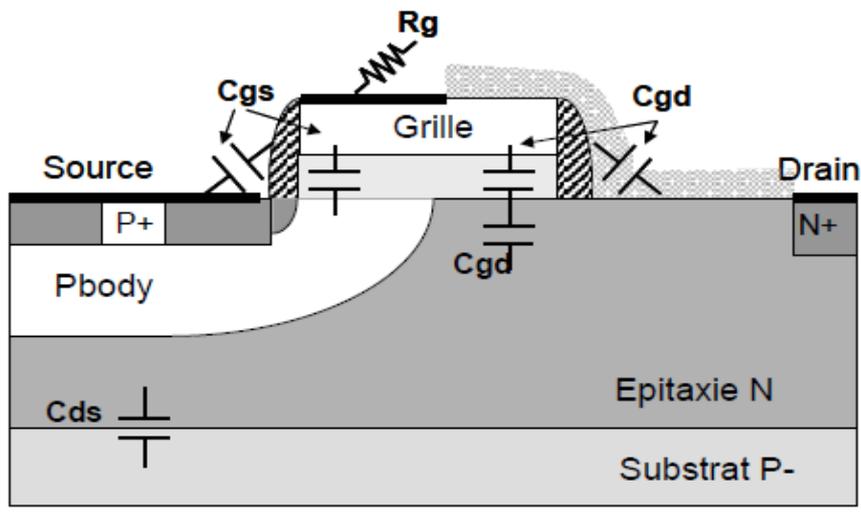


Figure III.21.Représentation des éléments capacitifs du transistor LDMOS

Des éléments parasites sont à prendre en compte éventuellement comme les capacités d'interconnexions et la R_g (résistance de grille) mais dépendent de la topologie des prises de contacts et du montage transistor dans l'application de type « circuit ».

III.9. Avantages des transistors DMOS

Par rapport aux transistors VMOS, les avantages des transistors DMOS sont multiples. On peut citer les plus intéressants :

- Ils peuvent être fabriqués aisément sur n'importe quelle orientation cristalline du Silicium, alors que le VMOS est obligé d'avoir son canal le long d'une surface. Ainsi, le choix d'une surface orientée pour un transistor DMOS permet une amélioration de 20% de la mobilité des électrons dans la couche inversée et de 15% de la vitesse limite de saturation de ces électrons dans cette même couche inversée par rapport aux transistors VMOS équivalents. Il en résulte une plus faible résistance à l'état passant et une plus importante transconductance par unité de surface dans les structures DMOS.
- La densité Q_{ss} de charges fixes à l'interface silicium/oxyde est d'environ trois fois plus grande sur les plans par rapport aux plans. Par conséquent, pour une tension de seuil donnée, le pic de dopage dans le canal doit être plus élevé dans une structure VMOS que dans une structure DMOS. Or, la mobilité des électrons dans une couche inversée diminue quand le dopage augmente [19], donc la mobilité dans la structure VMOS est à nouveau dégradée par rapport aux structures DMOS.
- Les difficultés technologiques évoquées pour le transistor VMOS n'existent pas pour les transistors DMOS. Ces derniers font appel à une technologie plus simple à haut rendement. Par exemple, le procédé d'auto-alignement des diffusions par une grille en poly-silicium est relativement simple à mettre en œuvre dans les transistors DMOS. Ce point évidemment essentiel explique le développement industriel considérable de ces composants.
- La densité d'intégration - rapport périmètre/surface - qu'il a été possible d'obtenir avec des transistors DMOS multicellulaires est rapidement devenue plus grande que celle des transistors VMOS.

III.10. Influence la température de jonction (T_J) sur le comportement du transistor

Les semi-conducteurs étant très sensibles à la température, il est primordial de comprendre les phénomènes qu'elle entraîne [17].

III.10.1.Modification de la caractéristique statique

III.10.1.1.Caractéristique $I_D = f(V_{DS}, V_{GS})$

Pour mesurer cette caractéristique on définit, à une valeur de température de jonction initiale, les tensions V_{GS} et V_{DS} ainsi que de la durée de l'impulsion appliquée.Cette caractéristique est illustrée par Figure III.22 [17].

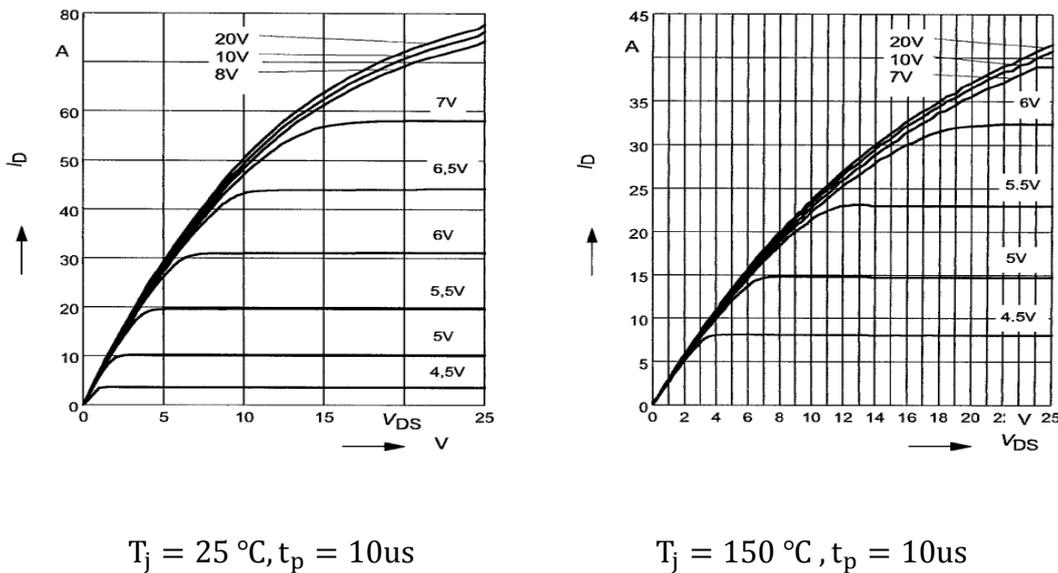


Figure III.22.Caractéristique $I_D = f(V_{DS}, V_{GS}, T_J)$

On peut conclure aisément que cette caractéristique est fortement dépendante de la température de jonction. On voit que pour des tensions $V_{GS} < 6\text{V}$, le coefficient en température est positif ce qui signifie qu'une augmentation de température entraîne une augmentation du courant I_D dans la zone active.

III.10.1.2.Caractéristique de transfert $I_D = f(V_{GS})$

Cette caractéristique n'a de signification que pour la région active, c'est-à-dire lorsque le MOSFET fonctionne en source de courant contrôlé en tension. Pour s'assurer que nous nous trouvons en région active, il faut respecter la condition suivante sur la tension V_{DS}

$$V_{DS} > 2I_D \cdot R_{DS(on)} \tag{III.29}$$

La pente de cette courbe représente la transconductance. Cette dernière est donnée par la relation [17]:

$$g_{fs} = \left[\frac{\Delta I_D}{\Delta V_{GS}} \right]_{V_{DS}, I_D} \quad (III.30)$$

La transconductance est un paramètre important dans le calcul des temps de commutation.

Cette caractéristique est mesurée pour des impulsions de tension Grille – Source de $t_p=10\mu s$. Pour de forts courants, l'augmentation de la température de jonction durant la mesure (10 μs) provoque une diminution de la transconductance. On remarque ce phénomène sur la Figure III.23.

$$g_{fs}(T_j) = g_{fs(300^\circ K)} \left(\frac{T_j}{300} \right)^{-\alpha} \text{ Avec } T_j \text{ en } [^\circ K] \quad (III.31)$$

Avec $\alpha = \begin{cases} 2.3 & \text{pour une structure trench} \\ 1.5 & \text{pour une structure cool Mos} \end{cases}$

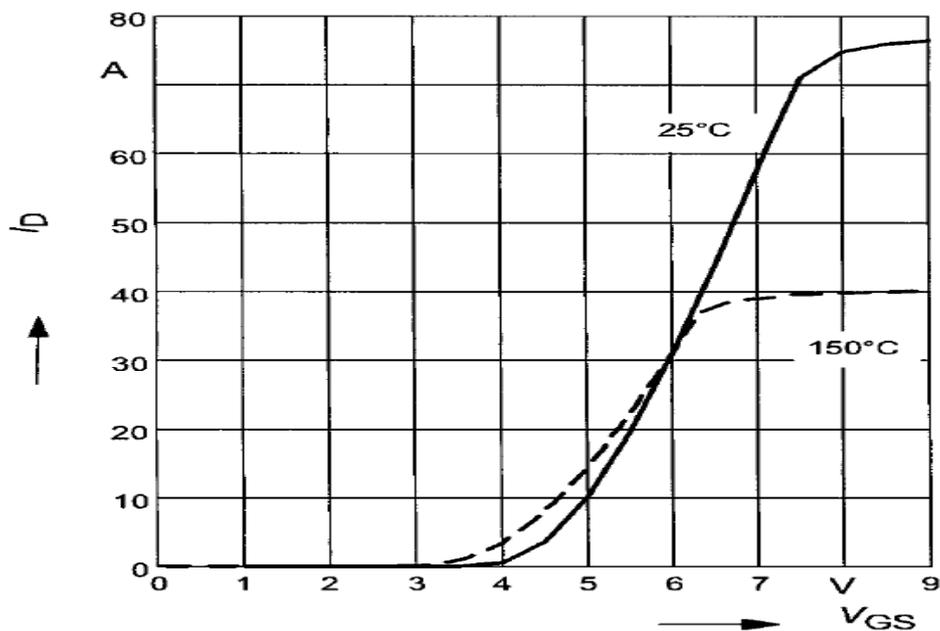


Figure III.23. Caractéristique de transfert $I_D = f(V_{GS})$

Les caractéristiques dynamiques des transistors MOSFET ne varient pas significativement avec la température : les capacités parasites ne sont que peu modifiées. Seule la diode interne voit son courant de recouvrement augmenter, ce qui peut s'avérer dangereux.

III.11.Conclusion

Dans ce chapitre, une vue d'ensemble d'architecture du transistor MOSFET de puissance aux diapositifs radiofréquence a été présentée. Deux architectures principales de transistor de puissance : transistor MOSFET latéral double diffusé LDMOS et le transistor MOSFET vertical double diffusé VDMOS.

Nous avons ensuite vu les spécificités des transistors destinés aux applications radiofréquences de puissance, qui font l'objet de notre étude : les caractéristiques I_D en fonction de V_{DS} et leur tenue aux régimes extrêmes (forte température, avalanche répétitive).

Le prochain chapitre fera l'objet de la modélisation des transistors de puissance appliqués aux dispositifs radiofréquences par la méthode neuromimétique.

Chapitre IV

MODELISATION DES TRANSISTORS FET APPLIQUES AUX DISPOSITIFS RADIOFREQUENCES EN UTILISANT L'APPROCHE NEURONALE

IV.1. Introduction

L'utilisation de simulateurs est conditionnée par l'existence de modèles non linéaires précis et fiable. Le modèle doit donner une précision meilleure sur la caractéristique courant-tension dans une large gamme de polarisation et il doit être simple pour l'implémentation.

Il existe plusieurs techniques de modélisation des composants ou dispositifs radiofréquences, chacune peut offrir un modèle non linéaire complètement différent. Les trois types de modélisation sont les suivantes :

- La modélisation physique basée sur les équations physiques du composant et son procédé technologique. L'inconvénient majeur de ce type de modélisation réside dans son temps de développement qui peut durer longtemps. Par conséquent le temps de calcul rend son utilisation difficile dans les simulateurs.
- La modélisation électrique utilise un circuit électrique équivalent qui peut être simple ou compliqué ; Il est obtenu après une analyse physique du dispositif à modéliser. Ce type de modélisation comporte une procédure d'extraction des éléments du circuit rigoureuse.
- La modélisation par des approches mathématiques de type 'boite noire' présente beaucoup d'avantages, vu qu'il ne dépend pas de la technologie du composant. La modélisation de type 'boite noire' utilise un fichier de mesures prédéterminé. L'inconvénient majeur de ce type de modélisation est la taille de la table qui représente la non linéarité.

Le but primordial recherché dans ce travail est de trouver un autre type de modélisation pour lever ces limitations, il s'agit de l'approche neuronale.

Les réseaux de neurones peuvent modéliser des fonctions non linéaires à plusieurs degrés ainsi que leurs dérivées. L'avantage principal de ce modèle est son apprentissage des informations et sa généralisation.

Dans les transistors MOSFET de puissance le courant de drain I_D est le non linéarité la plus importante qu'il faut modéliser par un modèle neuronal.

La construction d'un modèle à base de réseaux de neurones artificiels RNA dédié au MOSFET de puissance du dispositif radiofréquence, a pour but la reproduction fidèle de sa

réponse lors de la variation des différents paramètres électriques du transistor MOSFET de puissance.

IV.2. Modélisation du transistor MOSFET de puissance

Les réseaux de neurones de type Feedforward peuvent modéliser n'importe quelle fonction non linéaire ; Ils utilisent la retro propagation du gradient dans la phase d'apprentissage.

Pour créer le modèle du MOSFET de puissance au moyen des réseaux de neurones on choisit une base de données caractérisée, dans notre cas, par les paramètres électriques (la tension de drain V_{DS} , la tension de grille V_{GS}) et la température (T_j) de l'environnement où le MOSFET de puissance est placé. Dans une deuxième étape on sépare la base d'apprentissage et celle de test, puis on fait l'entraînement d'un réseau de neurones sur la base d'apprentissage avec l'algorithme de rétro propagation et finalement on mesure la performance du modèle obtenu avec la base de test.

IV.2.1. Choix de la base de données

La gamme des différents paramètres appliqués sur le transistor MOSFET de puissance IFR7413 sont présentés sur les équations suivantes:

- Dans la zone ohmique le courant du drain obéit à l'équation suivante

$$I_D = K_P \left[(V_{GS} - V_{TH}) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (\text{IV.1})$$

- En régime de saturation le courant de drain est donné par la formule IV.2 ci-dessous:

$$I_D = \frac{K_P}{2} (V_{GS} - V_{TH})^2 \quad (\text{IV.2})$$

Ou

K_P : Représente la transconductance du MOSFET ($A.V^{-2}$),

V_{GS} : Tension Grille-Source(V)

V_{TH} : Tension de seuil du MOSFET(V)

V_{DS} : Tension Drain-Source(V)

A partir du data sheet du composant radiofréquence IRF7413 (transistor MOSFET de puissance) utilisé, la figure IV.1 illustre les caractéristique de sortie I_D en fonction de V_{DS} pour des tensions grille source V_{GS} constantes.

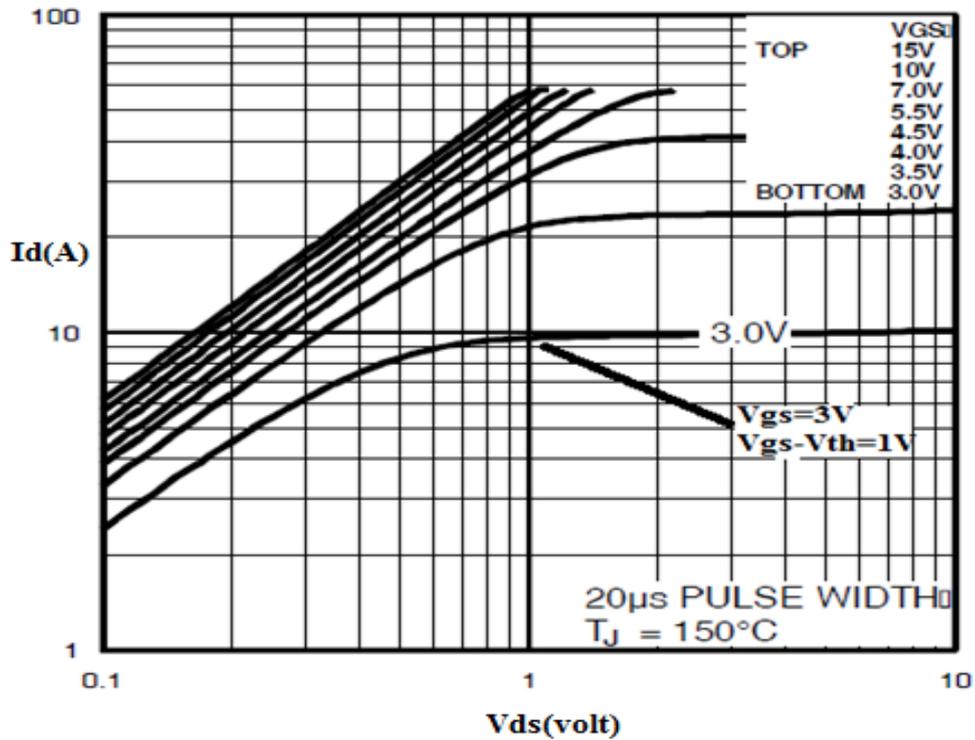


Figure IV.1. Caractéristiques I_D en fonction de V_{DS} pour $T_j = 150^\circ\text{C}$

La tension de seuil de ce transistor est d'environ 2V ($V_{TH} = 2V$).

On notera que les caractéristiques sont bien horizontales même à fort courant ce que traduit une impédance interne élevée.

En faisant varier la température on a:

Premier cas : $T_j = 150^\circ\text{C}$

I_D reste constant à 10A et V_{GS} est égale à 3V, la valeur de K_P est donnée par la formule IV.2 :

$$K_P = \frac{2 \cdot I_D}{(V_{GS} - V_{TH})^2} \quad \text{Donc} \quad K_P = 20 \left(\frac{\text{A}}{\text{volt}^2} \right)$$

On remplace la valeur de K_P dans les équations (IV.1) et (IV.2), il vient :

- Pour $V_{GS} > V_{TH}$

$$\begin{cases} I_D = 20 \left[(V_{GS} - 2)V_{DS} - \frac{V_{DS}^2}{2} \right] \text{ pour } V_{DS} < V_{GS} - 2 \\ I_D = \frac{20}{2} (V_{GS} - 2)^2 \text{ pour } V_{DS} > V_{GS} - 2 \end{cases} \quad (\text{IV.3})$$

- Pour $V_{GS} < V_{TH}$

$$I_D = 0 \quad (\text{IV.4})$$

On trace I_D en fonction de V_{DS} pour $K_P = 20 \left(\frac{A}{\text{volt}^2} \right)$ à partir de l'équation (IV.3)

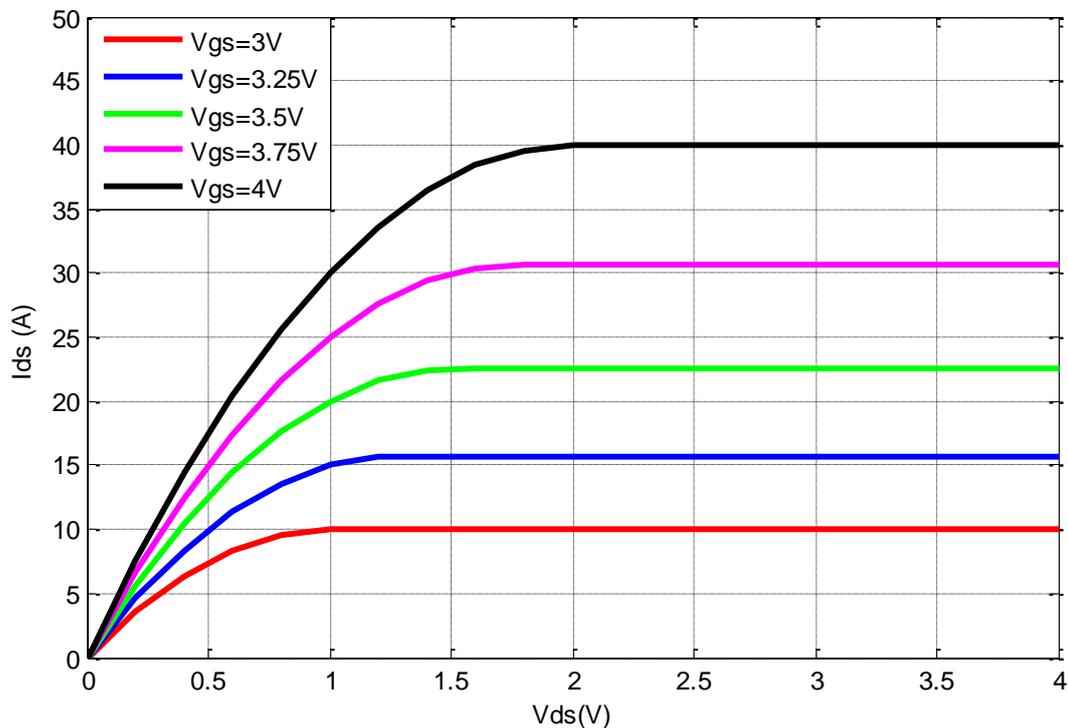


Figure IV.2. Caractéristiques I_D en fonction de V_{DS} pour $T_j = 150^\circ\text{C}$

La Figure IV.2 représente la caractéristique du courant de drain I_D en fonction de la tension Drain-Source (V_{DS}) [$I_D = f(V_{DS})$], pour des différentes valeurs de la tension de grille à une température ($T_j = 150^\circ\text{C}$).

Deuxième cas : $T_j = 25^\circ\text{C}$

La caractéristique courant- tension du transistor est illustrée dans ce cas par la figure IV.3 ci-dessous :

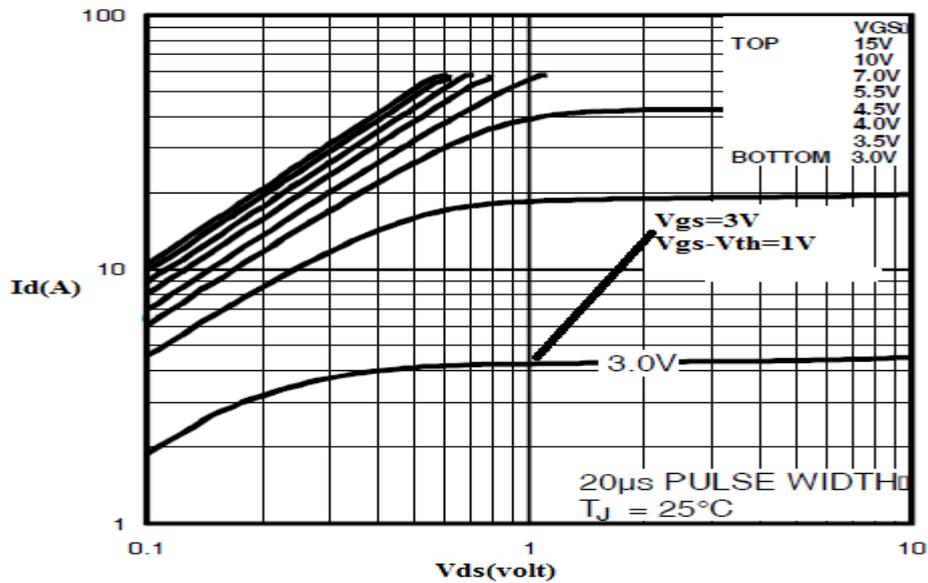


Figure IV.3.Caractéristiques I_D en fonction de V_{DS} pour $T_j = 25^\circ\text{C}$

Dans le régime de saturation I_D reste constant a 4.5A et $V_{GS} = 3\text{V}$, K_P est régi à partir de l'équation (IV.2) par la formule suivante

$$K_P = \frac{2 \cdot I_D}{(V_{GS} - V_{TH})^2} \quad \text{Donc} \quad K_P = 36 \left(\frac{\text{A}}{\text{volt}^2} \right)$$

On remplace K_P dans les équations (IV.1) et (IV.2), il vient:

-Pour $V_{GS} > V_{TH}$ on a :

$$\begin{cases} I_D = 36 \left[(V_{GS} - 2.5)V_{DS} - \frac{V_{DS}^2}{2} \right] & \text{pour } V_{DS} < V_{GS} - 2.5 \\ I_D = \frac{36}{2} (V_{GS} - 2.5)^2 & \text{pour } V_{DS} > V_{GS} - 2.5 \end{cases} \quad (\text{IV.5})$$

-Pour $V_{GS} < V_{TH}$, on a :

$$I_D = 0 \quad (\text{IV.6})$$

On trace I_D en fonction de V_{DS} pour $K_P = 36 \left(\frac{\text{A}}{\text{volt}^2} \right)$ à partir de l'équation (IV.5)

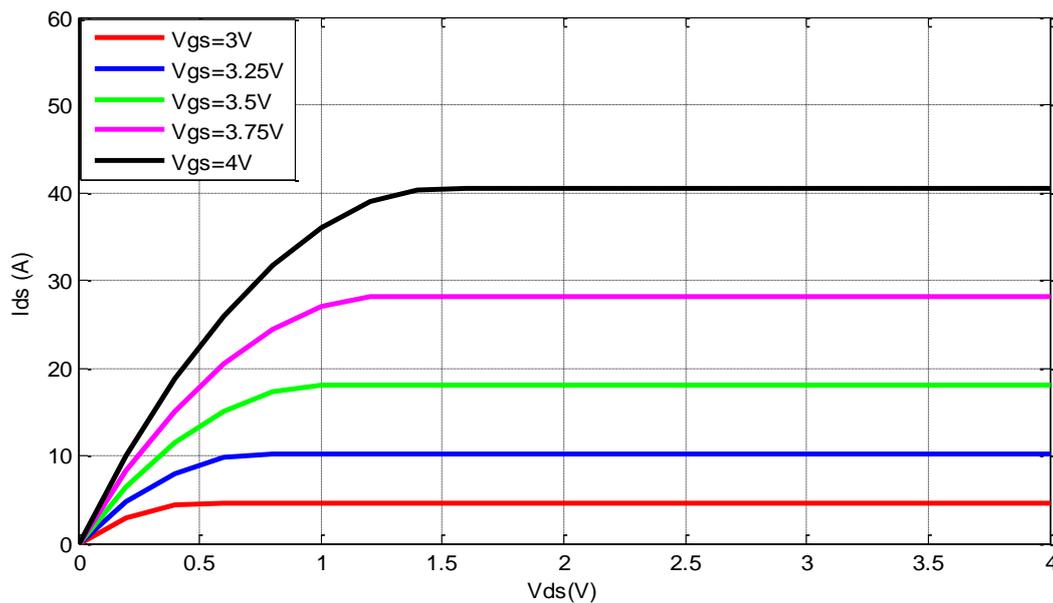


Figure IV.4.Caractéristiques I_D en fonction de V_{DS} pour $T_j = 25^\circ\text{C}$

La Figure IV.4 représente la caractéristique de I_D en fonction de V_{DS} , pour des différentes valeurs de la tension de grille avec ($T_j = 25^\circ\text{C}$).

La base de données doit couvrir cet ensemble des valeurs. En se basant sur les valeurs des différents paramètres notre base de données possèdera donc $5 \times 5 \times 21$ éléments, et l'apprentissage du réseau de neurones est fait avec une base de données de 525 éléments. Ces éléments sont répartis en sous-ensembles. Un sous ensemble d'entrée contenant les deux tensions V_{DS} et V_{GS} , l'autre sous ensemble contient la sortie désiré du courant I_D

Il est important de ne pas utiliser aucun élément de la base de test pendant toute la durée de l'apprentissage. Cette base est réservée uniquement à la mesure finale de la performance. Autrement dit, elle sert à vérifier si le réseau de neurones a une bonne performance sur les exemples non appris « base de test » avec les réseaux de neurones,

IV.2.2. Apprentissage du réseau de neurone

Comme nous avons vu précédemment, l'apprentissage nécessite une base de données, une base de tests, un nombre de couches et des neurones utilisés dans chaque couche.

En général un perceptron multicouche est constitué d'une couche d'entrée, de plusieurs couches cachées et d'une couche de sortie. La fonction d'activation joue un rôle prépondérant dans la

phase d'apprentissage. Ainsi le nombre d'itérations pour la convergence du réseau est aussi important pour avoir un seuil d'estimation E qui représente la valeur minimale de l'erreur quadratique moyenne.

La Figure IV.5 représente l'organigramme principal du calcul du modèle neuronal du courant de drain. Le logiciel utilisé dans toute la programmation est le Matlab :

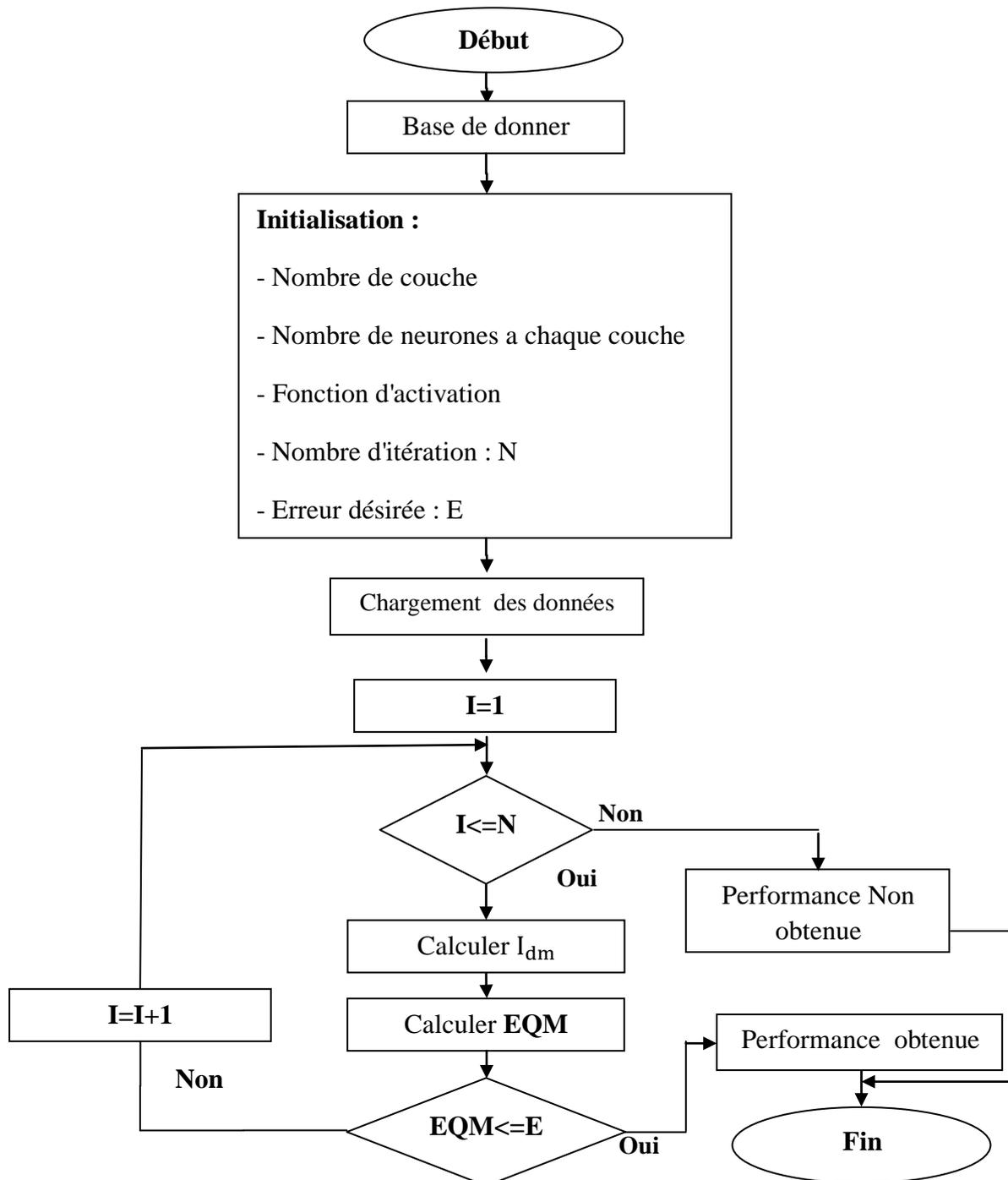


Figure IV.5. Organigramme pendant la phase d'apprentissage

Dans notre cas, puisque le nombre des neurones dans la couche de sortie est déterminé par le nombre de sorties du système à modéliser, le MOSFET de puissance possède une seule sortie I_D donc on a besoin d'un neurone pour la couche de sortie.

La structure du modèle neuronal du courant de drain est illustrée par le schéma de la figure IV.6.

Notons que la structure détaillée de chaque couche sera faite plus loin.

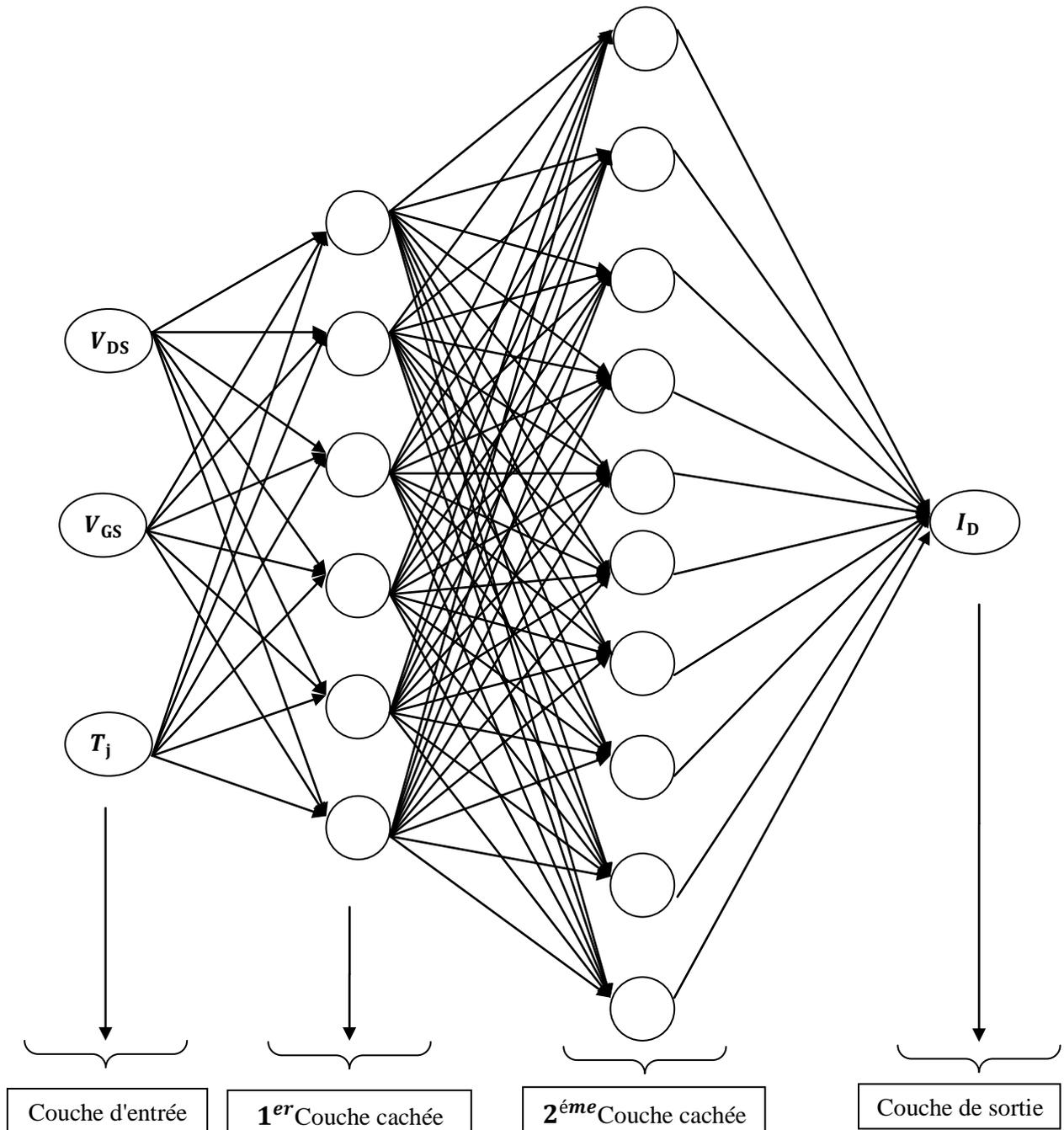


Figure IV.6. Schémas des différentes couches de neurones de la structure du modèle neuronal

Chapitre IV MODELISATION DES TRANSISTORS FET APPLIQUES AUX DISPOSITIFS RADIOFREQUENCES EN UTILISANT L'APPROCHE NEURONALE

Le choix du nombre de neurones est lié à la fonction non linéaire à modéliser. Pour cela on a testé plusieurs réseaux en faisant varier le nombre de neurones cachés et notre choix est fixé sur l'architecture qui a donné une erreur quadratique minimale (EQM) pendant la phase d'apprentissage. Une variation du momentum est fixée entre 0.1 et 0.9. Le momentum permet l'introduction de l'ancienne valeur de chaque paramètre du réseau dans le calcul de sa nouvelle valeur. Il est utilisé pour accélérer la convergence du réseau neuronal.

Le tableau IV.1 représente les différents résultats obtenus après plusieurs tests sur différentes architectures ; L'erreur converge vers le minimum dans la vingtième architecture neuronale, ce qui justifie notre choix antérieur.

	Nombre de neurones 1 ^{er} couche	Nombre de neurones 2 ^e couche	EQM
1	4	5	0.0285
2	4	6	0.0357
3	4	7	0.0224
4	4	8	0.0225
5	4	9	0.0464
6	4	10	0.156
7	5	4	0.0231
8	5	5	0.0297
9	5	6	0.174
10	5	7	0.634
11	5	8	0.0507
12	5	9	0.0511
13	5	10	0.0213
14	6	4	0.0722
15	6	5	0.0109
16	6	6	0.410
17	6	7	0.0714
18	6	8	0.413
19	6	9	0.0202
20	6	10	0.0195
21	7	4	0.0225
22	7	5	0.0553

Chapitre IV MODELISATION DES TRANSISTORS FET APPLIQUES AUX DISPOSITIFS
RADIOFREQUENCES EN UTILISANT L'APPROCHE NEURONALE

23	7	6	0.0387
24	7	7	0.0985
25	7	8	0.0622
26	7	9	0.1614
27	7	10	0.0423
28	8	4	0.0476
29	8	5	0.0324
30	8	6	0.0768
31	8	7	0.0436
32	8	8	0.0634
33	8	9	0.0762
34	8	10	0.0332
35	9	4	0.0271
36	9	5	0.0693
37	9	6	0.0807
38	9	7	0.0626
39	9	8	0.0354
40	9	9	0.0792
41	9	10	0.0543
42	10	4	0.0371
43	10	5	0.0372
44	10	6	0.0377
45	10	7	0.0641
46	10	8	0.1471
47	10	9	0.1502
48	10	10	0.554

Tableau IV.1. Variation de l'erreur du test en fonction des différentes architectures

IV.2.3. Le Choix nombre d'itération

L'erreur quadratique moyenne est donnée par la formule IV.7 ci-dessous :

$$EQM = \sum_k (d_k - y_k^{[s]})^2 \quad (IV.7)$$

k varie de 1 jusqu'à N nombre total d'exemples dans la base d'apprentissage. d_k représente la réponse désirée du réseau et y_k étant la réponse actuelle générée par le réseau neuronal.

Le test et l'observation, de la variation de l'erreur, pour des différentes valeurs est la solution pour avoir les meilleurs résultats. Afin d'évaluer l'influence de l'erreur désirée E sur le nombre d'itérations nécessaires pour obtenir $EQM < E$, nous avons choisi plusieurs valeurs pour E. La figure IV.7 représente l'évolution de l'**EQM** en fonction de nombre d'itérations pour une erreur désirée $E=10^{-3}$.

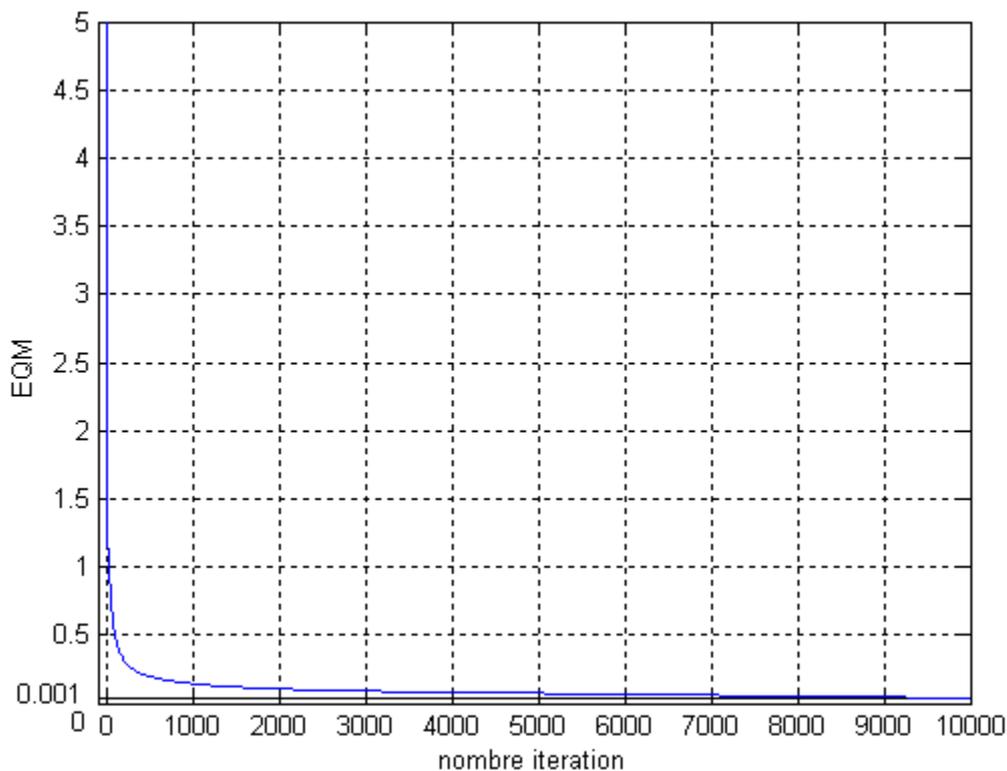


Figure IV.7. EQM en fonction du nombre d'itérations

En augmentant le nombre d'itération l'erreur EQM diminue ce qui assure la convergence du réseau.

IV.2.4. Structure du modèle neuronal du courant de drain et normalisation de la base d'apprentissage

La structure du modèle neuronal du transistor MOSFET de puissance utilisé est :

- L'architecture utilisée est le perceptron multicouche de type (Feed-forward MLP). Il peut modéliser n'importe quelle fonction non linéaire et les dérivées de cette fonction.
- Le perceptron est entraîné par l'algorithme de la retro propagation du gradient avec momentum.
- Le nombre de couches cachées est fixé à 2
- Le nombre de neurones dans la couche d'entrée est fixé à 3, vu qu'on a trois entrées : (V_{DS}, V_{GS}, T_j) .

La première couche cachée contient 06 neurones, la deuxième couche cachée contient 10 neurones et la couche de sortie contient un neurone, vu qu'on a une seule sortie (I_D).

La fonction de transfert d'un neurone représente son état d'activation et elle représente une importance particulière. Elle peut avoir plusieurs formes, dans notre cas on a utilisé :

Une fonction sigmoïde dans les deux couches cachées, sa formule est donnée par l'équation IV.8 et sa forme représentative est illustrée par la figure IV.8.

$$F(x) = \frac{1}{1+e^{-x}} \quad (IV.8)$$

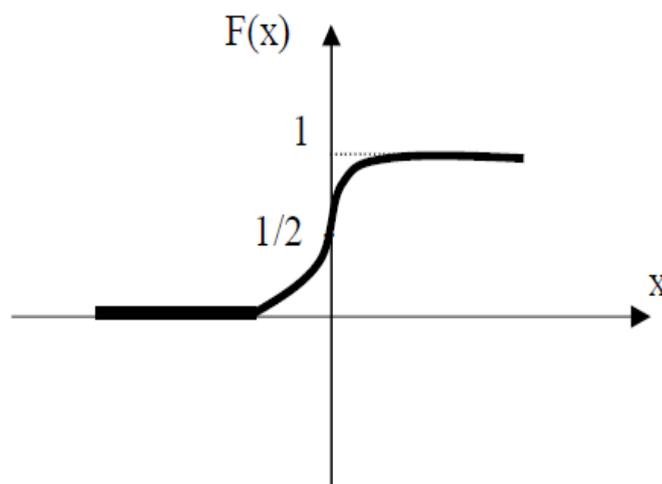


Figure IV.8. Allure d'une Fonction sigmoïde

La couche de sortie est représentée par une fonction linéaire ; l'équation et la forme de la fonction sont données ci-dessous :

$$F(x)=x \quad (IV.9)$$

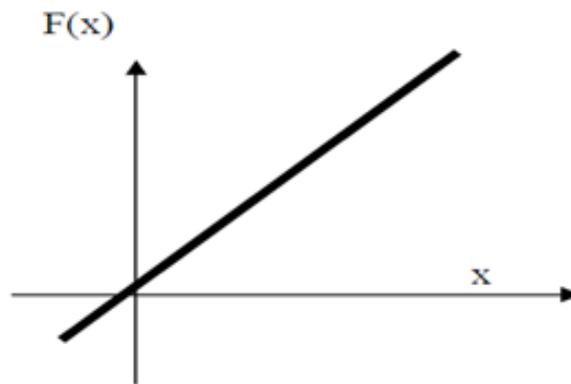


Figure IV.9. Fonction de transfert de type linéaire

L'utilisation des fonctions sigmoïdes nécessitent une normalisation des valeurs des entrées et sorties :

$$V_{DS}(V) : 0-4 \text{ V}, \quad V_{GS}(V) : 3-4 \text{ V}$$

$$T_j(^{\circ}\text{C}) : 25- 150^{\circ}\text{C}$$

$$I_D (A) : 0-45A$$

La base d'apprentissage contient 525 éléments tandis que la base de test 100 éléments.

La convergence est assurée pour une EQM de 0.0195.

IV.3. Organigramme des programmes élaborés

IV.3.1 Organigramme de la phase d'apprentissage

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. L'organigramme est donné par la figure IV.10 ci-dessous :

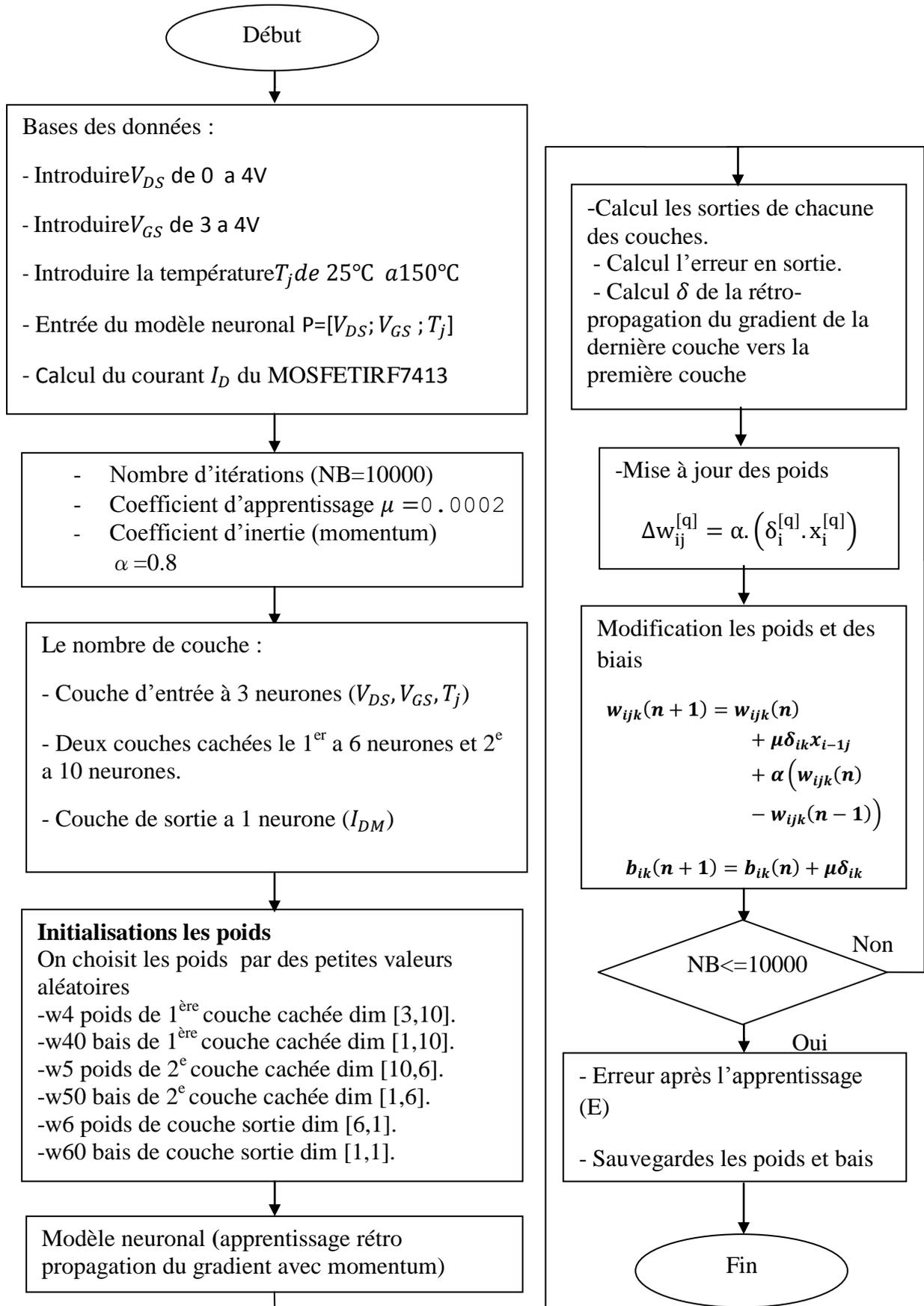


Figure IV.10. Organigramme de la phase d'apprentissage

Après plusieurs exécutions du programme, on choisit les valeurs des coefficients d'apprentissage μ (0.0002) et d'inertie (momentum) α (0.8) qui donnent l'erreur désirée avec un nombre d'itération fixe.

w_{ijk} est le poids de la connexion qui existe entre la $i^{\text{ème}}$ entrée et le $j^{\text{ème}}$ neurone cachée

IV.3.2. Programme de calcul du courant I_D du MOSFET IRF7413

Ce programme permet de calculer le courant de sortie I_D du MOSFET I_D en fonction V_{DS} , V_{GS} et T_j

On utilise le courant de sortie I_D du programme d'apprentissage (apprentissage supervisé) pour calculer l'erreur quadratique moyenne (EQM).

L'organigramme de calcul I_D est donné par la formule IV.11

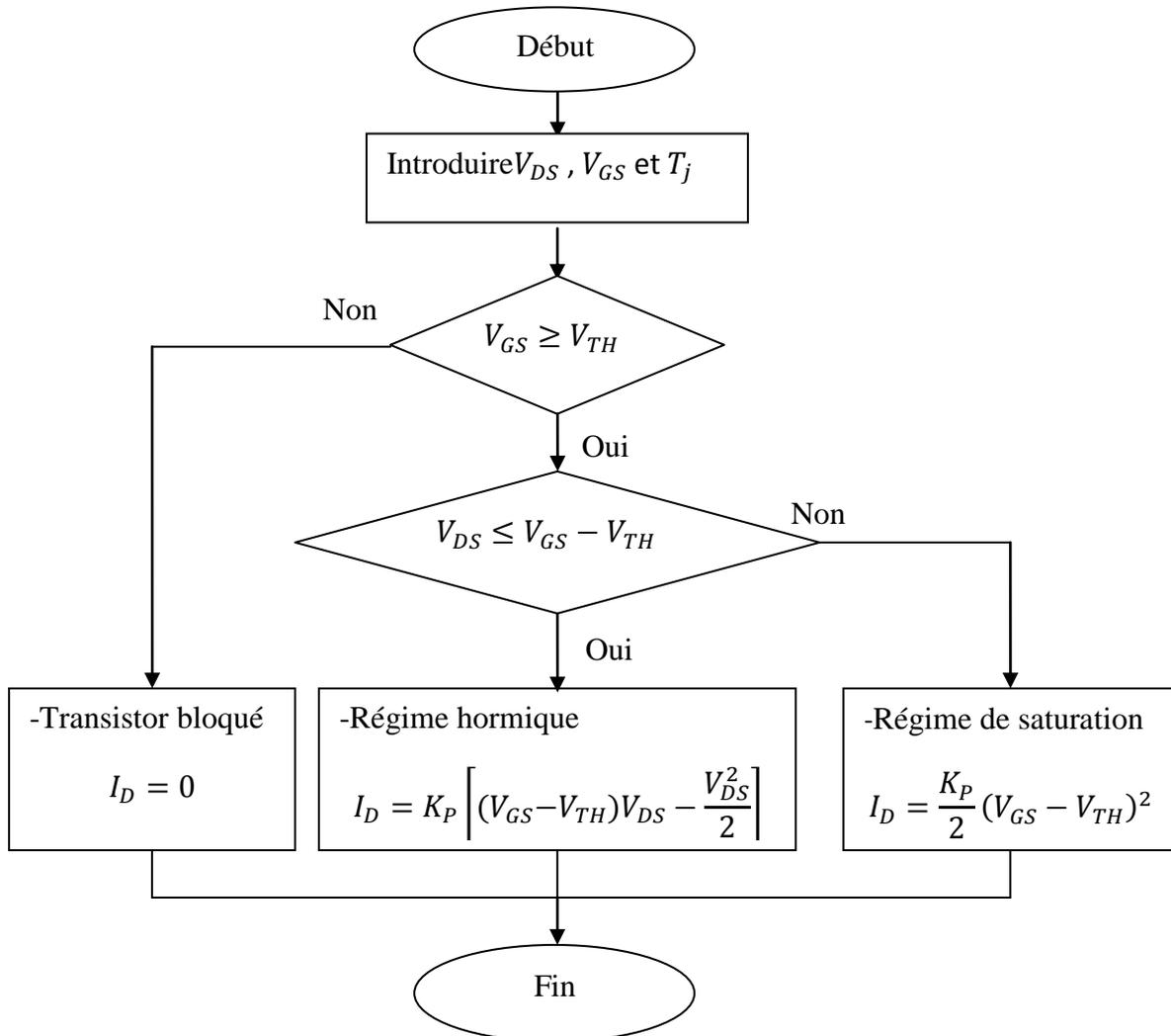


Figure IV.11. Organigramme de calcul du courant de drain I_D

IV.3.3. Programme de validation

L'évaluation de la capacité de généralisation du réseau est réalisée sur la base de validation, en utilisant des couples entrée-sortie n'appartenant pas à la base d'apprentissage; c'est-à-dire après avoir effectué l'apprentissage et avoir trouvé les poids du réseau nécessaires au calcul des différents paramètres ; Il faut donc estimer l'erreur sur la base de test.

On effectue une comparaison entre la base de données initiale et celle obtenue après l'apprentissage et ensuite tracer la courbe de courant modéliser I_{DM} pour différentes valeurs de V_{DS} , V_{GS} et T_j . L'organigramme de validation est donné par la figure IV.12 ci-dessous :

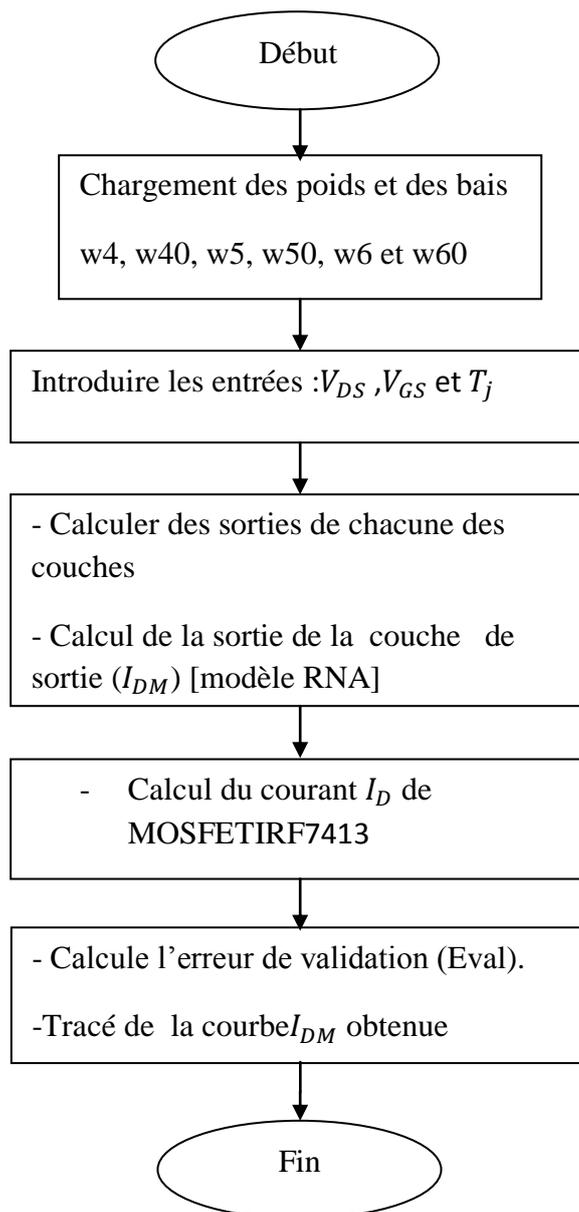


Figure IV.12. Organigramme de la validation du modèle

IV.4. Mesure la performance du modèle obtenu

La comparaison entre la base de données initiale et celle obtenue après l'apprentissage, utilisant la base de test, indique que notre modèle exprime fidèlement la variation de la réponse du MOSFET de puissance.

Les figures IV.13 et IV.14 présentent la performance du modèle RNA obtenue pour différentes valeurs de V_{GS} avec ($T_j = 25^\circ\text{C}$)

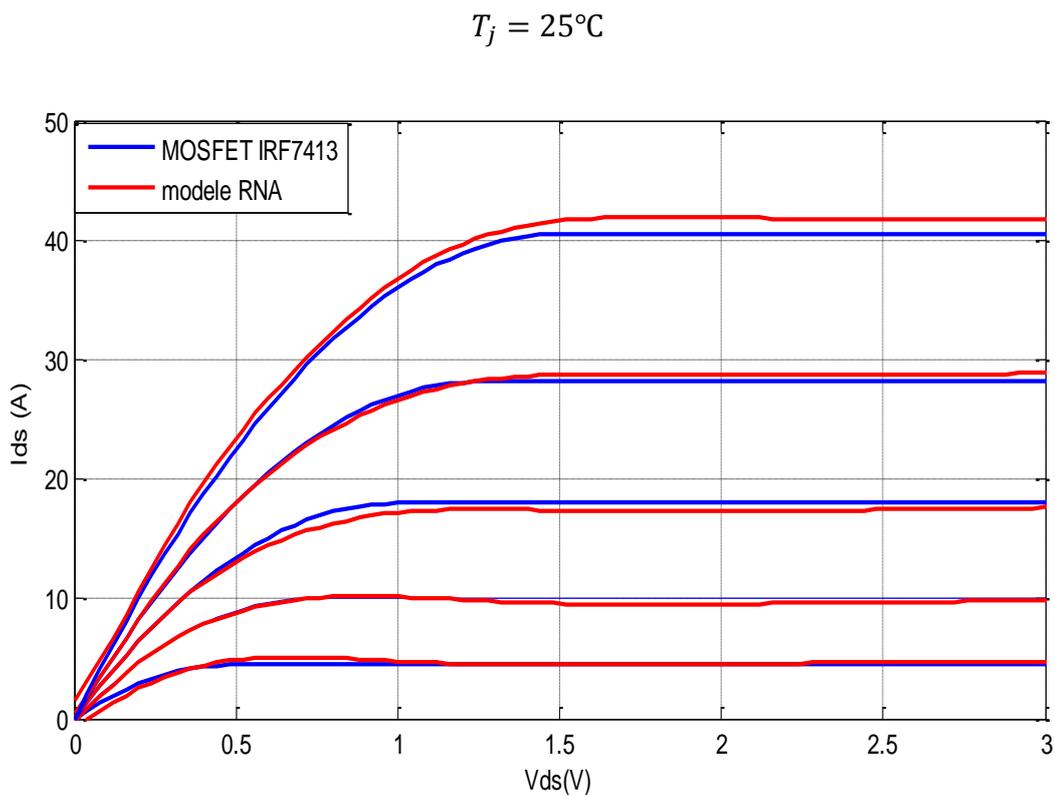


Figure IV.13. Performance du modèle neuronal (RNA)

Pour le régime ohmique on a pour une température $T_j = 25^\circ\text{C}$, l'évolution du courant de drain est illustrée par la figure IV.14 ci-dessous :

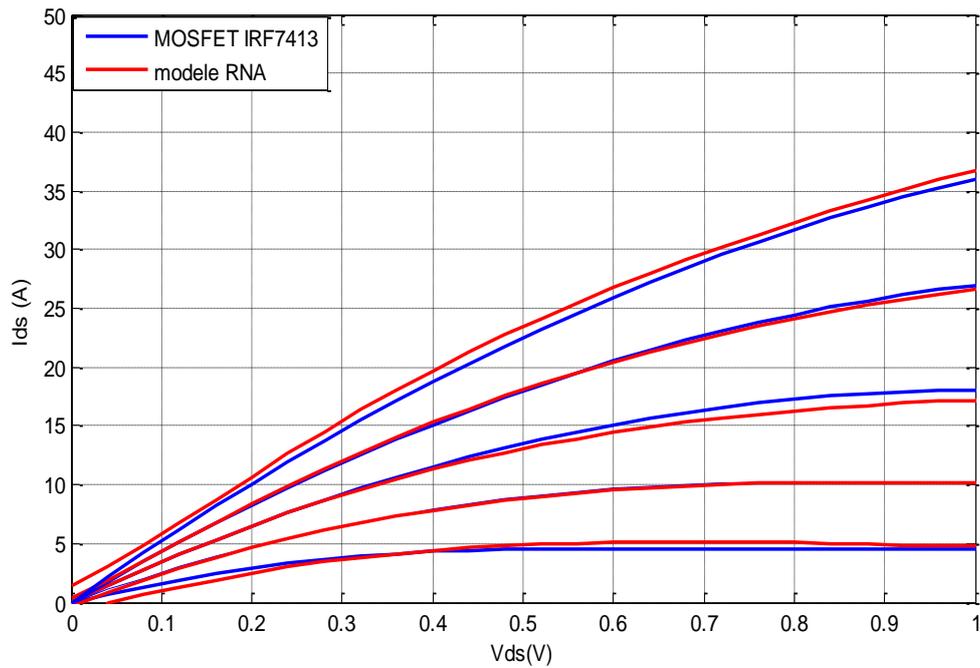


Figure IV.14. Performance du modèle RNA obtenu pour le régime ohmique

Les figures IV.15 et IV.16 présentent la performance du modèle RNA obtenue pour différentes valeurs de V_{GS} avec cette fois une température du transistor fixée à: $T_j = 50^\circ\text{C}$

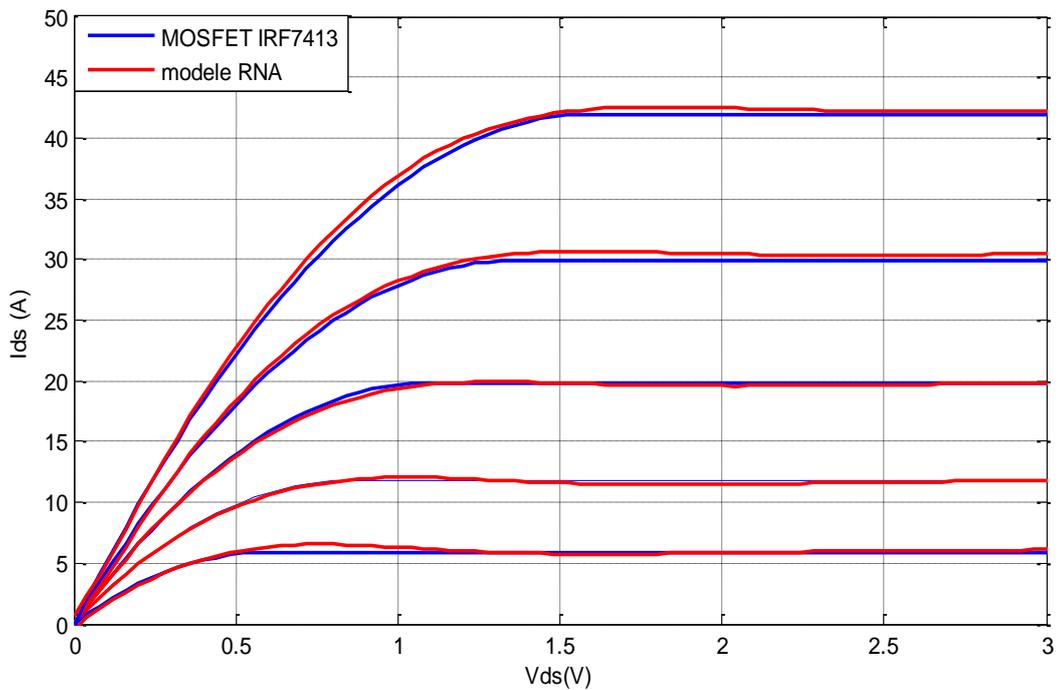


Figure IV.15. La performance du modèle RNA

Pour le régime ohmique on a pour une température $T_j = 50^\circ\text{C}$:

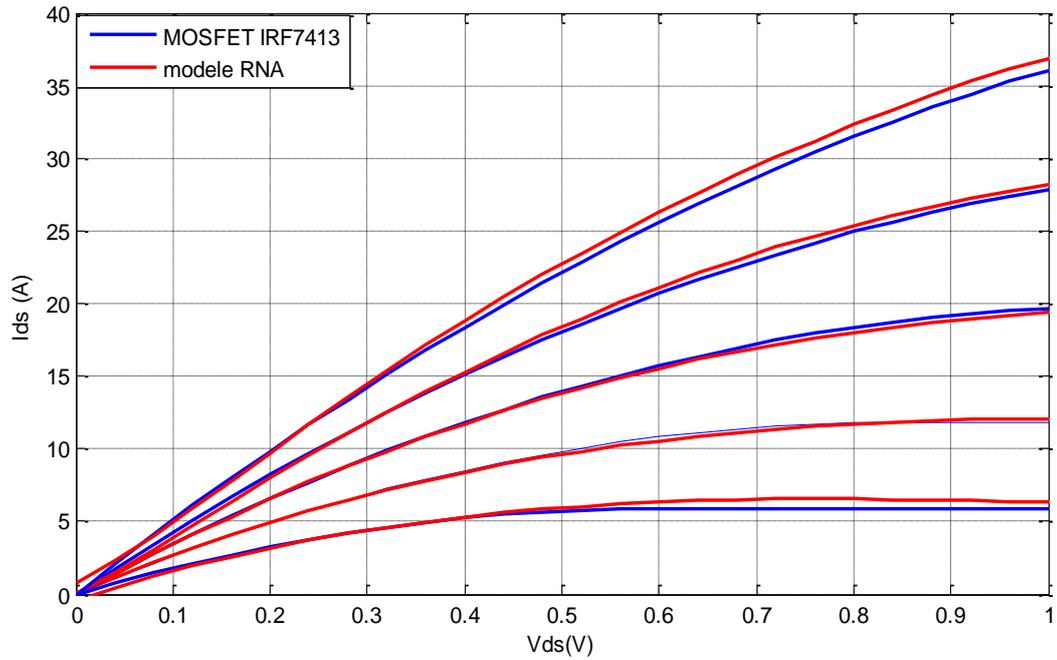


Figure IV.16.Performance du modèle RNA obtenu pour régime ohmique

Les figures IV.17 et IV.18 présentent la performance du modèle RNA obtenue pour différentes valeurs de V_{GS} avec une température de : $T_j = 75^\circ\text{C}$

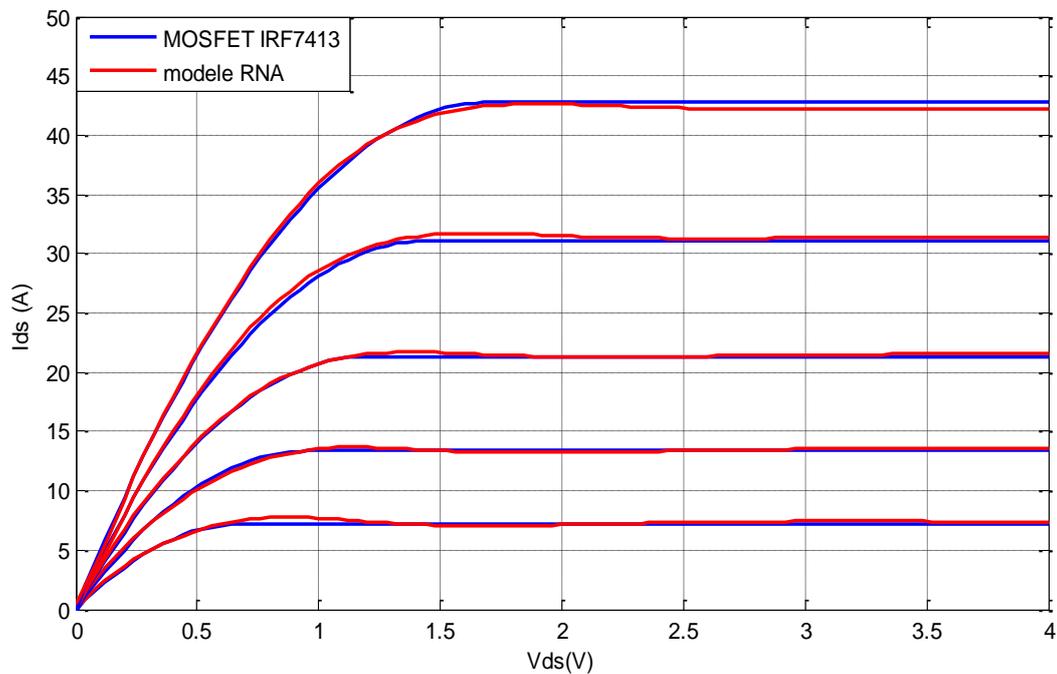


Figure IV.17.La performance du modèle RNA

Pour le régime ohmique on a pour $T_j = 75^\circ\text{C}$, la variation du courant de drain est illustrée par la figure IV.18

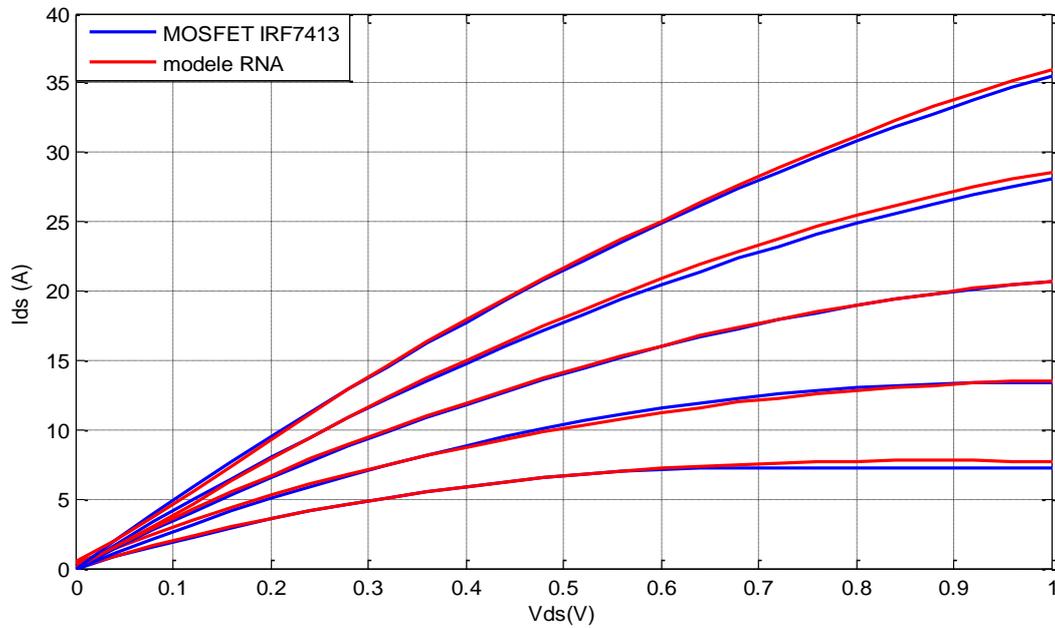


Figure IV.18. La performance du modèle RNA obtenu pour régime ohmique

La figure IV.19 présente la performance du modèle RNA obtenu pour différentes valeurs de V_{GS} avec ($T_j = 150^\circ\text{C}$)

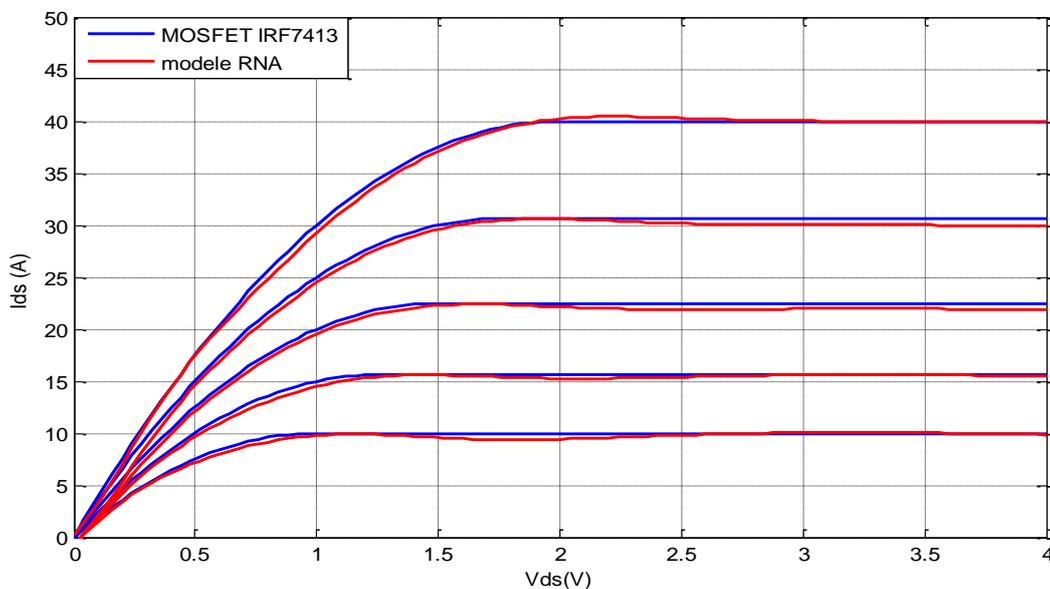


Figure IV.19. La performance du modèle RNA obtenu

En analysant les différentes courbes du courant drain source de sortie du transistor MOSFET on peut conclure que les résultats obtenus par le modèle neuronal sont en bon accord avec ceux obtenus théoriquement. Cette conclusion est vraie pour une variation de la tension grille source ou une variation de la température du transistor.

IV.5.Variation de la température T_j du MOSFET

Une augmentation de la température locale dans le composant entraîne la réduction de la mobilité et par conséquent la diminution de la vitesse des porteurs. En effet à mesure que la température augmente le courant de sortie décroît. De plus cet effet qui se présente à la fois dans les régimes de saturation et de quasi-saturation est seulement dépendant du niveau de puissance continue DC de l'entrée. La figure IV.20 montre l'effet de cette augmentation sur la caractéristique I_D du courant de drain.

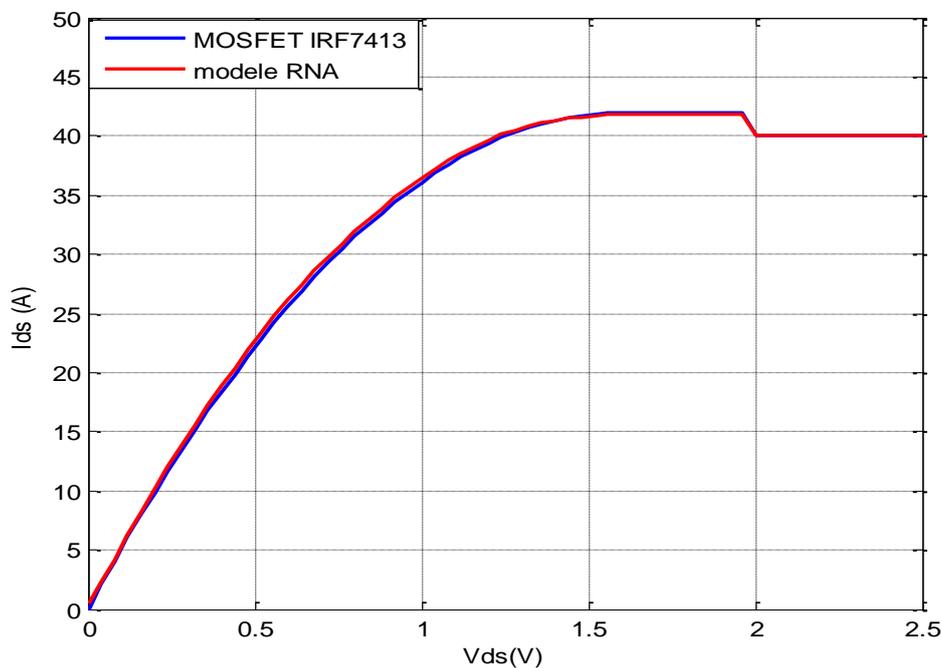


Figure IV.20. Modèle RNA pour une augmentation de température T_j de 50°C à 150°C

Le modèle neuronal suit de la même manière pour une diminution de la température. L'effet de cette diminution est illustré par la figure IV.21.

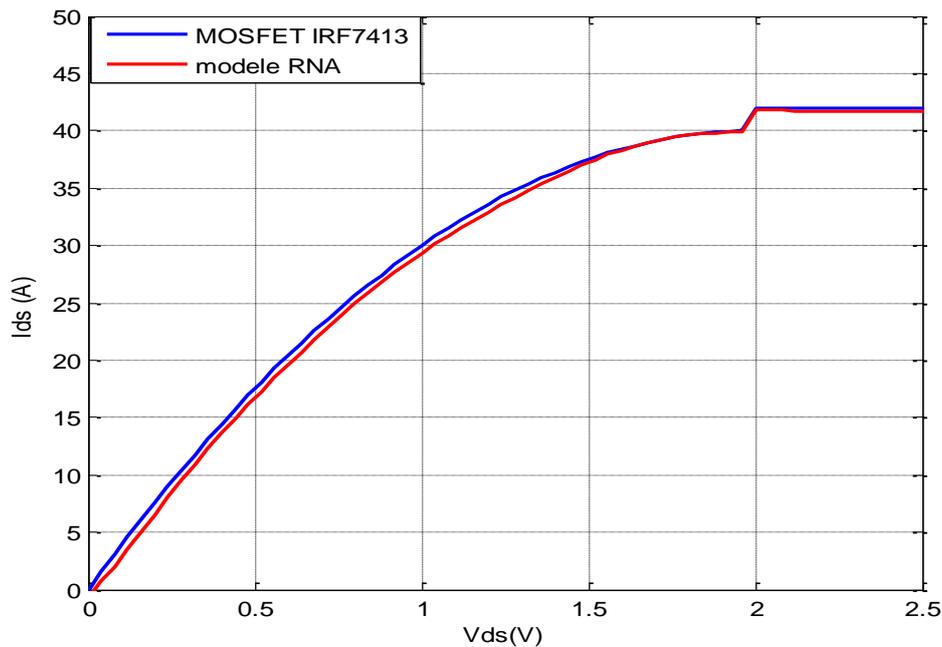


Figure IV.21. Modèle RNA pour une diminution de température T_j de 150°C a 50°C

On peut conclure qu'une diminution ou augmentation de la température n'influe pas sur l'allure de la caractéristique courant-tension.

IV.6. Effet de la variation rapide de la tension V_{GS}

- En faisant subir au modèle neuronal (RNA) une augmentation rapide de la tension V_{GS} de 3.3V a 4V pendant un court temps .Cette augmentation rapide produit une augmentation rapide courant du système. Le modèle RNA(ANN) élaboré se comporte parfaitement et suit convenablement ce changement.

Le comportement du modèle est similaire pour une diminution de cette tension. Les deux figures IV.22 et IV.23 illustrent ce phénomène.

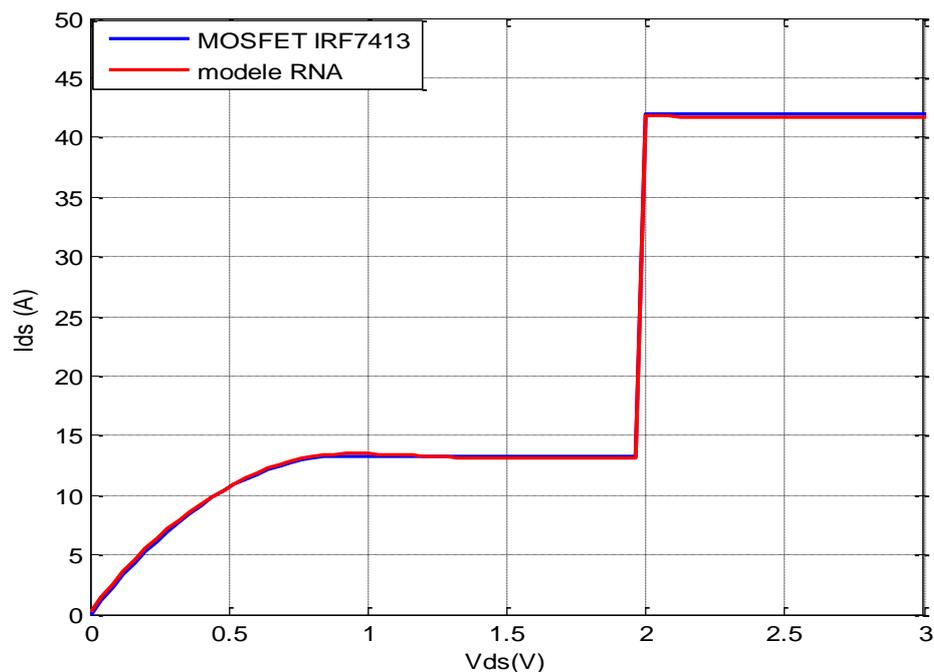


Figure IV.22. Le modèle RNA pour une augmentation rapide de la tension V_{GS} de 3.3V a 4V

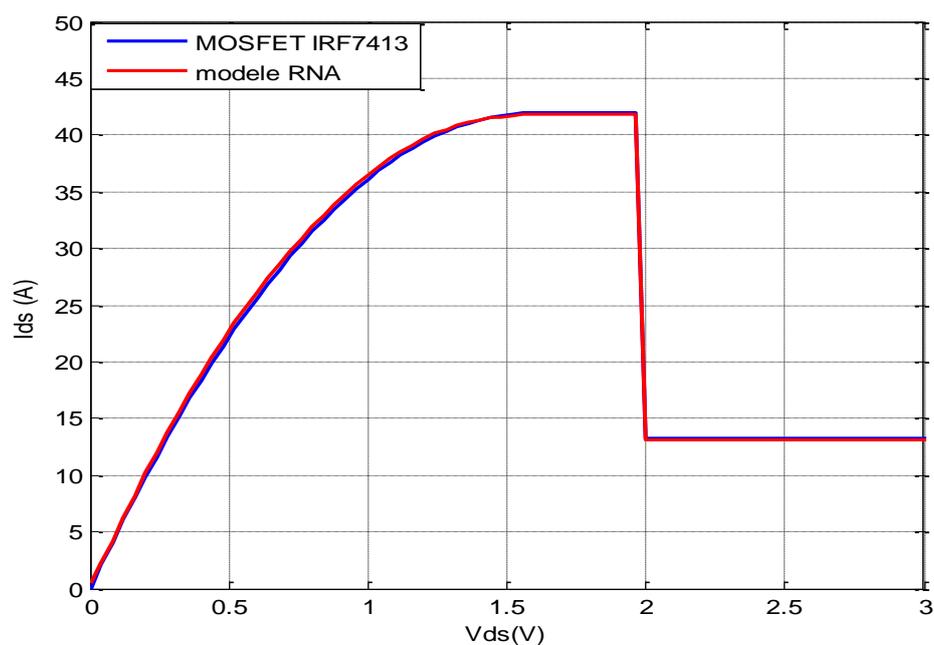


Figure IV.23. Le modèle RNA pour une diminution rapide de la tension V_{GS} de 4V a 3.3V

IV.7. Modélisation neuronale de la conductance de sortie du transistor MOSFET de puissance

La conductance de sortie joue un rôle très important dans le fonctionnement du transistor. En effet le réseau de caractéristiques I_D en fonction de V_{DS} donne des informations sur le courant de drain maximum que peut délivrer le composant et sur sa tension de pincement. On a jugé utile d'utiliser une approche neuronale pour la modélisation de ce paramètre.

La conductance extrinsèque du canal (g_{DS}) est donnée par :

$$g_{DS} = \frac{\partial I_D}{\partial V_{DS}} = \begin{cases} K_P [V_{GS} - V_{TH} - V_{DS}] & \text{en régime ohmique} \\ 0 & \text{en régime de saturation} \end{cases} \quad (\text{IV.10})$$

Ou

K_P : Transconductance du MOSFET ($A.V^{-2}$),

V_{GS} : Tension Grille-Source(V)

V_{TH} : Tension de seuil du MOSFET(V)

V_{DS} : Tension Drain-Source(V)

Lors des mesures de ces paramètres, le transistor est soumis aux lentes variations : régime quasi-statique, souvent dit "statique".

La structure du model neuronal est la suivante :

- Un neurone pour la couche de sortie.
- Deux entrées (V_{DS} et V_{GS}).
- Nombre de couches cachée : 2 ; La première couche contient 6 neurones, la deuxième couche contient 4 neurones.
- Les fonctions de transfert utilisées sont des sigmoïdes pour les deux couches cachées et une fonction linéaire pour la couche de sortie

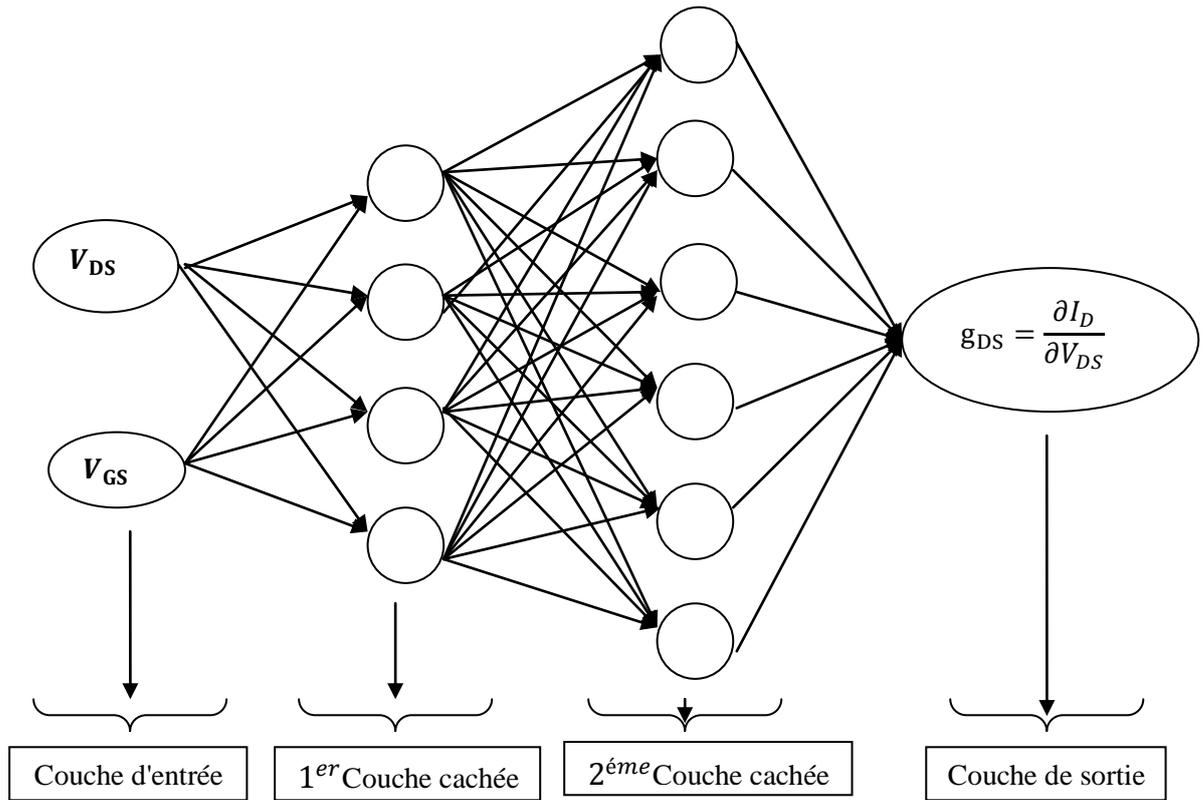


Figure IV.24. Structure du modèle neuronal de la conductance ($\frac{\partial I_D}{\partial V_{DS}}$)

L'utilisation des fonctions sigmoïdes nécessitent une normalisation des données, on a donc pour:

Les deux entrées : $V_{DS}(V)$: min=0V max=4V
 $V_{GS}(V)$: min=2V max=4.5V

La sortie : $g_{DS} = \frac{\partial I_D}{\partial V_{DS}} (\Omega^{-1})$: min=0 Ω^{-1} max=50 Ω^{-1}

La base d'apprentissage contient 246 éléments et la base de test contient 41 éléments. L'erreur quadratique moyenne EQM est de 0.0043 avec une erreur désirée de 10^{-3} .

La structure est le perceptron multicouche de type Feedforward entraîné par la rétro propagation du gradient.

Le modèle neuronal élaboré est illustré par les figures IV.25, 26 et 27 qui montrent un bon accord entre ce modèle neuronal et le modèle théorique.

Pour $V_{GS}=2.5V$

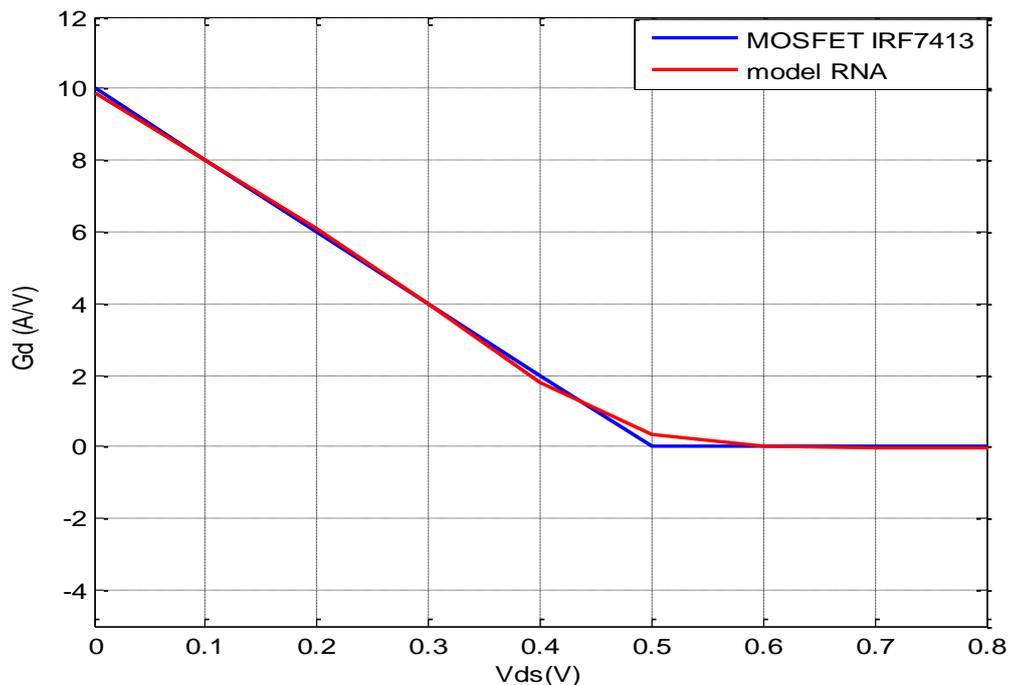


Figure IV.25. Caractéristique $I_{DS} = f(V_{DS})$ pour $V_{GS}=2.5V$

Pour $V_{GS}=3.5V$

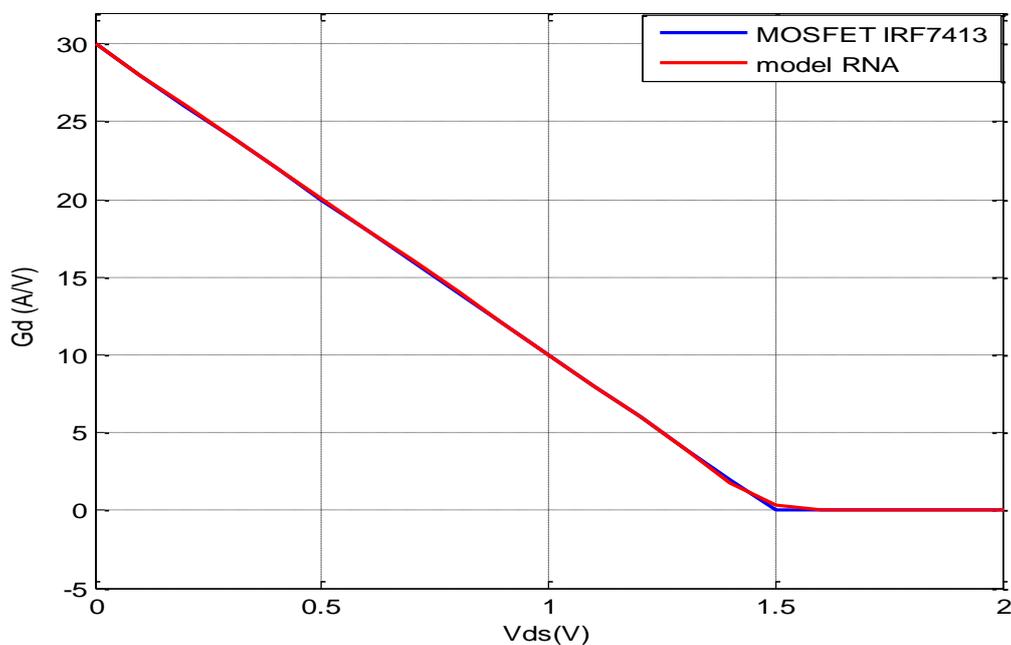


Figure IV.26. Caractéristique $I_{DS} = f(V_{DS})$ pour $V_{GS}=3.5V$

Pour $V_{GS}=4.5V$

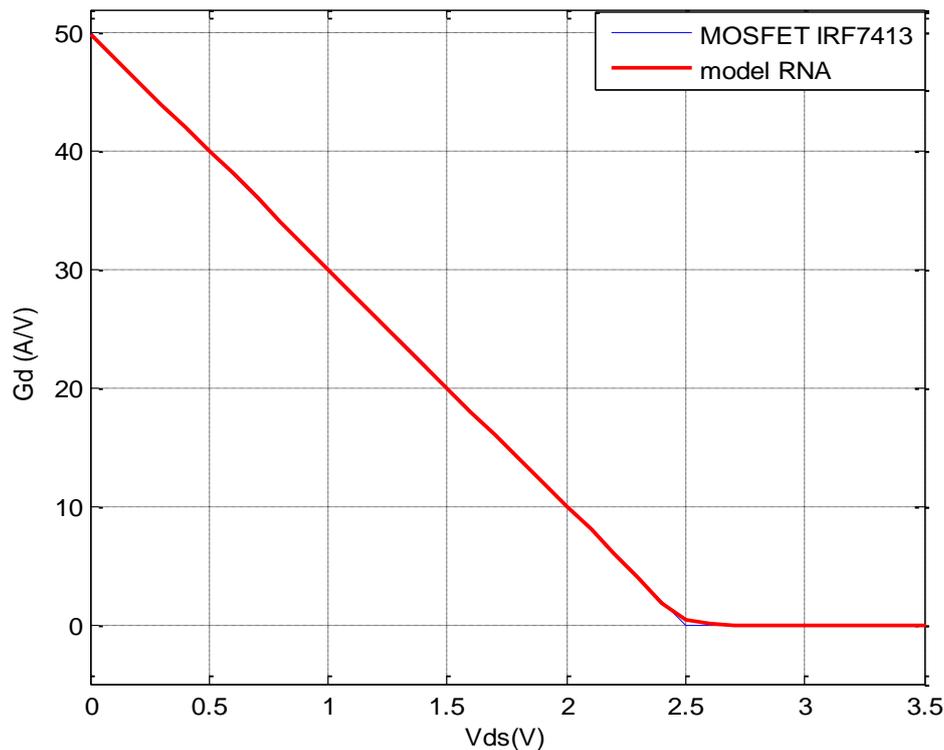


Figure IV.27. Caractéristique $g_{DS} = f(V_{DS})$ pour $V_{GS}=4.5V$

IV.8. Modèle neuronal (RNA) du MOSFET de puissance sur simulink matlab

Le transistor MOSFET sous Simulink possède 20 boîtes dont la répartition de leurs fonctions est la suivante :

- Les trois boîtes V_{DS} , V_{GS} et T_j sont utilisées pour la normalisation des entrées.
- Les sept boîtes O1, O2, O3, O4, O5 et O6 sont utilisées pour la première couche cachée.
- Les six boîtes O7, O8, O9, O10, O11, O12, O13, O14, O15 et O16 sont utilisées pour la deuxième couche cachée.
- La boîte O17 est utilisée pour la couche de sortie.

La figure IV.28 représente l'implémentation du model neuronal.

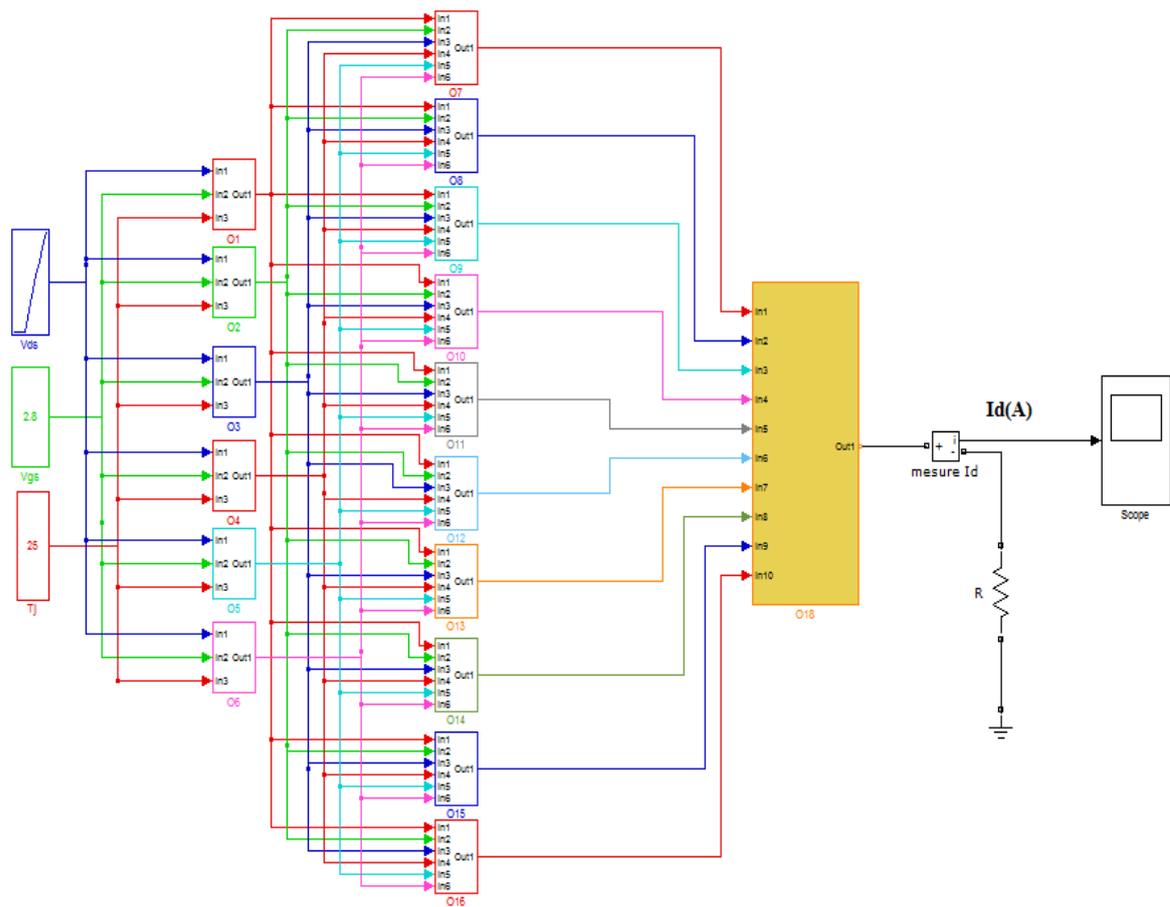


Figure IV.28. Implantation du modèle RNA sur Simulink matlab

Le schéma équivalent de ce modèle est illustré par la figure IV.29 ci-dessous :

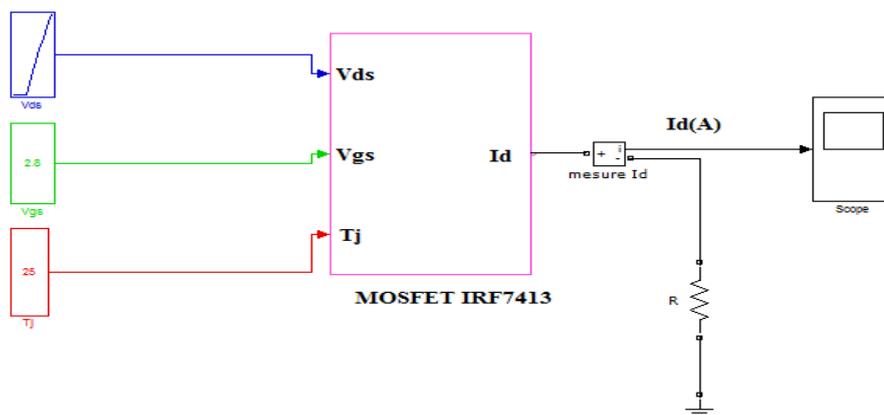


Figure IV.29. Schéma équivalent du modèle RNA sur Simulink matlab

Chacune des boites O1 à O17 est l'équivalent à un neurone caractérisé par son équation représentative, l'équation de O1 par exemple est donnée par la formule IV.11.:

$$f = 1 / \left(1 + \exp \left(- (W_{11} * V_{DS} + W_{12} * V_{GS} + W_{13} * T_j - b_1) \right) \right) \quad (IV.11)$$

La forme exponentielle dans l'équation précédente est due au choix de la fonction de transfert sigmoïde (*Logsig*) dans la première couche cachée, b_1 est le premier biais pour la première couche cachée ; Dans la matrice du biais « b_{ij} », W_{11} , W_{12} et W_{13} sont respectivement du premier au troisième poids pour la première couche cachée, dans la matrice des poids « W_{ij} ».

Le schéma interne de la composition de la boite O1 est donné par le schéma de la figure IV.30 :

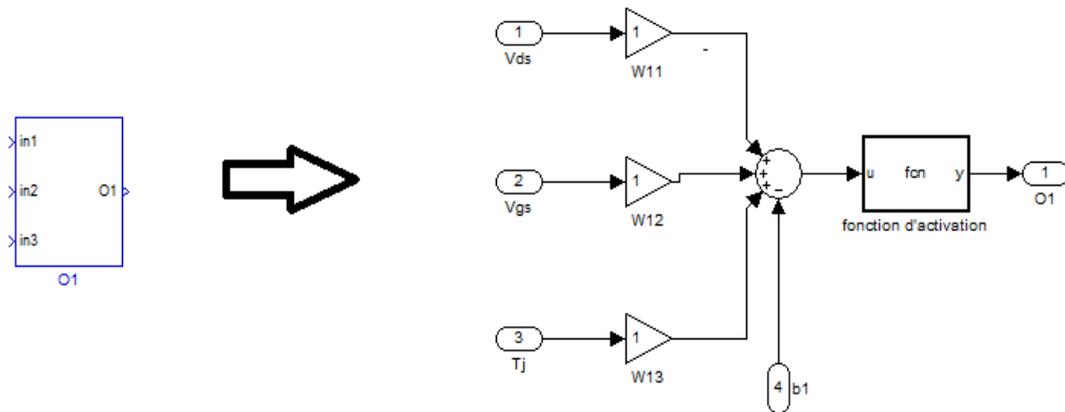


Figure IV.30. Schéma interne de boîte O1

La caractéristique $I_D = f(V_{DS})$ du modèle RNA en utilisant le logiciel Simulink à une température $T_j = 25^\circ\text{C}$ est représentée dans la figure IV.31

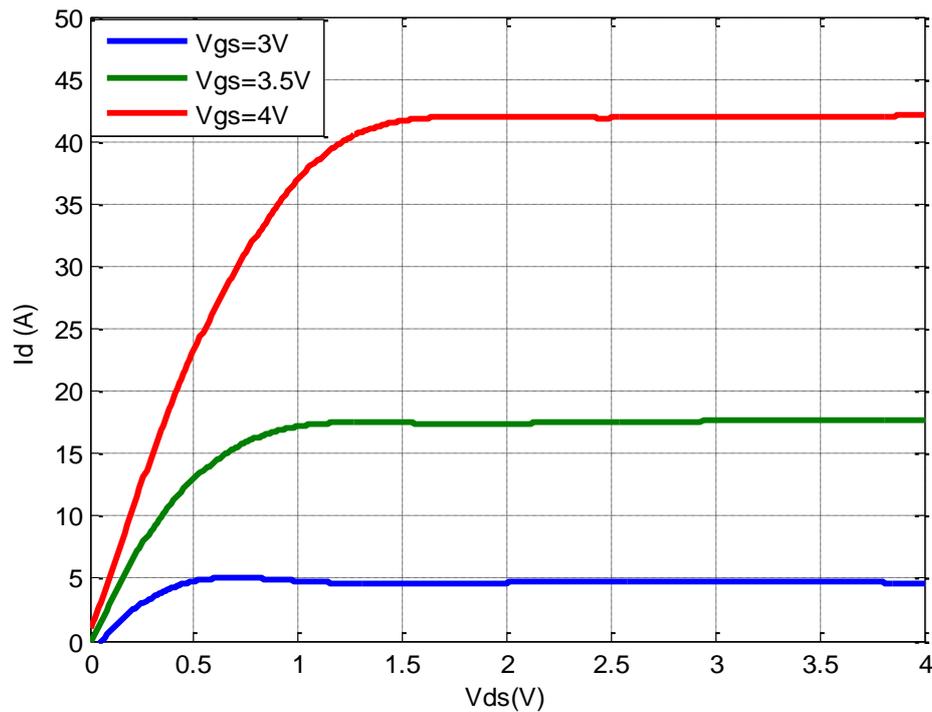


Figure IV.31.Caractéristique $I_D = f(V_{DS})$ du modèle neuronal (RNA)

IV.9 Conclusion

On a utilisé dans ce travail un nouveau type de modélisation pour les transistors MOSFET de puissance, les réseaux neurones en l'occurrence.

Ce type de modélisation est précis et a donné de bons résultats et en bon accord avec le modèle théorique. On a constaté une bonne compatibilité entre les deux réseaux de caractéristiques.

On peut conclure qu'en utilisant une approche neuronale on peut modéliser une fonction non linéaire et sa dérivée.

Nous avons utilisé une programmation sous Matlab et l'autre sous Simulink

Conclusion générale

Conclusion Générale

L'objectif primordial recherché dans ce travail est d'utiliser l'approche neuronale pour modéliser les transistors MOSFET de puissance appliqués aux dispositifs radiofréquences. Le choix de cette technique est justifié par le fait que les méthodes basées sur la physique du composant ou sur un schéma électrique équivalent présentent certains inconvénients.

La modélisation par réseaux de neurones est un outil très efficace pour la conception de modèles non linéaires appliqués à des composants de puissance. Les réseaux de neurones présentent des avantages considérables tels que l'apprentissage et la généralisation.

On a construit un modèle neuronal pour le courant de drain du transistor MOSFET de puissance, vu que ce paramètre représente la non linéarité la plus importante. A cet effet on a utilisé la structure multicouche de type Feedforward entraîné par l'algorithme de la rétro propagation. Dans ce travail nous avons seulement utilisé un réseau multicouche non récurrent à cause de sa structure simple, son utilisation dans une grande variété de problèmes et de son aptitude à réaliser une transformation arbitraire des entrées vers les sorties.

On doit noter qu'on a fait plusieurs apprentissages afin de trouver le modèle présentant une erreur quadratique moyenne la plus faible. Ce modèle neuronale a donné des résultats satisfaisants. On a utilisé le logiciel Matlab, pour effectuer les différentes simulations pour montrer l'efficacité des réseaux de neurones particulièrement dans le domaine de la modélisation des composants radiofréquences.

Bibliographie

Bibliographie

- [1] G.DREYFUS " **les réseaux de neurones**" Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI), Laboratoire d'Electronique.
- [2] CHEKROUN Soufyane " **Commande neuro-floue sans capteur de vitesse d'une machine asynchrone triphasée**" Mémoire de Magister, Ecole Normale Supérieure d'Enseignement Technologique d'Oran, p.7-8, 25 Octobre 2009.
- [3] Sami BADRA "**Utilisation de la méthode de l'équation intégrale de champ électrique et des réseaux de neurones artificiels pour l'étude d'une antenne imprimée bicouche fonctionnent en mode TM**" Mémoire de Magister, Université Batna, pp 24-26,2011.
- [4] Elassd CHEBAKI" **Modélisation neuronale du transistor GAAMOSFET nanométrique**" Mémoire de Magister, Université Batna, 2010.
- [5]" **Les réseaux de neurones**" p.44-46,
- [6] M.IBNNKAHLA "**Application of neural networks to digital communication** ", Signal Processing 80, pp 11585-1215,2000.
- [7] B.MULGREW "**Applying radial basis function net works**" IEEE Signal Processing Magazine, pp.50-65, Mar.1996.
- [8] S.HAYKIN "**Neural Networks: A comprehensive Foundation 2^{and} Ed**,New York: Macmillan College Publishing",1998.
- [9]Vania ANDRONOVA "**Utilisation de données météo et des réseaux de neurones pour la prédiction de vitesses de vent**" Mémoire de Master, Université de Corse – Pasqual Paoli, pp.29-32, juillet 2006,
- [10] Zahra ZERDOUMI " **Application des réseaux de neurones artificiels a la poursuite des non linéarités fluctuantes des systèmes satellitaires**" Mémoire de Magister, Université Mohamed Boudiaf M'sila, 21 / 06 / 2006.
- [11] Yann MORERE "**Identification par réseaux de neurones**" Mémoire de D.E.A. Université de Valenciennes et du Hainaut-Cambrésis, 1996.

- [12] Kary FRÄMLING "**Les réseaux de neurones comme outils d'aide à la décision floue**" Mémoire de Rapport de D.E.A (Ingénierie Informatique), Ecole Nationale Supérieure des Mines de Saint-Etienne, France, Juillet 1992, pp.18-19.
- [13] "**Introduction aux réseaux de neurones**" Mémoire de Master ,2006.
- [14] Toufik BENDIB" **Modélisation et simulation du transistor DG MOSFET en utilisant les Algorithmes Génétiques**" Mémoire de Magister, Université Batna, 2008.
- [15] A. Galadi "**Etude de propriétés physiques et nouvelle modélisation spice de Transistor FLIMOS de puissance** " Mémoire de Doctorat de l'Université de Toulouse Délivre par l'Université de Toulouse 3- Paul Sabatier 2008.
- [16] W. Shockley "**Problems related to p-n junctions in silicon**" Solid State Electronics, vol. 2, no. 1, pp. 35–67, 1961.
- [17] MESSAADI Lotfi "**Macro modélisation du transistor MOSFET**" Mémoire de Magister, Université Batna.
- [18] Nidhal ABDELMALEK" **Etude et modélisation du transistor VSG MOSFET nanométrique** " Mémoire de Magister, Université Batna, 2010.
- [19] Dorothée MULLER "**Optimisation des potentialités d'un transistor LDMOS pour l'intégration d'amplification de puissance FR sur silicium**" Université de Limoges, Ecole Doctorale, Science - Technologie - Santé Faculté des Science et Techniques, 12 Octobre 2006.
- [20] Anouar Essadate AOUF "**Modélisation et simulation du Nano-transistor de puissance**" Mémoire de Magister, Université Batna, 2009.
- [21] ahriche AIMAD" **Réalisation d'une commande numérique avec différentes stratégies et identification par les réseaux de neurones artificielles, application et réglage de température** " Mémoire de Magister, Université Boumerdes ,2008.
- [22] Abdelaziz AOUCHE "**Rejection des perturbations dans les systèmes non linéaires : étude comparative**" Mémoire de Magister, Université Mohamed Boudiaf M'sila, 2006.

- [23] P. Masson "**Etude par pompage de charge et par mesures de bruit basse fréquence de transistors MOS à oxytriture de grille ultra-minces**" Mémoire de Doctorat, INSA Lyon, France, 1999.
- [24] S. M. Sze "**Physics of Semiconducteur Devices**" Wiley & Sons, New York, 2nd edition, 1981.
- [25] A. Galadi "**Etude de propriétés physiques et nouvelle modélisation spice de transistor FLIMOS de puissance**" Mémoire de Doctorat de l'université de Toulouse délivré par l'université de Toulouse 3- Paul Sabatier 2008.
- [26] H.R Chang "**Numerical and Experimental Comparison of 60V vertical Double-Diffused MOSFETs with a Trench-Gate structure**" Solid-State Electronics, Vol. 32, No 3,P 247-251 , 1989.
- [27] T.D. Gilio "**Etude de la fiabilité porteurs chauds et des performances des technologies CMOS 0.13 μm - 2nm**" Mémoire de Doctorat, Université de Provence, France, 2006.
- [28] D. E. Rumelhart and J. L. Mc Clelland. "**Parallel Distributed Processing: Exploration in The Micro Structure of Cognition, MIT Press**" vol.1, pp.318-362, Cambridge, 1986.
- [29] S. Siu, G.J. Gibson,and C.F.N. Cowan" **Decision feedback equalisation using neural network structures and performance comparison with standard architecture**" IEE Proceedings, Vol.137, 1990.
- [30] M. Faiza Belhachat "**Commande neuro-floue d'un hacheur MPPT**" Mémoire de Magister, Ecole Nationale Polytechnique, Alger, juin 2007.
- [31] Ferhat Lahouazi" **Mise en œuvre d'une stratégie de commande neuro floue : Application à un pendule inversé**" Mémoire de Magister, Université Mouloud mammeri,Tizi-Ouzou, 2011.
- [32] k.kara "**Application des réseaux de neurones à l'identification des systèmes non linéaire**" Mémoire de magister Université de Constantine ,1995.

[33] GUITANI .ISSAM "**commande adaptation neuronale par retour de sortie des systèmes non linéaire**" Mémoire de magister, université Mentouri de Constantine, 2007

[34] AHRICHE.Aimad "**réalisation d'une commande numérique avec différentes stratégies de commande et identification par réseaux de neurones artificiels application au réglage de température** " Mémoire de magister université Boumerdes, 2008.

[35] Zidani Ghania "**Exécution de trajectoire pour robot mobile d'intérieur –réseaux de neurones-**" Mémoire de magister, Université de Batna, 2009

Modélisation des transistors FET appliqué aux dispositifs radiofréquences en utilisant l'approche neuronale

ملخص

الموضوع الرئيسي من هذا العمل هو نمذجة الترانزستور ذو التأثير المجالي في نهج شبكة العصبونات الاصطناعية. والهدف من ذلك هو اعطاء نموذج لتيار المخرج ، الذي يمثل دالة غير خطية من الترانزستور الأكثر أهمية، ثم تطوير نموذج العصبونات الاصطناعية على أساس نموذج متعدد الطبقات، وتدريبه على خوارزمية الرجعية والذي أعطى نتائج جيدة بالمقارنة مع النموذج النظري.

كلمات البحث: الشبكات العصبية، النمذجة، المحاكاة، خوارزمية الرجعية و نموذج متعدد الطبقات

Résumé

Le thème essentiel de ce travail est la modélisation des transistors MOSFET de puissance par l'approche neuronale. Le but recherché est la modélisation du courant de drain qui représente le non linéarité la plus importante du transistor. Le modèle neuronale élaboré est basé sur le perceptron multicouche, entraîné par la retro propagation du gradient avec momentum et a donné de bons résultats en comparaison au modèle théorique.

Mots clés : réseaux de neurones, modélisation, simulation, algorithme de la retro propagation, perceptron multicouche

Abstract

The main theme of this work is the modeling of power MOSFETs in the neural approach. The aim is to model the drain current, which represents the largest non-linearity of the transistor. The neural model is developed based on the multilayer perceptron, trained by the spread of retro gradient with momentum and gave good results in comparison with the theoretical model.

Keywords: neural networks, modeling, simulation, retro propagation algorithm, multilayer perceptron.