



**Εθνικό Μετσόβιο Πολυτεχνείο**

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής  
Εργαστήριο Σχεδίασης Μικροηλεκτρονικών Κυκλωμάτων

**Αναλογικά Ολοκληρωμένα Κυκλώματα  
Ταξινομητών Χαμηλής Κατανάλωσης  
Βασισμένων στο Gaussian Mixture Model για  
Βιο-ιατρικές Εφαρμογές**

**Διπλωματική Εργασία**

του

**Γεωργίου Γέννη**

**Επιβλέπων:** Παύλος Π. Σωτηριάδης  
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2022



## Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής  
Εργαστήριο Σχεδίασης Μικροηλεκτρονικών Κυκλωμάτων

# Αναλογικά Ολοκληρωμένα Κυκλώματα Ταξινομητών Χαμηλής Κατανάλωσης Βασισμένων στο Gaussian Mixture Model για Βιο-ιατρικές Εφαρμογές

Διπλωματική Εργασία

του

Γεωργίου Γέννη

Επιβλέπων: Παύλος Π. Σωτηριάδης  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11<sup>η</sup> Φεβρουαρίου  
2022:

.....  
Παύλος Π. Σωτηριάδης    Νεκτάριος Κοζύρης    Γεώργιος Ματσόπουλος  
Καθηγητής                    Καθηγητής                    Καθηγητής  
Ε.Μ.Π.                            Ε.Μ.Π.                            Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2022

.....  
**Γεώργιος Γέννης**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών, Ε.Μ.Π.

Copyright © Γεώργιος Γέννης, 2022.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.



# Περίληψη

Ο σκοπός αυτής της διπλωματικής είναι η ανάπτυξη αναλογικών ταξινομητών βασισμένων στη Γκαουσιανή συνάρτηση. Συγκεκριμένα, θα περιγραφούν δύο γενικές αρχιτεκτονικές: μια ενός Bayesian ταξινομητή με Γκαουσιανές συναρτήσεις πυκνότητας πιθανότητας και μία ενός Gaussian Mixture Model ταξινομητή. Αυτές οι καθαρά αναλογικές αρχιτεκτονικές είναι προσαρμόσιμες σε μεγάλο εύρος εφαρμογών. Ο τρόπος εκπαίδευσης, λειτουργίας και προσαρμογής τους περιγράφεται αναλυτικά. Παράλληλα, παρουσιάζονται και τα κυκλώματα που χρησιμοποιούνται στις ευρύτερες αρχιτεκτονικές. Στην διπλωματική θα παρουσιαστούν 4 διαφορετικές αρχιτεκτονικές Bump κυκλωμάτων και 2 αρχιτεκτονικές Winner-Take-All κυκλωμάτων. Τέλος, για την επιβεβαίωση των κυκλωμάτων έχουν χρησιμοποιηθεί ένα τεχνητό και τέσσερα πραγματικά σετ δεδομένων καθώς και έχουν γίνει οι απαραίτητες προσομοιώσεις για την ευαισθησία των κυκλωμάτων.

**Λέξεις Κλειδιά:** αναλογικές αρχιτεκτονικές, Γκαουσιανά Μοντέλα Μίξης, κυκλώματα Γκαουσιανής συνάρτησης, Μπεϋζιανοί ταξινομητές, περιοχή υπο-κατωφλίου, πλήρως προσαρμόσιμη υλοποίηση, σχέδια πολύ χαμηλής κατανάλωσης, ταξινόμηση σε τσιπ, layout, Winner-Take-All κυκλώματα.



# Abstract

The purpose of this thesis is the development of analog classifiers based on the Gaussian function. Specifically, two general architectures will be discussed: one of a Bayesian classifier with a Gaussian probability density function and one of a Gaussian Mixture Model-based classifier. These fully analog architectures are adaptable to a wide range of applications. The training, the operation and the adaptation, to different applications, procedures are described in detail. Additionally, the basic building circuits used for these architectures are presented. In this thesis 4 different Bump circuit architectures and 2 Winner-Take-All circuit architectures will be analyzed. A toy and 4 real-world classification problems have been used to confirm the proposed architectures as well as the sensitivity of the circuits.

**Keywords:** analog hardware architecture, Bayesian Classifier, Gaussian Mixture Model, Gaussian function circuit, fully tunable implementation, layout, on-chip classification, subthreshold region, Ultra-low power design, Winner-Take-All circuit.





# Ευχαριστίες

Η ολοκλήρωση της παρούσας Διπλωματικής Εργασίας σηματοδοτεί και την ολοκλήρωση των προπτυχιακών μου σπουδών. Στο σημείο αυτό θα ήθελα να ευχαριστήσω την οικογένειά μου, τους φίλους μου και τους ανθρώπους που στάθηκαν δίπλα μου σε αυτό το κομμάτι της ζωής μου.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Παύλο Πέτρο Σωτηριάδη για την πολύτιμη καθοδήγησή του, τις συμβουλές του καθώς και το πραγματικό ενδιαφέρον που επέδειξε για μένα ως φοιτητή και για το αντικείμενο της Διπλωματικής μου Εργασίας. Μου δόθηκε έτσι η πολύτιμη ευκαιρία να ασχοληθώ με ένα ιδιαίτερα ενδιαφέρον ερευνητικό θέμα σε ένα πολύ υποστηρικτικό περιβάλλον.

Επίσης, θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Βασίλειο Αλιμήση για την ανεκτίμητη συνεισφορά του. Η Διπλωματική αυτή Εργασία όπως και οι σχετικές δημοσιεύσεις σε μεγάλο μέρος τους οφείλονται σε αυτόν διότι αποτελούν το αποτέλεσμα της καθημερινής μας συνεργασίας. Οπότε όσες ευχαριστίες και να του γράψω θα είναι πραγματικά λίγες. Θέλω εξίσου να ευχαριστήσω τον διδάκτορα Χρήστο Δήμα, τους υποψήφιους διδάκτορες Κωνσταντίνο Τουλούπα και Νικόλαο Τέμενο και τον διπλωματούχο ηλεκτρολόγο μηχανικό Μάριο Γουρδουπάρη. Μαζί τους συγκροτήθηκε μια ξεχωριστή ομάδα με άριστη συνεργασία η οποία συνέβαλε καθοριστικά στην ολοκλήρωση της εργασίας αυτής.

Γέννης Γιώργος,  
Φεβρουάριος 2022



# Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Ευρετήριο Εικόνων	13
Κατάλογος Πινάκων	15
<b>1 Εισαγωγή</b>	<b>17</b>
1.1 Analog Computing . . . . .	17
1.2 Μηχανική Μάθηση με Αναλογικά Κυκλώματα . . . . .	18
1.3 Gaussian Mixture Models . . . . .	19
<b>2 Βασικά κυκλώματα</b>	<b>21</b>
2.1 Bump κυκλώματα . . . . .	21
2.1.1 Υπόβαθρο . . . . .	21
2.1.2 Cascode Bump . . . . .	23
2.1.3 Neuron Bump . . . . .	25
2.1.4 Άλλες Αρχιτεκτονικές . . . . .	28
2.2 Winner-take-all κυκλώματα . . . . .	29
2.2.1 Υπόβαθρο . . . . .	29
2.2.2 Cascaded Winner-take-all . . . . .	30
2.3 Περιορισμοί των κυκλωμάτων . . . . .	32
<b>3 Αρχιτεκτονική Συστήματος</b>	<b>33</b>
3.1 Υπόβαθρο . . . . .	33
3.1.1 Απλό Μπεϋζιανό Μοντέλο . . . . .	33

3.1.2	Γκαουσιανό Μοντέλο Μίξης . . . . .	34
3.2	Αναλογικός Μπεϋζιανός Ταξινομητής . . . . .	35
3.3	Αναλογικός Ταξινομητής Βασισμένος στο GMM . . . . .	35
3.4	Εξαγωγή Παραμέτρων . . . . .	37
3.4.1	Εκπαίδευση . . . . .	38
3.4.2	Προσαρμοστικότητα . . . . .	38
<b>4</b>	<b>Εφαρμογές και Προσομοιώσεις</b>	<b>41</b>
4.1	Διάγνωση Ασθενειών Θυρεοειδή . . . . .	41
4.1.1	Θυρεοειδής . . . . .	41
4.1.2	Επίδοση του Αναλογικού Ταξινομητή . . . . .	42
4.2	Πρόβλεψη Επιληπτικών Κρίσεων . . . . .	43
4.2.1	Επιληπτικές Κρίσεις . . . . .	44
4.2.2	Επίδοση του Αναλογικού Ταξινομητή . . . . .	45
4.3	Άλλες Εφαρμογές . . . . .	47
4.3.1	Δοκιμαστικό Δισδιάστατο Σύνολο Δεδομένων . . . . .	48
4.3.2	Πραγματικό Σύνολο Δεδομένων 1 . . . . .	50
4.3.3	Πραγματικό Σύνολο Δεδομένων 2 . . . . .	52
4.4	Σύγκριση . . . . .	53
<b>5</b>	<b>Συμπεράσματα και Μελλοντική Δουλειά</b>	<b>57</b>

# Ευρετήριο Εικόνων

2.1	Delbruck's Simple Bump. Η τάση $V_m$ και το ρεύμα πόλωσης $I_{bias}$ ελέγχουν την μέση τιμή και το ύψος της Γκαουσιανής εξόδου $I_{out}$ . . . . .	22
2.2	Cascode Bump. Οι τάσεις $V_r$ και $V_c$ και το ρεύμα πόλωσης $I_{bias}$ ελέγχουν την μέση τιμή, την διασπορά και το ύψος της Γκαουσιανής εξόδου $I_{out}$ . . . . .	24
2.3	Η έξοδος του Cascode Bump, για $I_{bias} = 16nA$ και $V_r = V_c = 0V$ . . . . .	25
2.4	(4 πάνω: Παραμετρικές προσομοιώσεις, ως προς το $V_r$ (μεταβολή μέσης τιμής), του ρεύματος εξόδου του Cascode Bump, για $V_c = 300\mu$ και (αριστερά) $I_{bias} = 16nA$ (δεξιά) $I_{bias} = 1nA$ . 4 κάτω: Παραμετρικές προσομοιώσεις, ως προς το $V_c$ (μεταβολή διασποράς), του ρεύματος εξόδου του Cascode Bump, για $V_r = 0V$ και (αριστερά) $I_{bias} = 16nA$ (δεξιά) $I_{bias} = 1nA$ . . . . .	26
2.5	Neuron Bump. Το ρεύμα $I_r$ , η τάση $V_c$ και το ρεύμα πόλωσης $I_{bias}$ ελέγχουν την μέση τιμή, την διασπορά και το ύψος της Γκαουσιανής εξόδου $I_{out}$ . . . . .	27
2.6	Παραμετρικές προσομοιώσεις του ρεύματος εξόδου του Neuron Bump. $M$ είναι ένας συντελεστής που πολλαπλασιάζει την αναλογία $W/L$ των τρανζίστορ $M_{n2}-M_{n5}$ . (πάνω αριστερά) ως προς το $I_{bias}$ (κλιμάτωση ύψους) για $I_r = 5nA$ , $V_c = 0V$ , $M = 1$ (πάνω δεξιά) ως προς το $I_r$ (μεταβολή μέση τιμής) για $I_{bias} = 12nA$ , $V_c = 0V$ , $M = 1$ (κάτω αριστερά) ως προς το $V_c$ (μεταβολή διασποράς) για $I_{bias} = 12nA$ , $V_r = 0V$ , $M = 1$ (κάτω δεξιά) ως προς το $M$ (μεταβολή διασποράς) για $I_{bias} = 12nA$ , $V_r = 5nA$ , $V_c = 0V$ . . . . .	29
2.7	Δύο παραλλαγές Bump κυκλώματος. Οι τάσεις $V_r$ και $V_c$ και το ρεύμα πόλωσης $I_{bias}$ ελέγχουν την μέση τιμή, την διασπορά και το ύψος της Γκαουσιανής εξόδου $I_{out}$ . . . . .	30
2.8	Ένα κλασικό WTA κύκλωμα με δύο PMOS νευρώνες. . . . .	31

2.9	Η γενική μορφή του <i>Cascaded WTA</i> που αποτελείται από 2 NMOS-WTA και 1 PMOS-WTA. Η παραλλαγή με 1 NMOS-WTA και 2 PMOS-WTA φτιάχνεται αντίστοιχα. . . . .	31
2.10	Τα όρια απόφασης για το Lazzaro WTA και για το <i>Cascaded WTA</i> . Η έξοδος του <i>Cascaded WTA</i> είναι καλύτερης ποιότητας εφόσον έχει πολύ μειωμένη γραμμική περιογή, το οποίο είναι ιδανικό για προβλήματα ταξινόμησης. . . . .	32
3.1	Αναλογικός Bayesian ταξινομητής με 3 κλάσεις και 5-D εισόδους. (αριστερά) πολυδιάστατα κυκλώματα Bump (δεξιά) κύκλωμα WTA. . . . .	36
3.2	Αναλογικός ταξινομητής που βασίζεται στο GMM με $N_{cla}$ κλάσεις, $N_{clu}$ υπο-κατηγορίες ανά κλάση και $N_d$ -D εισόδους. . . . .	40
4.1	Layout του Bayesian ταξινομητή. . . . .	43
4.2	Αποτελέσματα προσομοίωσης Μοντε άρλο μετά τη διάταξη. . . . .	44
4.3	Layout του GMM ταξινομητή, για την πρόβλεψη των επιληπτικών κρίσεων. . . . .	45
4.4	Σύγκριση αρχιτεκτονικής. (α) Πλήρως ψηφιακό σύστημα. (β) Εξαγωγή αναλογικών χαρακτηριστικών. (γ) Αναλογικό wake-up κύκλωμα, όπου το ψηφιακό back-end ενεργοποιείται και απενεργοποιείται με βάση την έξοδο του αναλογικού ταξινομητή χαμηλής ισχύος. . . . .	46
4.5	Οι συναγερμοί που ενεργοποιούνται από τον ταξινομητή για τέσσερις ασθενείς σε ένα 24ωρο. Η ιδανική συμπεριφορά είναι η άνοδος τουλάχιστον ενός συναγερμού σε κάθε pre-ictal περίοδο, χωρίς να ακούγονται συναγερμοί κατά τη διάρκεια των inter-ictal. Οι ictal και post-ictal περιοχές είναι αδιάφορες για τον ταξινομητή. . . . .	47
4.6	Αποτελέσματα προσομοίωσης ανάλυσης ευαισθησίας Μοντε-άρλο μετά τη διάταξη. . . . .	48
4.7	Διάταξη μιας προτεινόμενης αρχιτεκτονικής GMM (Proposed I). . . . .	49
4.8	Δισδιάστατα όρια απόφασης (αριστερά) Ground Truth (μέση) τα αποτελέσματα του ταξινομητή Proposed I (δεξιά) τα αποτελέσματα του ταξινομητή Proposed II . . . . .	50
4.9	Αποτελέσματα προσομοίωσης Monte-Carlo (post-layout) της αρχιτεκτονικής Proposed I στο σύνολο δεδομένων CWRU. . . . .	52
4.10	Αποτελέσματα προσομοίωσης Monte-Carlo (post-layout) της αρχιτεκτονικής Proposed I στο σύνολο δεδομένων VSBD. . . . .	53

# Κατάλογος Πινάκων

2.1	Διαστάσεις τρανζίστορ (Σχήμα 2.2). . . . .	24
2.2	Διαστάσεις τρανζίστορ (Σχήμα 2.5). . . . .	28
2.3	Διαστάσεις τρανζίστορ (Σχήμα 2.7). . . . .	28
2.4	ΜΟΣ Τρανζιστορσ' Διμενσιονς (Φιγ. 2.9). . . . .	31
4.1	Αποτελέσματα ακρίβειας (για 20 επαναλήψεις). . . . .	43
4.2	Αποτελέσματα μείωσης κατανάλωσης ισχύος (για 20 επαναλήψεις). . . . .	47
4.3	Ιδιότητες συνόλου δεδομένων . . . . .	48
4.4	Ακρίβεια ορίων απόφασης. . . . .	49
4.5	Εξαγόμενα features [1] . . . . .	51
4.6	Αποτελέσματα ακρίβειας στο σύνολο δεδομένων CWRU (για 20 επαναλήψεις). . . . .	51
4.7	Αποτελέσματα ακρίβειας στο σύνολο δεδομένων VSBDB (για 20 επαναλήψεις). . . . .	53
4.8	Σύνοψη αλγορίθμων μηχανική μάθησης. . . . .	55





# Κεφάλαιο 1

## Εισαγωγή

Η συνεχιζόμενη πρόοδος στις τεχνολογίες ολοκληρωμένων κυκλωμάτων (Integrated Circuits (IC)) έχει οδηγήσει σε πολύπλοκα και ενεργειακά αποδοτικά συστήματα που αντιμετωπίζουν τις προκλήσεις διαφόρων εφαρμογών Internet of Things (IoT) και Μηχανικής Μάθησης (Machine learning (ML)) [2].

### 1.1 Analog Computing

Το IoT έχει σαν στόχο να επιτρέψει την αυτόνομη ανταλλαγή χρήσιμων πληροφοριών μεταξύ διαφορετικών συσκευών του πραγματικού κόσμου [3]. Οι IoT συσκευές αυξάνονται εκθετικά, με πολυάριθμες εφαρμογές στην υγειονομική περίθαλψη, τον αυτοματισμό, τις φορητές συσκευές και πολλές άλλες [4, 5]. Παράλληλα όμως, αυξάνονται και οι απαιτήσεις αυτών των συστημάτων τόσο σε υπολογιστική ισχύ όσο και σε μειωμένη επιφάνεια [6]. Για αυτό τον λόγο, ειδικά υπολογιστικά συστήματα, επιταχυντές, όπως GPUs, Field-Programmable Gate Arrays (FPGAs) ή Digital ASICs είδη αντικαθιστούν τα παραδοσιακά υπολογιστικά συστήματα [7, 8, 9, 10, 11].

Ένας μεγάλος τομέας του IoT είναι τα smart sensor συστήματα για φορητές συσκευές, οι οποίες εξάγουν πληροφορία από το περιβάλλον συνεχώς. Ωστόσο, η ανάγκη απρόσκοπτης λειτουργίας με μπαταρία λιθίου ή χρήσης συλλέκτη ενέργειας θέτει περιορισμούς: η εξ ολοκλήρου ψηφιακή επεξεργασία σήματος και η ψηφιακή εξαγωγή συμπερασμάτων απαιτούν μεγάλα ποσά ενέργειας. Έτσι, οι παραδοσιακοί ψηφιακοί επεξεργαστές αγωνίζονται να παράσχουν την απαιτούμενη απόδοση δεδομένων των περιορισμών ισχύος και περιοχής [12]. Για να αυξηθεί η αυτονομία των φορητών συσκευών η συνολική κατανάλωση ενέργειας πρέπει να μειωθεί. Ένας πολλά υποσχόμενος υποψήφιος για την άμβλυνση αυτού του περιορισμού είναι η χρήση αναλογικών υπολογιστικών μονάδων. Τα

αναλογικά κυκλώματα, γενικά, προσφέρουν υψηλές υπολογιστικές ικανότητες με πολύ μειωμένες ενεργειακές απαιτήσεις. Μερικά παραδείγματα πλήρως αναλογικών κυκλωμάτων [13, 14].

Η κατανάλωση των αναλογικών κυκλωμάτων μπορεί να μειωθεί περαιτέρω με την λειτουργία τους στην sub-threshold περιοχή [15]. Σε αυτή την περίπτωση, η μείωση της τάσης τροφοδοσίας σε πολύ χαμηλά επίπεδα (και λιγότερο από 1V) οδηγεί στην δυνατότητα πόλωσης των τρανζίστορ με ρεύματα της τάξεως των nano-Ampere. Αυτός ο συνδυασμός μειώνει σημαντικά την κατανάλωση των κυκλωμάτων που λειτουργούν στην sub-threshold περιοχή. Τα ρεύματα των τρανζίστορ σε αυτήν την περιοχή είναι εκθετικά ως προς τις τάσεις που εφαρμόζονται στο τρανζίστορ και δίνονται από τις σχέσεις:

$$I_{pmos} = I_{o_p} e^{\kappa_p(V_w - V_G)/V_T} \left( e^{(V_S - V_w)/V_T} - e^{(V_D - V_w)/V_T} \right), \quad (1.1)$$

$$I_{nmos} = I_{o_n} e^{\kappa_n(V_G - V_w)/V_T} \left( e^{(V_w - V_S)/V_T} - e^{(V_w - V_D)/V_T} \right). \quad (1.2)$$

Εδώ,  $\kappa_p$  και  $\kappa_n$  είναι οι συντελεστές κλίσης για τα τρανζίστορ PMOS και NMOS, αντίστοιχα, τα  $V_G$ ,  $V_S$ ,  $V_D$  και  $V_w$  είναι η τάση πύλης, η τάση πηγής, η τάση υποδοχής και η τάση σώματος, αντίστοιχα,  $V_T$  είναι η θερμοκή τάση και  $I_{o_p}$  και  $I_{o_n}$  είναι τα χαρακτηριστικά ρεύματα (προεκθετικό ρεύμα) για τρανζίστορ PMOS και NMOS, αντίστοιχα [16]. Αυτή η σχέση δίνει την δυνατότητα κατασκευής διαφορετικών διατάξεων, σε σχέση με τα κυκλώματα που λειτουργούν σε above threshold περιοχές, όπως κυκλώματα βασισμένα στο Translinear Principle [17, 18] και άλλα [19, 15].

## 1.2 Μηχανική Μάθηση με Αναλογικά Κυκλώματα

Ο σχεδιασμός μηχανών που θα μπορούσαν να αναπτύξουν τη δική τους νοημοσύνη ήταν ένα από τα μεγαλύτερα όνειρα της ανθρωπότητας και μια από τις μεγαλύτερες επιστημονικές και τεχνολογικές προκλήσεις του περασμένου αιώνα. Ως ML ορίζεται η μελέτη αλγορίθμων και στατιστικών μοντέλων που χρησιμοποιούνται για την επιτυχή εκτέλεση εργασιών χωρίς να έχουν προγραμματιστεί ρητά να το κάνουν [20, 21]. Η προσέγγιση της μηχανικής μάθησης επιτρέπει στα προγράμματα υπολογιστών να δημιουργούν νέα γνώση χωρίς ένα συγκεκριμένο σύνολο εντολών, αλλά χρησιμοποιώντας ένα σύνολο δειγμάτων δεδομένων και εξάγοντας χρήσιμα μοτίβα από αυτό. Αυτή η γενίκευση γνώσης πραγματοποιείται χωρίς ανθρώπινη παρέμβαση και έχει ως αποτέλεσμα την αποτελεσματική πρόβλεψη ή ταξινόμηση νέων πληροφοριών από το σύστημα μηχανικής μάθησης. Το ML χρησιμοποιείται ως θεμελιώδες εργαλείο σε πληθώρα εφαρμογών

στον σύγχρονο κόσμο, που κυμαίνονται από βιοϊατρικές εφαρμογές [22] έως αναγνώριση ομιλίας [23], αυτόνομη οδήγηση [24] και χρηματιστηριακές συναλλαγές [25]. Πολλές από αυτές τις εφαρμογές είναι για IoT συστήματα.

Οι αλγόριθμοι μηχανικής μάθησης παραδοσιακά υλοποιούνται εξ ολοκλήρου σε λογισμικό. Ωστόσο, ο όγκος των δεδομένων που απαιτούνται από τις IoT εφαρμογές που χρησιμοποιούν ML συστήματα αυξάνεται σταθερά τα τελευταία χρόνια. Καθώς οι εργασίες ML περιλαμβάνουν τεράστιο όγκο υπολογισμών, οι αλγόριθμοι ML γίνονται όλο και πιο απαιτητικοί στη μνήμη και στους υπολογιστικούς πόρους [26, 27]. Συνεπώς, εδώ η ανάγκη για επιταχυντές είναι ακόμα μεγαλύτερη και τα οφέλη των αναλογικών υλοποιήσεων είναι σχεδόν απαραίτητα. Πολλές από τις εφαρμογές ML που περιγράφονται παραπάνω απαιτούν υπολογισμό σε πραγματικό χρόνο. Η χαμηλού κόστους επεξεργασία πληροφοριών στο τσιπ χωρίς την ανάγκη μεταφοράς δεδομένων ενισχύει σημαντικά την ανάπτυξη έξυπνων και αυτόνομων IoT συστημάτων χαμηλής καθυστέρησης και μεγάλης διάρκειας μπαταρίας [28, 29].

### 1.3 Gaussian Mixture Models

Τα μοντέλα μίξης (Mixture Models (MM)) είναι ένας απλός και εύκολος τρόπος για την αποτελεσματική περιγραφή πολύπλοκων δεδομένων και χρησιμοποιούνται στο ML για περισσότερο από μια δεκαετία [20, 30]. Το τεράστιο εύρος των εφαρμογών τους περιλαμβάνει πιθανολογικές προσεγγίσεις μοντέλων (οικονομικά μοντέλα [31, 32], ιατρικές εφαρμογές, αναγνώριση χειρογράφου [33, 34], διαχείριση σφαλμάτων [1]), αναγνώριση θορύβου [35, 36], διαχωρισμός εικόνων [37] και λοιπά. Η ακρίβεια που επιτυγχάνεται σε τέτοιες εφαρμογές μπορεί να ταιριάζει με αυτή ακόμη πιο περίπλοκων και εξελιγμένων συστημάτων, όπως των Support Vector Machines (SVM) και των νευρωνικών δικτύων [1]. Το πιο διαδεδομένο MM είναι το Gaussian Mixture Model (GMM), καθώς βασίζεται στην κανονική κατανομή [20, 30]. Κατά συνέπεια, τα GMM αποτελούν ένα άκρως ερευνημένο θέμα στη βιβλιογραφία, για διάφορα επιστημονικά πεδία.

Όταν τα GMM χρησιμοποιούνται χωρίς επίβλεψη, κάθε στοιχείο καταγράφει μία συγκεκριμένη υπο-κατηγορία του εξεταζόμενου συνόλου δεδομένων. Αυτό τα καθιστά κατάλληλα για προβλήματα ομαδοποίησης (clustering). Σε προβλήματα ταξινόμησης, όπως αυτά που θα αναλυθούν παρακάτω, χρησιμοποιούνται πολλαπλά GMM. Σε αυτήν την περίπτωση, για κάθε κλάση, χρησιμοποιείται ένα μόνο GMM για clusing των δεδομένων ανεξάρτητα από τις άλλες κλάσεις. Ο αριθμός των συστάδων επιλέγεται με βάση την πολυπλοκότητα της κατανομής του συνόλου δεδομένων. Γενικά, ένας μεγάλος αριθμός συστάδων αυξάνει την ακρίβεια της προσέγγισης αλλά και την πολυπλοκότητα

του συστήματος.

## Κεφάλαιο 2

# Βασικά κυκλώματα

Σε αυτό το κεφάλαιο θα αναλυθούν τα βασικά κυκλώματα που αναπτύχθηκαν και χρησιμοποιήθηκαν στο πλαίσιο αυτής της διπλωματικής. Όλες οι αρχιτεκτονικές έχουν σχεδιαστεί να λειτουργούν στην Sub-threshold περιοχή, με τις ράγες τροφοδοσίας να είναι  $V_{DD} = -V_{SS} = 0.3V$ , με σκοπό την μείωση της κατανάλωσης τους και του συστήματος στο οποίο θα χρησιμοποιηθούν [15]. Ο σχεδιασμός και η προσομοίωση όλων των κυκλωμάτων έγινε σε μια TSMC 90nm CMOS τεχνολογία χρησιμοποιώντας το Cadence IC design suite.

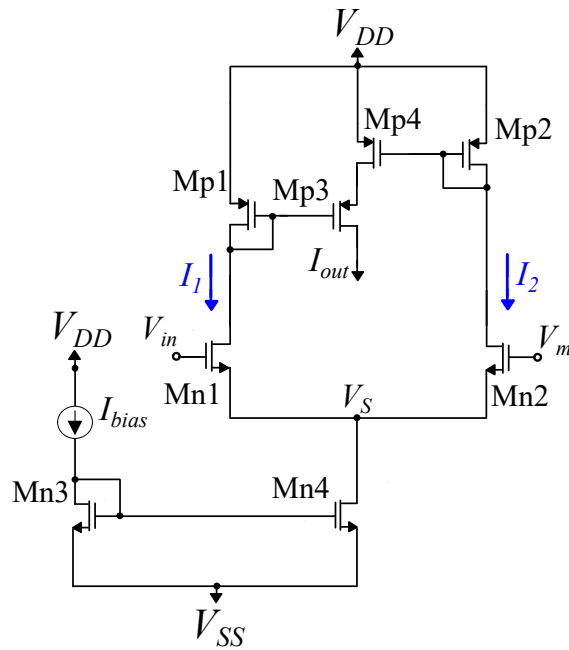
### 2.1 Bump κυκλώματα

Σε αυτήν την ενότητα εξηγούνται οι κύριες λειτουργίες των Bump κυκλωμάτων καθώς και οι παραλλαγές που αναπτύχθηκαν σε αυτήν την διπλωματική.

#### 2.1.1 Υπόβαθρο

Τα Bump κυκλώματα είναι απλές διατάξεις, ικανές να προσομοιάσουν την συμπεριφορά της Γκαουσιανής συνάρτησης με μικρό υλικό και ενεργειακό κόστος [38]. Το πρώτο κύκλωμα Bump (Delbruck's Simple Bump) εισήχθη από τον Delbruck [39, 19] το 1991 και φαίνεται στο Σχήμα 2.1. Η ανάπτυξη του υποκυκλώματος (τρανζίστορ  $M_{p1} - M_{p4}$ ) ικανού να υπολογίσει την συσχέτιση 2 ρευμάτων είναι ο ακρογωνιαίος λίθος, αυτουνού των περισσότερων μετέπειτα Bump κυκλωμάτων [38]. Το υπο-κύκλωμα αυτό ονομάζεται current correlator και συσχετίζοντας 2 στιγμοειδή ρεύματα η έξοδος του είναι μια καμπύλη μορφής καμπάνας που προσομοιάζει την Γκαουσιανή. Σημαντικό ρόλο επίσης έπαιξε ότι αυτό το υπο-κύκλωμα λειτουργεί στην sub-threshold περιοχή. Βασικά σημεία μελέτης και βελτίωσης τους έχουν αποτελέσει η ευρεία προσαρμογή

των βασικών χαρακτηριστικών μιας Γκαουσιανής συνάρτησης (μέση τιμή, διασπορά και ύψος), η μείωση του χώρου και της κατανάλωσης του κυκλώματος, η ομοιότητα της εξόδου με την ιδανική απόκριση και η προσαρμογή τους για τον υπολογισμό πολυδιάστατων συναρτήσεων [38]. Σε αυτή την διπλωματική ο κύριος στόχος σχετικά με την μελέτη πάνω στα Bump κυκλώματα είναι η μείωση της κατανάλωσης χωρίς την ύφεση των υπολοίπων χαρακτηριστικών.



Σχήμα 2.1: Delbruck's Simple Bump. Η τάση  $V_m$  και το ρεύμα πόλωσης  $I_{bias}$  ελέγχουν την μέση τιμή και το ύψος της Γκαουσιανής εξόδου  $I_{out}$ .

Στην βιβλιογραφία, έχουν προταθεί διαφόρων ειδών τεχνικές, σχετικά με την υλοποίηση ενός Bump κυκλώματος [38]: βασιζόμενα στο Translinear Principle [17], στην χρήση Bulk-controlled τρανζίστορ [40, 41], στις ιδιότητες των Floating Gate MOSFET [42], στην χρήση μόνο απλών διαφορικών ζευγών ή στην εκμετάλλευση μεικτών αρχιτεκτονικών [43] (χρήση αναλογικών και ψηφιακών κυκλωμάτων) [38]. Κάθε μια από αυτές τις τεχνικές προσφέρει διάφορα πλεονεκτήματα τα οποία χρειάζονται στις εφαρμογές για τις οποίες προορίζονταν. Έχοντας σαν στόχο την χαμηλή κατανάλωση, τα κυκλώματα που θα παρουσιαστούν παρακάτω, ακολουθούν τις Bulk-controlled υλοποιήσεις, και εκμεταλλεύονται το σώμα (Bulk) των τρανζίστορ για να παράγουν ρυθμιζόμενες Γκαουσιανές συναρτήσεις [40].

Ένα κύκλωμα Bump παράγει μία μονοδιάστατη Γκαουσιανή συνάρτηση. Σε εφαρμογές του πραγματικού κόσμου η χρήση πολυδιάστατων συναρτήσεων είναι απαραίτητη [20]. Ένα πολύ βασικό χαρακτηριστικών των Bump κυκλωμάτων είναι η δυνατότητα τους να υλοποιήσουν πολυδιάστατες Γκαουσιανές συναρτήσεις με πολύ εύκολο τρόπο, αποφεύγοντας την ανάγκη για επιπλέον κυκλώματα. Συγκεκριμένα, η σύνδεση δύο ή περισσότερων Bump κασχοδικά είναι ισοδύναμη με τον πολλαπλασιασμό των ρευμάτων τους [44]. Ακολουθώντας την εξίσωση (2.1), η έξοδος του τελευταίου Bump της σειράς ισοδυναμεί με την τιμή μιας πολυδιάστατης Γκαουσιανής συνάρτησης με διαγώνιο πίνακα συνδιασποράς. Σε αυτήν την τοπολογία, μόνο το πρώτο Bump πολώνεται με ένα προκαθορισμένο ρεύμα.

$$\mathcal{N}(\mathbf{X}|\mathbf{M}_i, \mathbf{\Sigma}_i) = \prod_{n=1}^N \mathcal{N}(x_n|\mu_n, \sigma_n^2), \quad (2.1)$$

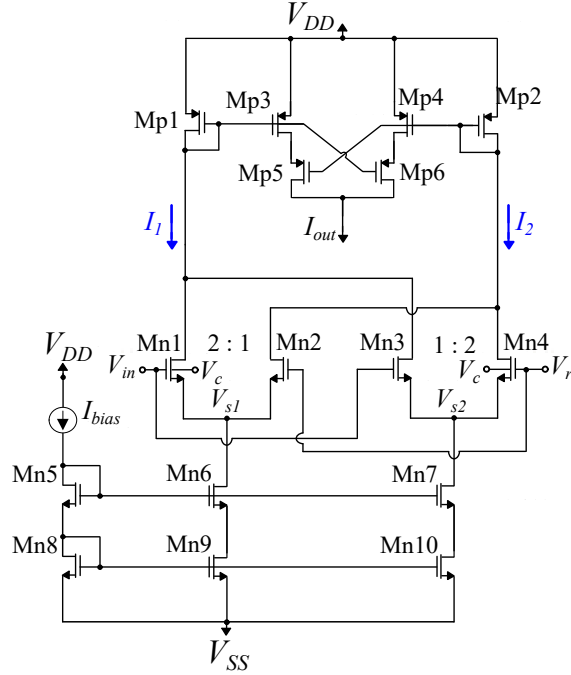
όπου  $x_n$ ,  $\mu_n$  και  $\sigma_n^2$  είναι οι  $n$ -οστές τιμές των διανυσμάτων  $\mathbf{X}$ ,  $\mathbf{M}_i$  και η  $(n,n)$  τιμή του πίνακα  $\mathbf{\Sigma}_i$ , αντίστοιχα.

### 2.1.2 Cascade Bump

Το πρώτο Bump κύκλωμα που παρουσιάζεται είναι το *Cascade Bump* που φαίνεται στο Σχήμα 2.2. Το κύκλωμα αυτό αναπτύχθηκε με σκοπό η έξοδος του να είναι μια υψηλής ποιότητας Γκαουσιανή, με μεγάλο εύρος μεταβολής των χαρακτηριστικών της και να είναι ικανό να παράξει μεγάλου βαθμού πολυδιάστατες Γκαουσιανές με μικρή παραμόρφωση στην έξοδο.

Το *Cascade Bump* αποτελεί μια τροποποίηση του [45], και συνεπώς και τα δύο είναι βασισμένα στο Delbruck's Simple Bump. Συγκεκριμένα, αντί για τον απλό current correlator που χρησιμοποιείται στο [45], προτιμάται ένας συμμετρικός (τρανζίστορ  $M_{p1}$ - $M_{p6}$  του Σχήματος 2.2. Το κίνητρο για αυτήν την τροποποίηση ήταν η ανάγκη για συμμετρικές καμπύλες Γκάους, κατά τη σύγκριση δύο τιμών ΣΙΠΙ. Στην πράξη, χρησιμοποιώντας έναν συμμετρικό current correlator, η συμμετρία για τις εισόδους γύρω από τη μέση τιμή διατηρείται ακόμη και για μικρά ρεύματα. Επιπλέον, ένας κασχοδικός καθρέφτης ρεύματος που αποτελείται από τα τρανζίστορ  $M_{n5}$ - $M_{n10}$  (Εικ. 2.2) αντικαθιστά τον απλό του [45]. Με αυτό το υπο-κύκλωμα, το ρεύμα πόλωσης του Bump καθρεφτίζεται χωρίς παραμόρφωση ακόμα και για μικρές τιμές ρεύματος. Οι διαστάσεις των τρανζίστορ συνοψίζονται στον Πίνακα 2.1.

Τα τρία κύρια χαρακτηριστικά μιας μονοδιάστατης Γκαουσιανής κατανομής (μέση τιμή, διασπορά, ύψος) ελέγχονται μέσω των παραμέτρων του κυκλώματος. Συγκεκριμένα, η παράμετρος τάσης  $V_r$  είναι ίση με τη μέση τιμή, και ως εκ



Σχήμα 2.2: *Cascode Bump*. Οι τάσεις  $V_r$  και  $V_c$  και το ρεύμα πόλωσης  $I_{bias}$  ελέγχουν την μέση τιμή, την διασπορά και το ύψος της Γκαουσιανής εξόδου  $I_{out}$ .

Πίνακας 2.1: Διαστάσεις τρανζίστορ (Σχήμα 2.2).

Διαφορικό ζεύγος	W/L ( $\mu m/\mu m$ )	Current correlator	W/L ( $\mu m/\mu m$ )
$M_{n1}, M_{n4}$	1.6/0.4	$M_{p1}, M_{p2}$	1.6/1.6
$M_{n2}, M_{n3}$	0.8/0.4	$M_{p3}-M_{p6}$	0.4/1.6
$M_{n5}-M_{n8}$	0.4/1.6	-	-
$M_{n9}, M_{n10}$	1.6/1.6	-	-

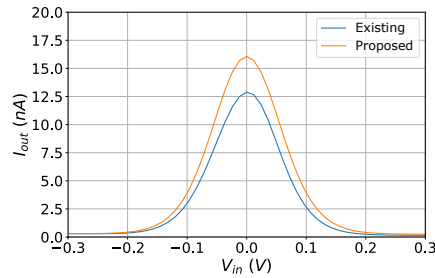
τούτου το μέγιστο της καμπύλης εξόδου επιτυγχάνεται όταν  $V_{in} = V_r$ . Σχετικά με την διασπορά, όμως, η σχέση μεταξύ της και της παραμέτρου τάσης  $V_c$  η οποία την ελέγχει, είναι μια σύνθετη μη γραμμική γνησίως αύξουσα συνάρτηση. Αξίζει να αναφερθεί ότι σε μια τυπική Γκαουσιανή συνάρτηση, το ύψος ορίζεται μονοσήμαντα από τη διακύμανση μέσω του όρου κανονικοποίησης α:

$$\alpha = \frac{1}{\sqrt{(2\pi)^{N_a} |\Sigma|}}. \quad (2.2)$$



Παρόλα αυτά, η αυτόνομη μεταβολή του ύψους είναι μία χρήσιμη ιδιότητα σε πολλές εφαρμογές. Για αυτόν τον λόγο το ύψος της Γκαουσιανής θέτεται ίσο με το ρεύμα πόλωσης  $I_{bias}$  και συνεπώς μπορεί να τοποθετηθεί σε οποιαδήποτε επιθυμητή τιμή. Αυτή η συμπεριφορά είναι παρόμοια για πολυδιάστατα κυκλώματα Bump.

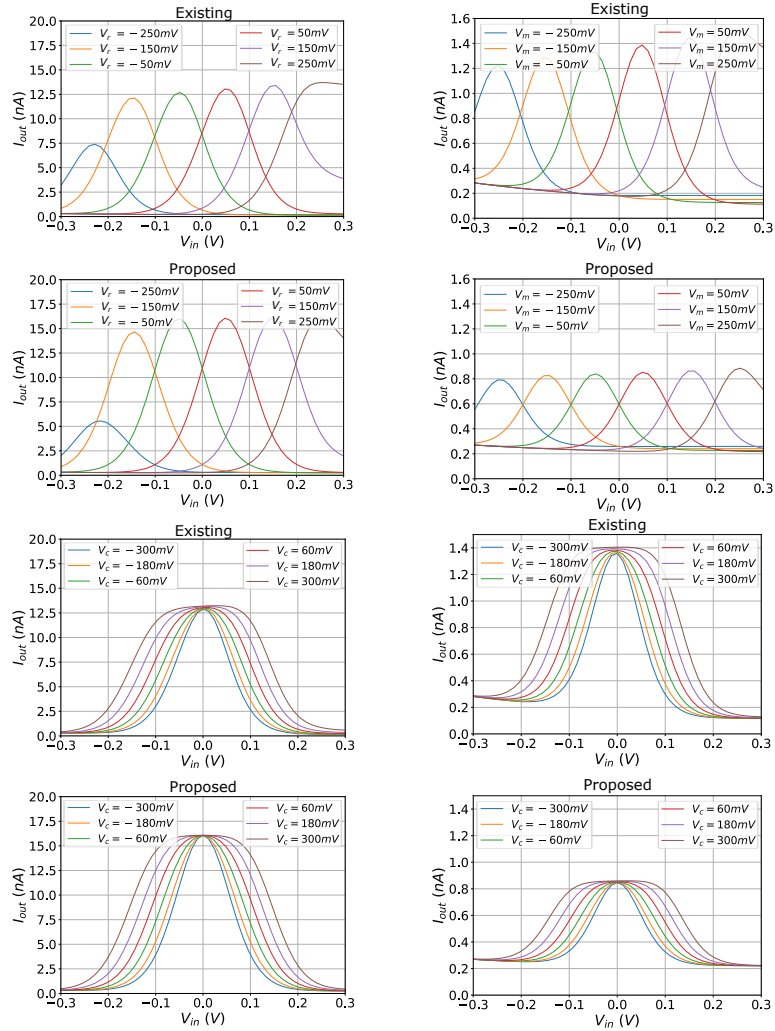
Προκειμένου να εξηγηθεί επαρκώς ο αντίκτυπος των προηγούμενων τροποποιήσεων, συγκρίνονται οι έξοδοι και των δύο κυκλωμάτων, που απεικονίζονται στα Σχήματα 2.3-2.4. Ένα όφελος του *Cascode Bump* μπορεί να παρατηρηθεί στην κορυφή της Γκαουσιανής καμπύλης, η οποία είναι πιο κοντά στις τιμές ρεύματος πόλωσης, σε σύγκριση με την έξοδο του [45], όπως φαίνεται στο σχήμα 2.3 Αυτή η συμπεριφορά είναι επιθυμητή επειδή η κορυφή του ρεύματος εξόδου κάθε κυκλώματος Bump σε ένα πολυδιάστατο Bump κύκλωμα πρέπει να είναι ίση με το ρεύμα πόλωσης τους. Σχετικά με την μεταβολή της μέσης τιμής και της διασποράς, εξετάζονται δύο περιπτώσεις για ρεύματα πόλωσης: ένα με σχετικά μεγάλο ρεύμα πόλωσης  $16nA$  και ένα με μικρότερο ρεύμα πόλωσης  $1nA$ . Δεδομένου ότι σε πολυδιάστατες τοπολογίες, τα πρώτα Bump είναι συνήθως πολωμένα με μεγαλύτερα ρεύματα από αυτά στα μεταγενέστερα στάδια, αυτή η σύγκριση επιτρέπει την αξιολόγηση της συμπεριφοράς τους και στις δύο περιπτώσεις. Το *Cascode Bump* προσφέρει μια πιο συμμετρική (γύρω από τη μέση τιμή) απόκριση σε όλο το εύρος του άξονα  $V_{in}$ , σε σύγκριση με το [45], όπως φαίνεται στο Σχήμα 2.4. Συγκεκριμένα, η καμπύλη του [45] έχει χαμηλότερες τιμές στον θετικό άξονα  $V_{in}$  σε σύγκριση με τον αρνητικό.



Σχήμα 2.3: Η έξοδος του *Cascode Bump*, για  $I_{bias} = 16nA$  και  $V_r = V_c = 0V$ .

### 2.1.3 Neuron Bump

Το δεύτερο Bump κύκλωμα που παρουσιάζεται είναι το *Neuron Bump* που φαίνεται στο Σχήμα 2.5. Το κύκλωμα αυτό αναπτύχθηκε με σκοπό την χρήση του σε συστήματα με πολλές πυκνά τοποθετημένες Γκαουσιανές κατανομές. Συγκεκριμένα, το κύκλωμα αυτό επιτρέπει μεγάλη μείωση στην διασπορά της

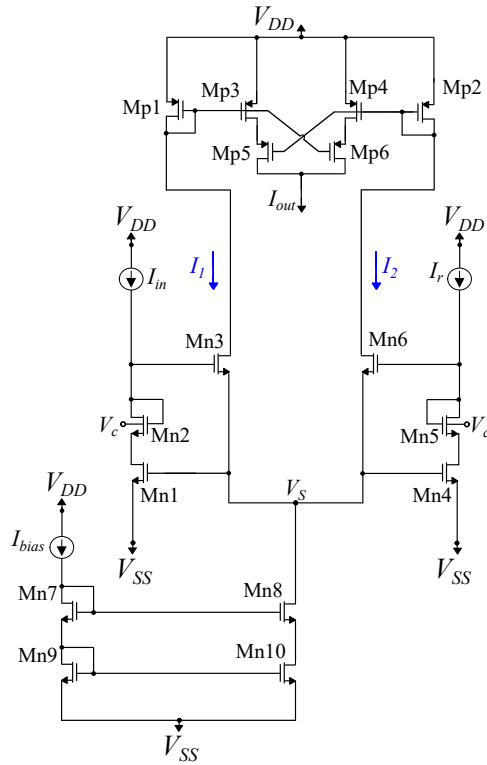


Σχήμα 2.4: (4 πάνω: Παραμετρικές προσομοιώσεις, ως προς το  $V_r$  (μεταβολή μέσης τιμής), του ρεύματος εξόδου του *Cascode Bump*, για  $V_c = 300\mu$  και (αριστερά)  $I_{bias} = 16nA$  (δεξιά)  $I_{bias} = 1nA$ .

4 κάτω: Παραμετρικές προσομοιώσεις, ως προς το  $V_c$  (μεταβολή διασποράς), του ρεύματος εξόδου του *Cascode Bump*, για  $V_r = 0V$  και (αριστερά)  $I_{bias} = 16nA$  (δεξιά)  $I_{bias} = 1nA$ .

Γκαουσιανής καθώς και μεγάλη μεταβολή στο εύρος των τιμών της εισόδου.

Το *Neuron Bump* είναι βασισμένο στο Delbruck's Simple Bump [19], Σχήμα 2.1, και στους νευρώνες του Winner-take-all κυκλώματος του Lazzaro [46]. Συγκεκριμένα, τα τρανζίστορ ενός απλού διαφορικού ζεύγους αντικαθι-



Σχήμα 2.5: *Neuron Bump*. Το ρεύμα  $I_r$ , η τάση  $V_c$  και το ρεύμα πόλωσης  $I_{bias}$  ελέγχουν την μέση τιμή, την διασπορά και το ύψος της Γκαουσιανής εξόδου  $I_{out}$ .

στούνται από τροποποιημένους νευρώνες. Το διαφορικό ζεύγος των νευρώνων παράγει 2 πολύ πιο απότομα σιγμοειδή ρεύματα και συνεπώς η διασπορά της Γκαουσιανής γίνεται πολύ μικρότερη. Στην συγκεκριμένη υλοποίηση τα τρανζιστορ  $Mn2$  και  $Mn5$  έχουν τοποθετηθεί για να ελέγχουν την κλίση των σιγμοειδών. Σημαντική αλλαγή είναι ότι η είσοδος  $I_{in}$  και η παράμετρος που ελέγχουν την μέση τιμή  $I_r$ , αντί για τάσεις, σε αυτό το κύκλωμα είναι ρεύματα. Εξαιτίας αυτού, οι τιμές τους δεν είναι περιορισμένες από την τάση τροφοδοσίας και συνεπώς το εύρος τους είναι ρυθμιζόμενο. Αντίστοιχα με το *Cascode Bump* χρησιμοποιείται ένας συμμετρικός current correlator και ένας κασχοδικός καθρέφτης ρεύματος. Οι διαστάσεις των τρανζιστορ συνοψίζονται στον Πίνακα 2.2.

Η λειτουργία αυτού του κυκλώματος είναι αντίστοιχη με του *Cascode Bump*. Στο Σχήμα 2.6 φαίνεται πως το ρεύμα πόλωσης  $I_{bias}$  και οι παράμετροι  $I_r$  και  $V_c$  επηρεάζουν το ύψος, την μέση τιμή και την διασπορά της Γκαουσιανής καμπύλης, αντίστοιχα. Για την περίπτωση του της διασποράς παρατίθενται δύο διαφορετικές γραφικές: μία ελεγχόμενη από το  $V_c$  και μία από τις διαστάσεις

Πίνακας 2.2: Διαστάσεις τρανζίστορ (Σχήμα 2.5).

Διαφορικό ζεύγος	W/L ( $\mu m/\mu m$ )	Current correlator	W/L ( $\mu m/\mu m$ )
$M_{n1}, M_{n4}$	1, 6/0, 4	$M_{p1}, M_{p2}$	1, 6/1, 6
$M_{n2}, M_{n3}$	0, 8/0, 4	$M_{p3}-M_{p6}$	0, 4/1, 6
$M_{n5}-M_{n8}$	0, 4/1, 6	-	-
$M_{n9}, M_{n10}$	1, 6/1, 6	-	-

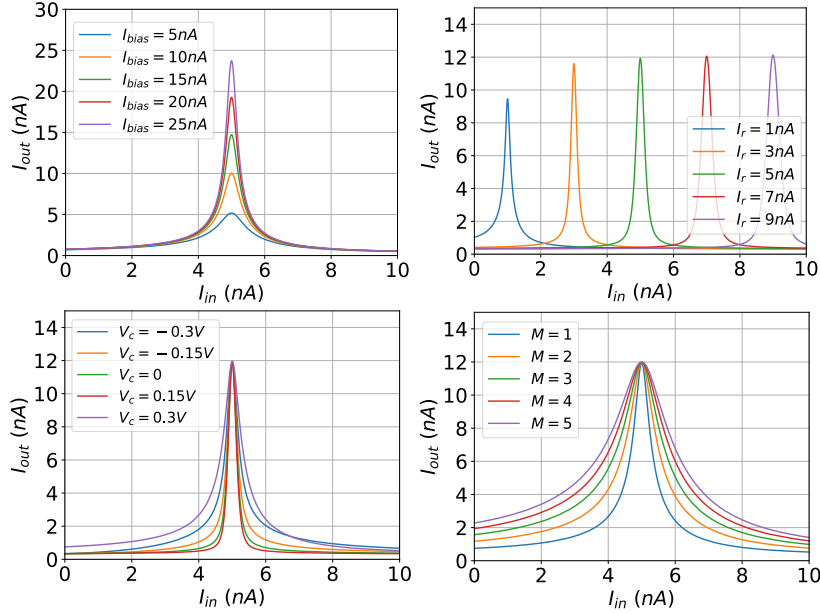
του διοδικά συνδεδεμένου τρανζίστορ στον νευρώνα. Με αυτόν τον τρόπο, το κύκλωμα μπορεί να σχεδιαστεί ανάλογα και σε περιπτώσεις όπου χρειάζεται μεγαλύτερη διασπορά, αλλάζοντας τις διαστάσεις των τρανζίστορ. Η Γκαουσιανή συνάρτηση στην περίπτωση αυτού του Bump, πολύ μικρή διασπορά, αρχίζει να μοιάζει με την κρουστική συνάρτηση, όπως και είναι αναμενόμενο. Επιπλέον, εδώ, η σχέση  $V_c$ -διασποράς δεν είναι μονότονη στο διάστημα  $[-0.3, 0.3]V$ , αλλά είναι αύξουσα στο  $[-0.3, 0.15]V$  και φθίνουσα στο  $[0.15, 0.3]V$

#### 2.1.4 Άλλες Αρχιτεκτονικές

Παράλληλα με τα προηγούμενα διατάξεις, έγινε πειραματισμός και σε άλλα κυκλώματα Bump. Τα κυκλώματα που φαίνονται στο Σχήμα 2.7 προέκυψαν κυρίως από εφαρμογή των βελτιώσεων που έγιναν στο *Cascode Bump* σε άλλες τοπολογίες Bump κυκλωμάτων [44, 47]. Με αυτές τις αλλαγές η ποιότητα της Γκαουσιανής βελτιώνεται σε μεγάλο βαθμό και για τα δύο κυκλώματα. Παρόλα αυτά, με εξαίρεση την επιφάνεια του κυκλώματος, το *Cascode Bump* υπερτερεί στις υπόλοιπες μετρικές. Οι διαστάσεις των τρανζίστορ συνοψίζονται στον Πίνακα 2.3.

Πίνακας 2.3: Διαστάσεις τρανζίστορ (Σχήμα 2.7).

Διαφορικό ζεύγος	W/L ( $\mu m/\mu m$ )	Current correlator	W/L ( $\mu m/\mu m$ )
$M_{n1}, M_{n4}$	1.6/0.4	$M_{p1}, M_{p2}$	1.6/1.6
$M_{n2}, M_{n3}$	0.8/0.4	$M_{p3}-M_{p6}$	0.4/1.6
$M_{n5}-M_{n8}$	0.4/1.6	-	-
$M_{n9}, M_{n10}$	1.6/1.6	-	-



Σχήμα 2.6: Παραμετρικές προσομοιώσεις του ρεύματος εξόδου του *Neuron Bump*.  $M$  είναι ένας συντελεστής που πολλαπλασιάζει την αναλογία  $W/L$  των τρανζίστορ  $M_{n2}-M_{n5}$ .

(πάνω αριστερά) ως προς το  $I_{bias}$  (κλιμάτωση ύψους) για  $I_r = 5nA$ ,  $V_c = 0V$ ,  $M = 1$   
 (πάνω δεξιά) ως προς το  $I_r$  (μεταβολή μέσης τιμής) για  $I_{bias} = 12nA$ ,  $V_c = 0V$ ,  $M = 1$   
 (κάτω αριστερά) ως προς το  $V_c$  (μεταβολή διασποράς) για  $I_{bias} = 12nA$ ,  $V_r = 0V$ ,  $M = 1$

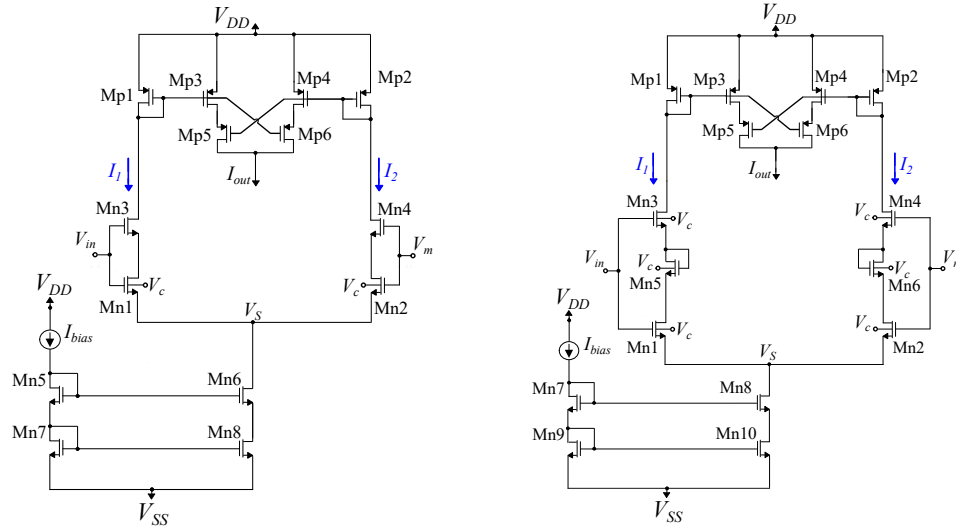
(κάτω δεξιά) ως προς το  $M$  (μεταβολή διασποράς) για  $I_{bias} = 12nA$ ,  $V_r = 5nA$ ,  $V_c = 0V$ .

## 2.2 Winner-take-all κυκλώματα

Σε αυτήν την ενότητα εξηγούνται οι κύριες λειτουργίες του πρώτου WTA κυκλώματος καθώς και μίας παραλλαγής που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής.

### 2.2.1 Υπόβαθρο

Ένα WTA κύκλωμα είναι ένα αναλογικό στοιχείο με  $N$  ζεύγη θυρών εισόδου-εξόδου  $[I_{in_i}, I_{out_i}]_{i=1}^N$ , που υλοποιεί τη συνάρτηση  $argmax$  [46]. Λέμε ότι κάθε ζεύγος  $(I_{in_i}, I_{out_i})$  αποτελεί έναν νευρώνα. Στην πράξη, όπως υποδηλώνει και το όνομα του κυκλώματος, με δεδομένο ένα σύνολο  $N$  σημάτων εισόδου και υποθέτοντας ότι υπάρχει ένα μόνο μέγιστο μεταξύ τους, που βρίσκεται στον

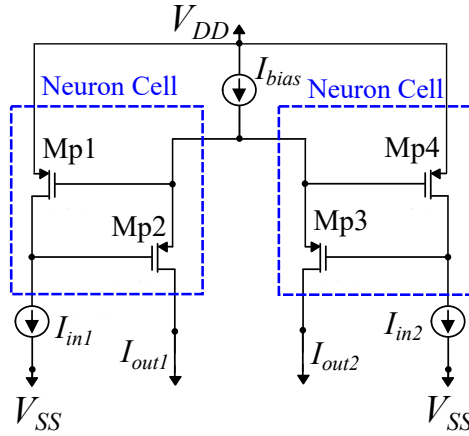


Σχήμα 2.7: Δύο παραλλαγές Bump κυκλώματος. Οι τάσεις  $V_r$  και  $V_c$  και το ρεύμα πόλωσης  $I_{bias}$  ελέγχουν την μέση τιμή, την διασπορά και το ύψος της Γκαουσιανής εξόδου  $I_{out}$ .

δείκτη  $j \leq N$ , η έξοδος  $I_{out_j}$  έχει μη μηδενική τιμή (νικητής), ενώ τα υπόλοιπα είναι μηδέν. Η έξοδος συνεπώς του κυκλώματος είναι ένα δυαδικό διάνυσμα με ιδανικά έναν μόνο άσσο. Εάν όμως δεν υπάρχει ξεκάθαρο μέγιστο, το WTA κύκλωμα λειτουργεί στη γραμμική περιοχή, όπου μπορεί να εμφανιστούν περισσότεροι από ένα νικητή. Το πρώτο δίκτυο WTA προτάθηκε από τον Lazzaro [46]. Ένα τυπικό παράδειγμα με δύο νευρώνες φτιαγμένους με PMOS τρανζίστορ, φαίνεται στο Σχήμα 2.8. Με παρόμοιο τρόπο μπορεί να φτιαχτεί ένα NMOS WTA.

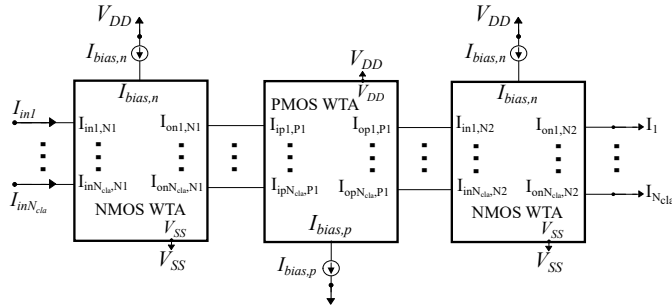
### 2.2.2 Cascaded Winner-take-all

Στα πιθανολογικά μοντέλα ταξινόμησης, οι πιθανότητες δύο ή περισσότερων κλάσεων γύρω από τα όρια απόφασης έχουν κοντινές τιμές. Για το Lazzaro WTA αυτό συνήθως οδηγεί σε πολλαπλούς νικητές. Για την αντιμετώπιση αυτού του ζητήματος, παρουσιάζεται το *Cascaded WTA* που φαίνεται στο Σχήμα 2.9. Το κύκλωμα αυτό αποτελείται από 3 διαδοχικά απλά WTA κυκλώματα, παρόμοια με το [48]. Σημαντικό χαρακτηριστικό αυτής της αρχιτεκτονικής είναι ότι εναλλάσσοντας τις NMOS και PMOS παραλλαγές, εξαλείφεται η ανάγκη για συνδετικά κυκλώματα μεταξύ δύο διαδοχικών WTA κυκλωμάτων. Όλες οι διαστάσεις των τρανζίστορ για τους NMOS και PMOS νευρώνες (Σχήμα 2.9)



Σχήμα 2.8: Ένα κλασικό WTA κύκλωμα με δύο PMOS νευρώνες.

συνοψίζονται στον Πίνακα 2.4.



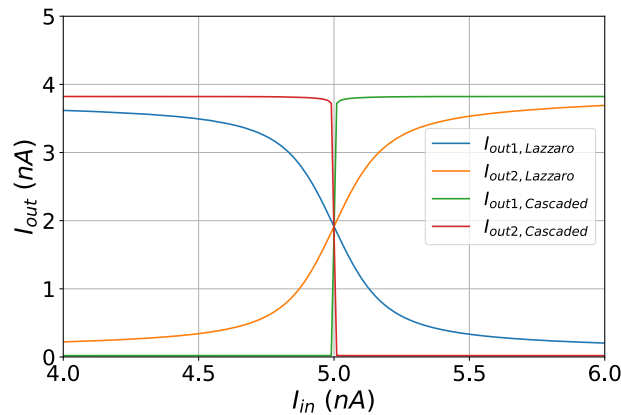
Σχήμα 2.9: Η γενική μορφή του *Cascaded WTA* που αποτελείται από 2 NMOS-WTA και 1 PMOS-WTA. Η παραλλαγή με 1 NMOS-WTA και 2 PMOS-WTA φτιάχνεται αντίστοιχα.

Πίνακας 2.4: ΜΟΣ Τρανζιστορς' Διμενσιονς (Φιγ. 2.9).

NMOS Νευρώνας	W/L ( $\mu m/\mu m$ )	PMOS Νευρώνας	W/L ( $\mu m/\mu m$ )
$M_{n1}-M_{n2}$	0.4/1.6	$M_{p1}-M_{p2}$	0.4/1.6

Το Σχήμα 2.10 φαίνονται τα όρια απόφασης ενός Lazzaro WTA και του *Cascaded WTA* για το ίδιο εικονικό πρόβλημα 1-D. Κάθε ρεύμα εξόδου αντιπροσωπεύει μια πιθανή κατηγορία. Η είσοδος  $I_{in_1}$  είναι ένα σταθερό ρεύμα

ίσο με  $5\text{nA}$ , ενώ η είσοδος  $I_{in2} = I_{in}$  αυξάνεται γραμμικά από τα  $4\text{nA}$  στα  $6\text{nA}$ . Τα κυκλώματα WTA συγκρίνουν και τις δύο εισόδους και παρέχουν τον δείκτη του μέγιστου. Τα θεωρητικά αποτελέσματα αυτού του προβλήματος ταξινόμησης δύο κατηγοριών θα υποδεικνύουν ότι η κλάση A κερδίζει για ρεύμα εισόδου  $I_{in} \in [4, 5)\text{nA}$ , ενώ η κατηγορία B κερδίζει για το ρεύμα εισόδου  $I_{in} \in (5, 6]\text{nA}$ . Το *Cascaded WTA* παρέχει το κατάλληλο όριο απόφασης με μεγαλύτερη ακρίβεια, σε σύγκριση με το *Lazzaro WTA*, μιας και έχει πολύ μικρότερη γραμμική περιοχή.



Σχήμα 2.10: Τα όρια απόφασης για το Lazzaro WTA και για το *Cascaded WTA*. Η έξοδος του *Cascaded WTA* είναι καλύτερης ποιότητας εφόσον έχει πολύ μειωμένη γραμμική περιοχή, το οποίο είναι ιδανικό για προβλήματα ταξινόμησης.

### 2.3 Περιορισμοί των κυκλωμάτων

Τα κυκλώματα που παρουσιάστηκαν έχουν κάποιους περιορισμούς όταν επρόκειτο να χρησιμοποιηθούν σε μεγάλα συστήματα. Ο μεγαλύτερος περιορισμός των *Bump* κυκλωμάτων είναι η επέκτασή τους σε περισσότερες διαστάσεις. Τα καλύτερα αποτελέσματα τα πετυχαίνει το *Cascade Bump*, φτάνοντας μέχρι τις 16 διαστάσεις με πολύ μικρές παραμορφώσεις. Οι περιορισμοί που θέτουν τα WTA κυκλώματα, αφορούν κυρίως τον αριθμό των εισόδων τους. Το *Lazzaro WTA* στις 4 εισόδους αρχίζει να εμφανίζει πολλαπλούς νικητές, ενώ το *Cascaded WTA* συνεχίζει να έχει πολύ καλής ποιότητας εξόδο. Το *Cascaded WTA* μπορεί σίγουρα να αυξήσει αρκετά παραπάνω τον αριθμό των εξόδων.



## Κεφάλαιο 3

# Αρχιτεκτονική Συστήματος

Σε αυτό το κεφάλαιο θα αναλυθούν τα μοντέλα και οι αρχιτεκτονικές των ταξινομητών που αναπτύχθηκαν στο πλαίσιο αυτής της διπλωματικής. Συγκεκριμένα, θα παρουσιαστεί ένας Μπεϋζιανός (Bayesian) ταξινομητής και ένας ταξινομητής Γκαουσιανού μοντέλου μίξης (GMM). Και οι δύο ταξινομητές χρησιμοποιούν τα κυκλώματα που παρουσιάστηκαν στο προηγούμενο κεφάλαιο, αλλά είναι ανεξάρτητοι από τις συγκεκριμένες υλοποιήσεις αυτών.

### 3.1 Υπόβαθρο

#### 3.1.1 Απλό Μπεϋζιανό Μοντέλο

Ένας Bayesian ταξινομητής βασίζεται στο θεώρημα του Bayes, εξίσωση (3.1), με σκοπό να απλοποιηθεί ο υπολογισμός ορισμένων πιθανοτήτων [20].

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}. \quad (3.1)$$

Συγκεκριμένα, ένας τέτοιος πιθανολογικός ταξινομητής συγκρίνει, για όλες τις κλάσεις, την *posterior* πιθανότητα  $p(C|X)$  μία κλάση  $C$  να είναι η κλάση στην οποία ανήκει η είσοδος  $X$  (δεδομένης της εισόδου  $X$ ) μέσω του θεωρήματος Bayes:

$$p(C|X) = \frac{p(C)p(X|C)}{p(X)}, \quad (3.2)$$

Εδώ,  $p(C)$  είναι η *prior* πιθανότητα της κλάσης  $C$ ,  $p(X|C)$  είναι η πιθανότητα του  $X$  δεδομένης μίας κλάσης  $C$  και  $p(X)$  είναι η *evidence* πιθανότητα του  $X$ . Στην πράξη, η  $p(C)$  υπολογίζεται από τα δεδομένα εκπαίδευσης (training

set), η  $p(X|C)$  υπολογίζεται με βάση την κατανομή της κλάσης  $C$  και η  $p(X)$  αγνοείται, εφόσον παραμένει σταθερή στην σύγκριση μεταξύ των κλάσεων.

Στον συγκεκριμένο ταξινομητή θεωρούμε Γκαουσιανή κατανομή σε κάθε κλάση και ανεξαρτησία μεταξύ διαστάσεων της εισόδου (διαγώνιο πίνακα συν-διακύμανσης). Συνεπώς, η πιθανότητα  $p(\mathbf{X}|C)$  υπολογίζεται ως:

$$p(\mathbf{X}|C) = \prod_{n=1}^N \frac{1}{\sqrt{(2\pi) \cdot \sigma_n^2}} e^{-\frac{1}{2} \cdot \frac{(x_n - \mu_n)^2}{\sigma_n^2}}, \quad (3.3)$$

όπου  $N$  είναι ο αριθμός των διαστάσεων τις κατανομής και  $\mu_n$  και  $\sigma_n$  είναι η μέση τιμή και η διακύμανση που αντιστοιχεί στο  $n$  χαρακτηριστικό της κλάσης  $C$ . Αντίστοιχα, το  $x_n$  είναι το  $n$  χαρακτηριστικό του διανύσματος εισόδου  $\mathbf{X}$ . Ο συνολικός ταξινομητής καθορίζει τη νικητήρια κλάση εφαρμόζοντας τον τελεστή  $\operatorname{argmax}$  στις πιθανότητες  $p(C_k|\mathbf{X})$  για όλες τις κλάσεις  $N_C$ , όπως ορίζεται στο:

$$y = \operatorname{argmax}_{k \in [1, N_C]} \{p(C_k|\mathbf{X})\} = \operatorname{argmax}_{k \in [1, N_C]} \{p(C_k)p(\mathbf{X}|C_k)\}. \quad (3.4)$$

### 3.1.2 Γκαουσιανό Μοντέλο Μίξης

Το GMM είναι ένα πιθανολογικό μοντέλο που αντιπροσωπεύει την πυκνότητα μιας  $N$ -διάστατης τυχαίας μεταβλητής ως σταθμισμένο άθροισμα  $K$  Γκαουσιανών υπο-κατανομών, προσφέροντας έτσι μεγαλύτερη εκφραστικότητα από μία μεμονωμένη Γκαουσιανή [20, 49, 50, 51]. Το μοντέλο ορίζεται μοναδικά από τον αριθμό των υπο-κατανομών  $K$ , τους παράγοντες βάρους  $[w_i]_{i=1}^K$ , τις μέσες τιμές  $[M_i]_{i=1}^K$  και τους  $N \times N$  πίνακες συν-διακύμανσης  $[\Sigma_i]_{i=1}^K$  κάθε Γκαουσιανής υπο-κατηγορίας. Στην πράξη ο αριθμός  $K$  είναι υπερπαραμέτρος που επιλέγεται ανάλογα με την εφαρμογή. Ένας σχετικά μεγάλος αριθμός υπο-κατηγοριών μπορεί να μοντελοποιήσει με ακρίβεια πολύπλοκες κατανομές, ενώ λιγότερα συμπλέγματα οδηγούν σε απλούστερες υλοποιήσεις. Τα GMMs χρησιμοποιούνται σε ταξινομητές επειδή είναι εξαιρετικά χρήσιμα σε προβλήματα ομαδοποίησης (clustering).

Ένας ταξινομητής βασισμένος στο GMM, είναι ένας Bayesian ταξινομητής όπου η πιθανότητα μιας κλάσης  $C$  υπολογίζεται όπως ορίζεται από το GMM. Συγκεκριμένα, η πιθανότητα της εισόδου  $X$  δεδομένης μιας κλάσης  $C$ , που αντιστοιχεί και στην κατανομή της  $C$  είναι:

$$p(\mathbf{X}|C) = \sum_{i=1}^K w_i^{(C)} \cdot \mathcal{N}(\mathbf{X} | \mathbf{M}_i^{(C)}, \Sigma_i^{(C)}). \quad (3.5)$$

Εδώ ισχύει ότι  $\sum_{i=1}^K w_i = 1$  και  $0 \leq w_i \leq 1$  για  $i = 1, 2, \dots, K$  και  $\mathcal{N}$  είναι η Γκαουσιανή κατανομή που υπολογίζεται όπως στην 3.3.  $M_i^{(C)}$  και  $\Sigma_i^{(C)}$  είναι οι πίνακες μέσης τιμής και συν-διακύμανσης της  $i$  υπο-κατανομής και της  $C$  κλάσης, αντίστοιχα. Συνεπώς, για κάθε κλάση χρησιμοποιείται ένα ξεχωριστό GMM. Όμοιως με την (3.4), η απόφαση αυτού του ταξινομητή υπολογίζεται ως:

$$y = \operatorname{argmax}_{c \in \{1, N_C\}} \{p(C_c)p(\mathbf{X}|C_c)\}. \quad (3.6)$$

### 3.2 Αναλογικός Μπεϋζιανός Ταξινομητής

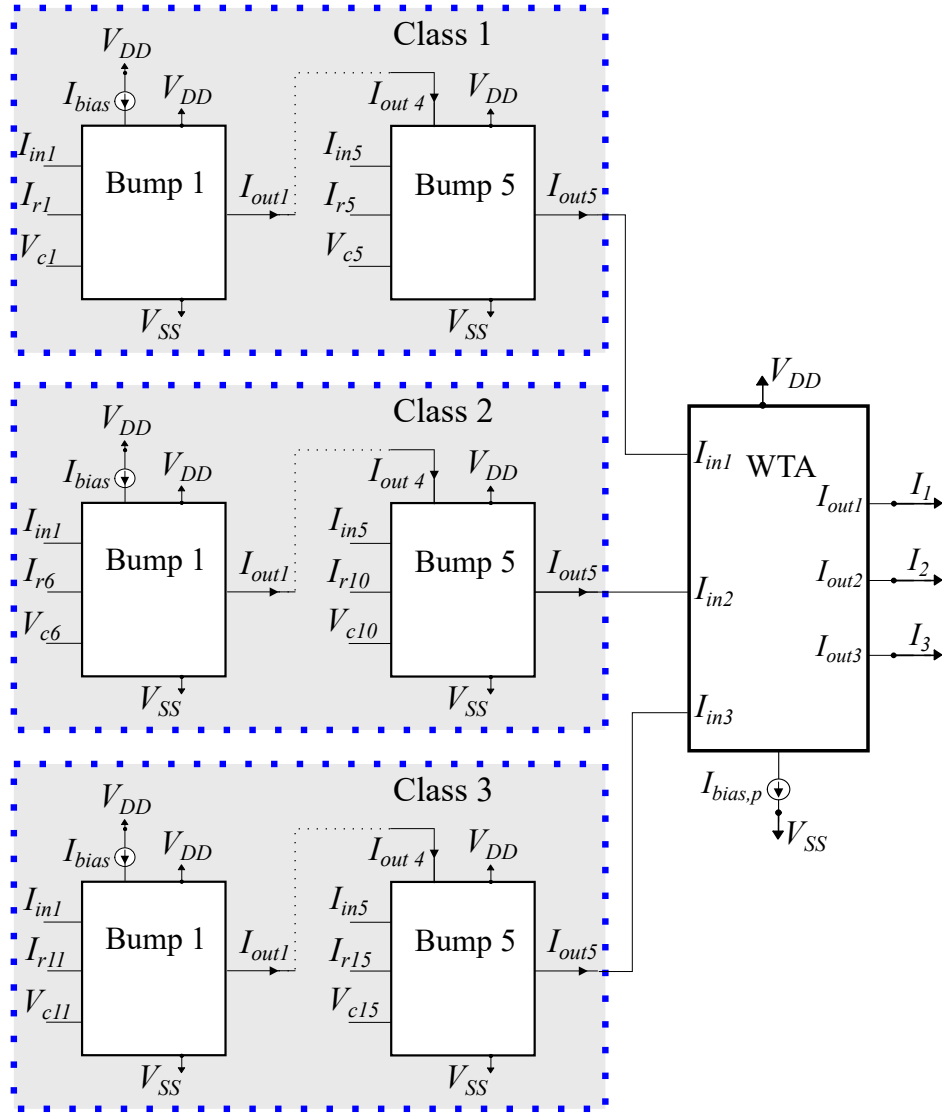
Σε αυτήν την ενότητα εξηγείται η αρχιτεκτονική ενός αναλογικού ταξινομητή Bayes. Με βάση τη εξίσωση (3.4), τα κύρια δομικά στοιχεία ενός Γκαουσιανού ταξινομητή Bayes είναι η Γκαουσιανή συνάρτηση και ο τελεστής  $\operatorname{argmax}$ . Το πρώτο υλοποιείται με τα Bump κυκλώματα και το δεύτερο με τα WTA κυκλώματα.

Ο ταξινομητής του σχήματος 3.1 έχει σχεδιαστεί για ένα πρόβλημα ταξινόμησης με 3 κλάσεις και 5-D εισόδους. Κάθε μπλοκ  $[\text{class}_i]_{i=1}^3$  αποτελείται από μία σειρά από 5 κασχοδικά συνδεδεμένων Bump κυκλωμάτων και αντιπροσωπεύει μία συγκεκριμένη κλάση. Με αυτόν τον τρόπο υλοποιείται η 5-D Γκαουσιανή κατανομή. Η έξοδος του κάθε μπλοκ αντιστοιχεί στην τιμή της κατανομής την οποία το WTA κύκλωμα συγκρίνει. Τα ρεύματα  $[I_i]_{i=1}^3$  υποδεικνύουν την νικήτρια κλάση και είναι και η τελική έξοδος του ταξινομητή.

Η υλοποίηση αυτή μπορεί να γενικευτεί για οποιοδήποτε αριθμό κλάσεων και διαστάσεων. Αυτό γίνεται εύκολα αλλάζοντας το πλήθος τον  $\text{class}_i$  μπλοκ και τον αριθμό των κασχοδικά συνδεδεμένων Bump κυκλωμάτων. Είναι σημαντικό να σημειωθεί ότι η αρχιτεκτονική αυτή δεν έχει, εκ φύσεως της, περιορισμό στον αριθμό των κλάσεων και των διαστάσεων. Οι περιορισμοί προέρχονται από τα βασικά δομικά κυκλώματα που θα χρησιμοποιηθούν. Χρησιμοποιώντας το *Cascade Bump* και το *Cascaded WTA*, ο ταξινομητής μπορεί να δουλέψει μέχρι 16-D εισόδους και για πάνω από 4 κλάσεις, χωρίς σημαντική πτώση στην ακρίβεια.

### 3.3 Αναλογικός Ταξινομητής Βασισμένος στο GMM

Σε αυτήν την ενότητα εξηγείται η αρχιτεκτονική ενός αναλογικού ταξινομητή βασισμένου στο GMM. Για γενικότητα και για να διευκρινιστεί ο συλλογισμός πίσω από αυτήν την αρχιτεκτονική, υποθέτουμε ένα πρόβλημα ταξινόμησης που



Σχήμα 3.1: Αναλογικός Bayesian ταξινομητής με 3 κλάσεις και 5-D εισόδους. (αριστερά) πολυδιάστατα κυκλώματα Bump (δεξιά) κύκλωμα WTA.

περιλαμβάνει  $N_{cla}$  κλάσεις και  $N_d$ -D εισόδους. Ο αριθμός των υπο-κατηγοριών ανά τάξη είναι μια υπερπαράμετρος και έχει επιλεγεί ως  $N_{clu}$ . Η αρχιτεκτονική αυτή μπορεί να εφαρμοστεί για οποιοδήποτε αριθμό από διαστάσεις εισόδου, κλάσεις ή υπο-κατηγορίες.

Η δομή της προτεινόμενης αναλογικής ολοκληρωμένης υλοποίησης ενός ταξινομητή που βασίζεται στο GMM φαίνεται στο Σχήμα 3.2. Με την προαναφερθείσα υπόθεση, ο ταξινομητής απαιτεί  $N_{cla}$  GMM cells, καθένα από τα οποία αποτελείται από  $N_{clu}$  cluster cells και ένα μπλοκ WTA με  $N_{cla}$  εισόδους. Τα cluster cells είναι στην πραγματικότητα πολυδιάστατα Gaussian κυκλώματα με  $N_d$ -D εισόδους. Κάθε κελί εξάγει την πιθανότητα ενός διανύσματος εισόδου  $\mathbf{X}$  να ανήκει σε μία συγκεκριμένη υποκατηγορία, που υπολογίζεται χρησιμοποιώντας την Gaussian ΣΠΠ της υπο-κατηγορίας, σύμφωνα με το (2.1). Αυτή η πιθανότητα πολλαπλασιάζεται επίσης με το αντίστοιχο βάρος ( $w_i$ ) της υπο-κατηγορίας.

Με βάση το (3.5), η πιθανότητα το  $\mathbf{X}$  να ανήκει σε μια συγκεκριμένη κλάση  $C$  υπολογίζεται προσθέτοντας τις πιθανότητες των υπο-κατηγοριών που συνθέτουν την κλάση. Αυτό εκτελείται μέσα σε ένα GMM cell χρησιμοποιώντας τα cluster cells και κασκοδικούς καθρέπτες ρεύματος προκειμένου να αυξηθεί η ποιότητα της άθροισης. Το WTA μπλοκ υλοποιεί τον τελεστή  $\text{argmax}$  και με βάση το (3.6) συγκρίνει τις πιθανότητες κλάσης για να υποδείξει τη μεγαλύτερη (winning class). Επιπλέον, με τη χρήση ενός τυπικού κυκλώματος WTA, η κλάση που κερδίζει υποδεικνύεται μέσω ενός ψηφιακού one-hot-vector  $[I_1, \dots, I_{N_{cla}}]$  (τα ρεύματα  $[I_i]_{i=1}^{N_{cla}}$  είναι σε δυαδική μορφή) [46]. Επομένως, ολόκληρη η έξοδος του ταξινομητή είναι ψηφιακή.

Η προτεινόμενη αρχιτεκτονική δεν έχει εγγενείς περιορισμούς. Ωστόσο, οι περιορισμοί επιβάλλονται από δομικά στοιχεία που χρησιμοποιούνται. Συγκεκριμένα, ο αριθμός των κλάσεων περιορίζεται από την ικανότητα του κυκλώματος WTA να συγκρίνει με ακρίβεια μεγάλο αριθμό εισόδων. Ομοίως, η ποιότητα του αθροίσματος του ρεύματος εξόδου του συμπλέγματος καθορίζει τον μέγιστο αριθμό υπο-κατηγοριών. Η άθροιση μεγάλου αριθμού μεμονωμένων ρευμάτων θα έχει ως αποτέλεσμα ανεπιθύμητες παραμορφώσεις. Ο αριθμός των διαστάσεων εισόδου εξαρτάται από το πραγματοποιηθέν κύκλωμα πολυδιάστατης συνάρτησης Gauss. Υπάρχουν πολλά κυκλώματα που παράγουν Gaussian συναρτήσεις, αλλά οι πολυδιάστατες Gaussian συναρτήσεις εξακολουθούν να αποτελούν ανοιχτό ερευνητικό θέμα [38]. Τα Bump κυκλώματα που αναπτύχθηκαν σε αυτήν την διπλωματική, είχαν σαν στόχο και την ορθή λειτουργία για πολυδιάστατες συναρτήσεις.

### 3.4 Εξαγωγή Παραμέτρων

Σε αυτή την ενότητα, εξηγείται ο τρόπος εκπαίδευσης του αναλογικού μοντέλου καθώς και παρουσιάζονται μερικές ενδιαφέρουσες ιδιότητες που έχει η αρχιτεκτονική σχετικά με την εύκολη προσαρμογή της σε διαφορετικά προβλήματα.

### 3.4.1 Εκπαίδευση

Για την λειτουργία του ταξινομητή χρειάζεται να έχει παραχθεί ένα μεγάλο πλήθος παραμέτρων (τάσεις  $V_r$ ,  $V_c$  και ρεύματα πόλωσης  $I_{bias}$ ). Η προτεινόμενη σχεδίαση όμως δεν είναι ικανή για on-chip εκπαίδευση. Για την παραγωγή αυτών των παραμέτρων, χρησιμοποιείται ένα μοντέλο που έχει αναπτυχθεί και εκπαιδευτεί σε λογισμικό. Το μοντέλο αυτό έχει ίδιο αριθμό κλάσεων, υποκατηγοριών ανά κλάση και διαστάσεων εισόδου με το αναλογικό μοντέλο. Τα σύνολα δεδομένων είναι διαθέσιμα σε ψηφιακή μορφή και τα χαρακτηριστικά τους υποβάλλονται σε προ-επεξεργασία για να ληφθεί υπόψη το λειτουργικό εύρος του κυκλώματος (σε αυτή την διπλωματική, το λειτουργικό εύρος για όλες τις τάσεις τέθηκε ως το  $[-0.1, 0.2]V$ ). Έπειτα, από το μοντέλο εξάγονται οι μέσες τιμές και οι διασπορές κάθε υπο-κατηγορίας και μετατρέπονται στις κατάλληλες τάσεις  $V_r$  και  $V_c$ , αντίστοιχα. Είναι σημαντικό ότι ενώ το  $V_r$  αντιστοιχεί στην μέση τιμή, η μετατροπή της διασποράς στην κατάλληλη τιμή  $V_c$  είναι μη-γραμμική και περιορίζεται από τα όρια του  $V_c$  (που συνήθως είναι τα όρια τροφοδοσίας).

Τα ρεύματα πόλωσης των  $N_d$ -D Bump έχουν πολλαπλό ρόλο: αντιπροσωπεύουν τις *prior* πιθανότητες της κάθε κλάσης, τα βάρη της κάθε υπο-κατηγορίας και τον όρο κανονικοποίησης της εξίσωσης (2.2). Αυτοί οι 3 όροι πολλαπλασιάζονται στο λογισμικό και οι τιμές που προκύπτουν πρέπει να κανονικοποιηθούν σε ένα αποδεκτό εύρος λειτουργίας για τον κύκλωμα. Για να οριστεί ένα κατάλληλο εύρος, πρέπει τα ρεύματα να είναι αρκετά μεγάλα ώστε το κύκλωμα να λειτουργεί σωστά και ταυτόχρονα όχι πολύ μεγάλα, ώστε να μην αυξηθεί υπερβολικά η κατανάλωση του κυκλώματος. Στους ταξινομητές που σχεδιάστηκαν σε αυτή την διπλωματική, το εύρος αυτό είναι στα  $16nA$ - $25nA$ . Δυστυχώς, αυτή η περιορισμένη κανονικοποίηση αφαιρεί μεγάλο μέρος τις πληροφορίας. Όλη η διαδικασία εκπαίδευσης χρειάζεται να εκτελεστεί μόνο μία φορά και οι παράμετροι που προκύπτουν εξάγονται και αποθηκεύονται σε μια αναλογική μνήμη [52].

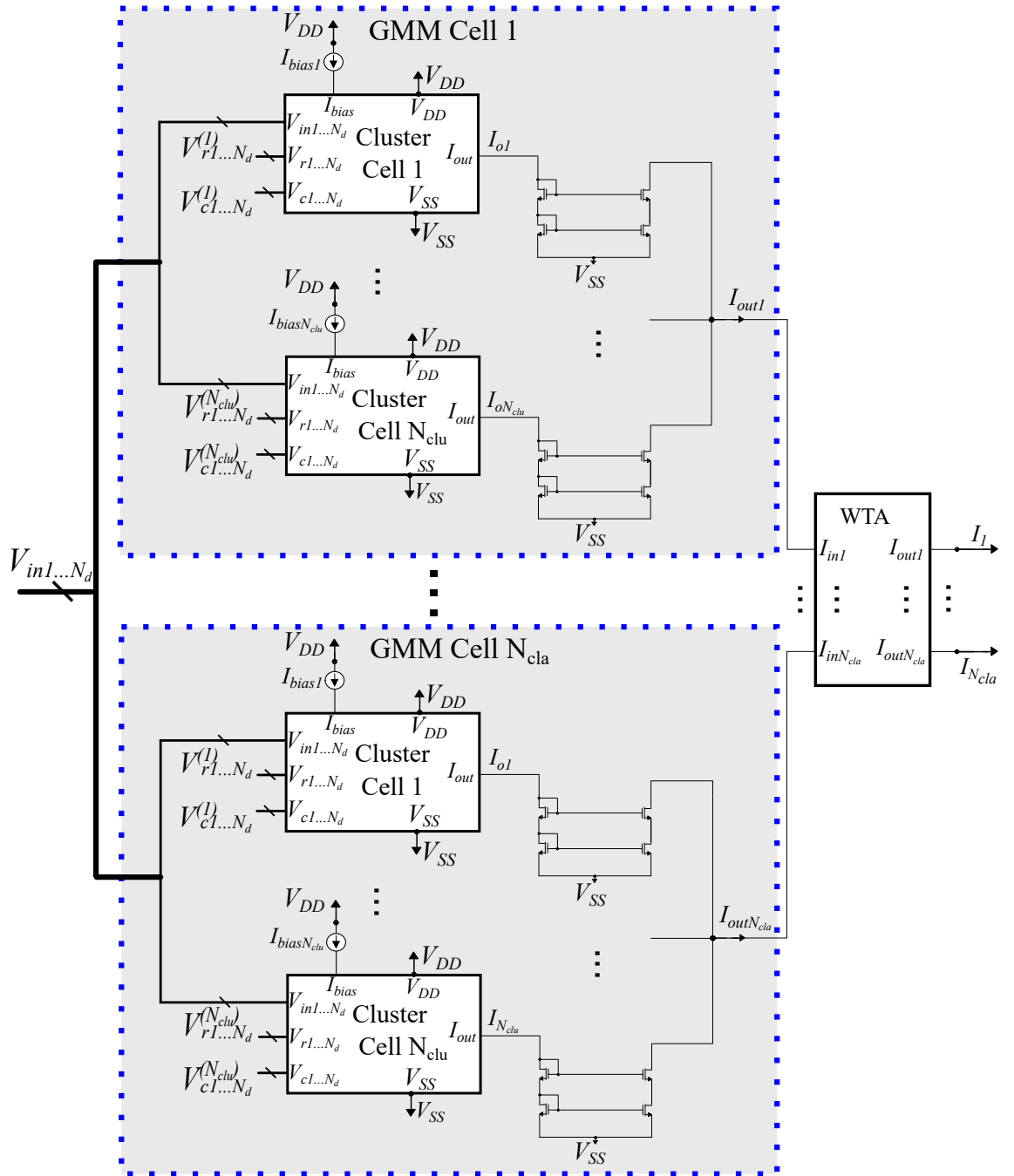
### 3.4.2 Προσαρμοστικότητα

Μία ενδιαφέρουσα ιδιότητα που μπορεί να προκύψει από την εκπαίδευση του μοντέλου είναι η προσαρμογή του ταξινομητή ακόμα και μετά τον αρχικό σχεδιασμό του. Συγκεκριμένα, με την σχεδίαση ενός κυκλώματος για ένα σχετικά μεγάλο σύστημα (μεγάλες τιμές στα  $N_{cla}$ ,  $N_{clu}$ ,  $N_d$ ), οποιοσδήποτε συνδυασμός  $m \leq N_{cla}$  κλάσεων,  $k \leq N_{clu}$  υπο-κατηγοριών και  $n \leq N_d$  διαστάσεων εισόδου, είναι δυνατός χρησιμοποιώντας το ίδιο layout. Η ιδιότητα αυτή προκύπτει άμεσα από τα κυκλώματα Bump: η μείωση στην διάσταση εισόδου (στον

αριθμό των εισόδων) γίνεται μέσω των παραμέτρων τάσης  $V_r$  και  $V_c$ , καθώς και της εισόδου  $V_{in}$ . Για τα πρώτα  $(N_d - n)$  από τα  $N_d$  διαδοχικά συνδεδεμένα Bumps, και οι τρεις τάσεις  $V_r$ ,  $V_c$ ,  $V_{in}$  θα πρέπει να οριστούν στην υψηλότερη αποδεκτή τιμή τάσης, συνήθως στην θετική τάση τροφοδοσίας. Με αυτόν τον τρόπο, τα  $(N_d - n)$  λειτουργούν σαν buffers ρεύματος, προκαλώντας μόνο μια μικρή μεταβολή στο ρεύμα πόλωσης. Συνεπώς, οι είσοδοι και οι παράμετροι του συστήματος κατευθύνονται στα υπόλοιπα  $n$  Bump, με εξαίρεση το ρεύμα πόλωσης  $I_{bias}$  το οποίο εισάγεται στο πρώτο Bump. Η μείωση στον αριθμό των διαστάσεων δυστυχώς δεν μειώνει την κατανάλωση του συστήματος.

Η μείωση του αριθμού των κλάσεων επιτυγχάνεται κυρίως μέσω του  $I_{bias}$ . Συγκεκριμένα, μηδενίζοντας το  $I_{bias}$  ένα  $N_d$ -D Bump θεωρείται αδρανές, και η έξοδος του είναι σταθερή σε μερικά  $\mu A$ . Επιπλέον, τα αποτελέσματα ενισχύονται αν όλες οι παράμετροι  $V_r$  και  $V_c$  ρυθμιστούν στη χαμηλότερη αποδεκτή τιμή τάσης (για παράδειγμα στην αρνητική τροφοδοσία) και όλες οι είσοδοι  $V_{in}$  στην υψηλότερη αποδεκτή τάση. Κατά συνέπεια, η απόρριψη μιας κλάσης γίνεται με μηδενισμό των ρευμάτων πόλωσης όλων των υπο-κατηγοριών της. Αντίθετα με τον αριθμό των διαστάσεων, σε περίπτωση ανενεργών κλάσεων ή υπο-κατηγοριών οι οδηγίες αυτές μειώνουν σημαντικά την κατανάλωση του ταξινομητή.

Από την ικανότητα του κυκλώματος να θέτει υπο-κατηγορίες σε αδράνεια, προκύπτουν δύο επιπλέον τρόποι λειτουργίας του ταξινομητή. Πρώτον, το κύκλωμα είναι ικανό να εισέλθει σε συνολική κατάσταση αδράνειας, στην οποία θα μειωθεί σημαντικά η κατανάλωση του. Αυτό μπορεί να γίνει και αν απλά τεθούν όλες οι τάσεις εισόδου  $V_{in}$  στη χαμηλότερη αποδεκτή τιμή (σε περίπτωση που δεν μεταβάλλονται εύκολα τα ρεύματα πόλωσης). Η δεύτερη ιδιότητα είναι η τοποθέτηση  $N_d$ -D Bump με διαφορετικές ιδιότητες, όπως να μπορούν να παράγουν Γκαουσιανές με διαφορετικές διασπορές. Έπειτα, επιλέγονται τα καταλληλότερα  $N_d$ -D Bump για την κάθε υπο-κατηγορία, ενώ τα υπόλοιπα θέτονται σε αδράνεια.



Σχήμα 3.2: Αναλογικός ταξινομητής που βασίζεται στο GMM με  $N_{cla}$  κλάσεις,  $N_{clu}$  υπο-κατηγορίες ανά κλάση και  $N_d$ -D εισόδους.



## Κεφάλαιο 4

# Εφαρμογές και Προσομοιώσεις

Σε αυτό το κεφάλαιο θα αναλυθούν τα αποτελέσματα των παραπάνω αρχιτεκτονικών σε διαφορετικές πραγματικές εφαρμογές. Για την εξαγωγή των αποτελεσμάτων έχει σχεδιαστεί το layout του κάθε ταξινομητή. Η υλοποίηση των layout βασίζεται στην τεχνική κοινού κέντρου (common-centroid technique) και χρησιμοποιούνται επιπλέον εικονικά τρανζίστορ για την αποφυγή των mismatches και κατασκευαστικών αστοχιών [53, 54, 55].

### 4.1 Διάγνωση Ασθενειών Θυρεοειδή

Σε αυτήν την ενότητα παρουσιάζονται τα αποτελέσματα του αναλογικού Bayes ταξινομητή σε μια πραγματική εφαρμογή διάγνωσης 2 βασικών ασθενειών που σχετίζονται με τον θυρεοειδή: υπερθυρεοειδισμός και υποθυρεοειδισμός. Η εφαρμογή αυτή χρησιμοποιείται ως proof-of-concept, αλλά τα αποτελέσματα του ταξινομητή είναι αρκετά υψηλά. Συνεπώς, το κύκλωμα θα μπορούσε να λειτουργήσει αυτόνομα ή συμβουλευτικά σε πραγματικές περιστάσεις.

#### 4.1.1 Θυρεοειδής

Ο θυρεοειδής είναι ένας πολύ μικρός αδένας σε σχήμα πεταλούδας στο μπροστινό μέρος του λαιμού [56, 57]. Ο θυρεοειδής είναι μέρος του ενδοκρινικού συστήματος, το οποίο αποτελείται από αδένες που παράγουν, αποθηκεύουν και απελευθερώνουν ορμόνες στην κυκλοφορία του αίματος, ώστε να φτάσουν στα κύτταρα του σώματος. Ο κύριος ρόλος του είναι η ρύθμιση του μεταβολισμού (έχει σημαντικούς ρόλους στη ρύθμιση πολλών μεταβολικών διεργασιών) πα-

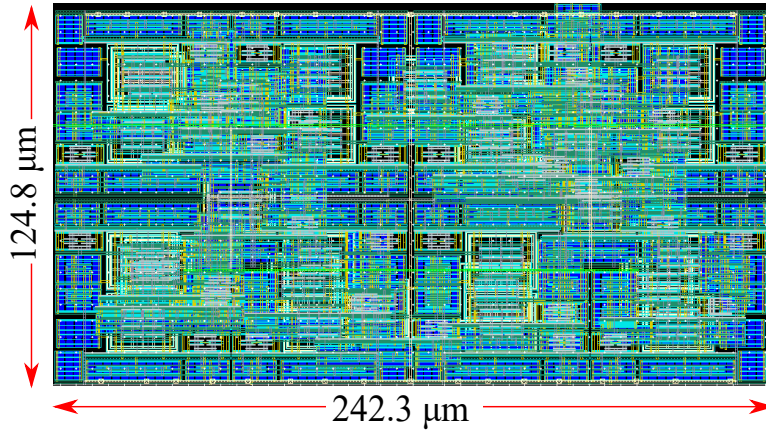
ράγοντας ορμόνες. Βοηθάει επίσης στον έλεγχο των μυών, στην ανάπτυξη του εγκεφάλου, στη ρύθμιση της διάθεσης και στη λειτουργία του πεπτικού συστήματος.

Οι διαταραχές του θυρεοειδούς είναι καταστάσεις που επηρεάζουν αυτό το όργανο και συνήθως οδηγούν στην ασυνήθιστη παραγωγή ορμονών [58]. Υπάρχουν διάφορες ασθένειες σχετικά με τον θυρεοειδή όπως θυρεοειδίτιδα, οζιδια θυρεοειδούς, βρογχοκήλη και καρκίνο του θυρεοειδούς. Οι δύο όμως πιο συνηθισμένοι τύποι θυρεοειδικής νόσου είναι ο υπερθυρεοειδισμός και ο υποθυρεοειδισμός. Ο υποθυρεοειδισμός προκύπτει από τον θυρεοειδή αδένα που παράγει ανεπαρκή ποσότητα θυρεοειδικών ορμονών. Τα κύρια συμπτώματά του περιλαμβάνουν κόπωση και κόπωση, προβλήματα ύπνου, κατάθλιψη, ευαισθησία στις χαμηλές θερμοκρασίες, ξηρό δέρμα και μαλλιά, δυσκολία συγκέντρωσης, συχνές και βαριές περιόδους και πόνο στις αρθρώσεις και τους μύες. Από την άλλη πλευρά, ο υπερθυρεοειδισμός περιγράφει την υπερβολική παραγωγή θυρεοειδικής ορμόνης, μια λιγότερο συχνή πάθηση από τον υποθυρεοειδισμό. Τα κύρια συμπτώματα είναι άγχος, ευερεθιστότητα ή διάθεση, υπερκινητικότητα, εφίδρωση ή ευαισθησία σε υψηλές θερμοκρασίες, τρέμουλο χεριών, απώλεια μαλλιών και απώλεια ή ελαφριά εμμηόρροια. Μετά την ανίχνευση, οι θεραπευτικές επιλογές εξαρτώνται από τη συγκεκριμένη μορφή της νόσου του θυρεοειδούς και περιλαμβάνουν φάρμακα, ραδιενεργό ιώδιο και μερικές φορές χειρουργική επέμβαση.

#### 4.1.2 Επίδοση του Αναλογικού Ταξινομητή

Σε αυτήν την ενότητα παρουσιάζονται τα αποτελέσματα του αναλογικού Bayes ταξινομητή σε μια πραγματική εφαρμογή διάγνωσης 2 ασθενειών του θυρεοειδή. Το σύνολο δεδομένων αποκτήθηκε από το αποθετήριο μηχανικής μάθησης του Πανεπιστημίου της Καλιφόρνια, Irvine (UCI) [57] και περιέχει 5 μετρήσεις εξετάσεων αίματος (που σχετίζονται με τον θυρεοειδή) για ασθενείς με φυσιολογικό θυρεοειδή, υποθυρεοειδισμό και υπερθυρεοειδισμό. Για την εφαρμογή αυτή, το layout του κυκλώματος με 3 κλάσεις και εισόδους 5 διαστάσεων φαίνεται στο σχήμα 4.1. Στο ταξινομητή αυτόν έχει χρησιμοποιηθεί το *Neuron Bump*, Σχήμα 2.5, και ένα Lazzaro WTA.

Για να ελεγχθεί ο ταξινομητής Bayes τόσο ως προς την ακρίβεια ταξινόμησης όσο και για τη συμπεριφορά του κυκλώματος σε παραλλαγές PVT πραγματοποιούνται δύο ξεχωριστά τεστ. Το πρώτο είναι μια σύγκριση μεταξύ της αναλογικής υλοποίησης και μιας που βασίζεται σε λογισμικό. Συγκεκριμένα, πραγματοποιούνται 20 ξεχωριστές επαναλήψεις εκπαίδευσης που βασίζονται σε λογισμικό για να αποφευχθούν τυχαίες επιρροές και τα αποτελέσματα συνοψίζονται στον Πίνακα 4.1. Η μέση ακρίβεια της αναλογικής αρχιτεκτονικής είναι



Σχήμα 4.1: Layout του Bayesian ταξινομητή.

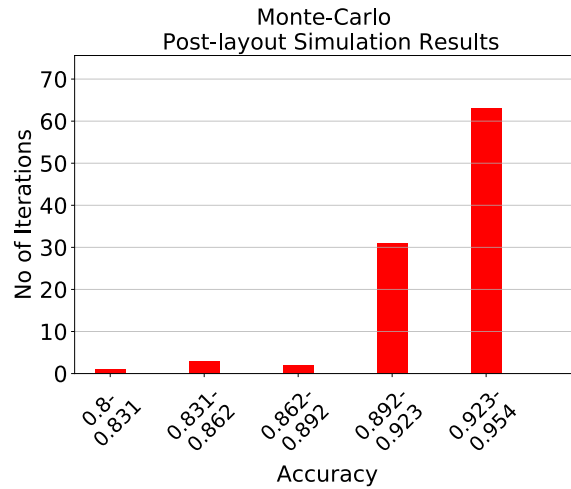
μόνο περίπου 2% χαμηλότερη από της υλοποίησης σε λογισμικό. Το δεύτερο τεστ, είναι μια ανάλυση Monte-Carlo για  $N = 100$  σημεία, για μία από τους προηγούμενους 20 δοκιμές. Το ιστόγραμμα της Monte-Carlo ανάλυσης φαίνεται στο Σχήμα 4.2. Η μέση τιμή του είναι  $\mu_M = 0.916$  με τυπική απόκλιση  $\sigma_M = 0.023$ . Τα δύο τεστ επιβεβαιώνουν τη σωστή απόδοση και λειτουργία της προτεινόμενης αρχιτεκτονικής.

Πίνακας 4.1: Αποτελέσματα ακρίβειας (για 20 επαναλήψεις).

Μέθοδος	Καλύτερο	Χειρότερο	Μέση τιμή	Τυπική απόκλιση
Λογισμικό	1.000	0.9231	0.9639	0.0184
Αναλογικά	0.9692	0.9231	0.9423	0.0149

## 4.2 Πρόβλεψη Επιληπτικών Κρίσεων

Σε αυτήν την ενότητα παρουσιάζονται τα αποτελέσματα του αναλογικού GMM ταξινομητή σε μια πραγματική εφαρμογή πρόβλεψης επιληπτικών κρίσεων. Λόγω της δυσκολίας αυτής της εφαρμογής, ο αναλογικός ταξινομητής δεν μπορεί να χρησιμοποιηθεί αυτόνομα. Αντί αυτού, η αρχιτεκτονική προτείνεται ως ένα κύκλωμα αφύπνισης ελάχιστης ισχύος.



Σχήμα 4.2: Αποτελέσματα προσομοίωσης Μοντε Καρλο μετά τη διάταξη.

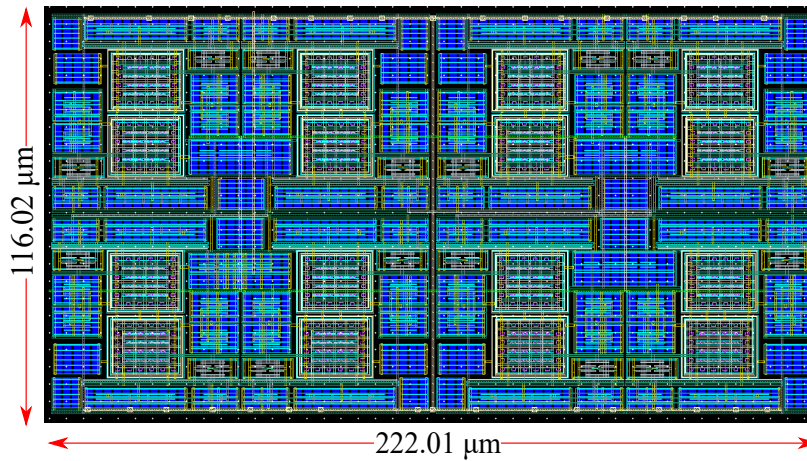
#### 4.2.1 Επιληπτικές Κρίσεις

Η επιληψία είναι μια χρόνια νευρολογική διαταραχή που επηρεάζει άτομα όλων των ηλικιών [59, 60]. Συνήθως προκαλείται από μια ξαφνική μη φυσιολογική εκκένωση εγκεφαλικών νευρώνων που οδηγεί σε προσωρινή δυσλειτουργία. Τα αποτελέσματα των επιληπτικών κρίσεων κυμαίνονται από σχεδόν απαρατήρητα έως παραλυτικά ή ακόμα και θανατηφόρα [61]. Η φαινομενικά απρόβλεπτη εμφάνισή τους δυσκολεύει την ζωή των ασθενών και καθιστά επικίνδυνες διάφορες καθημερινές δραστηριότητες. Επομένως, η πρόβλεψη των επιληπτικών κρίσεων είναι μείζονος σημασίας για την ποιότητα ζωής τους.

Σχετικά με τις επιληπτικές κρίσεις, υπάρχουν τέσσερις διαφορετικές καταστάσεις του ασθενούς: (α) pre-ictal, (β) ictal, (γ) post-ictal και (δ) inter-ictal [60, 59]. Τα στάδια (α)-(γ) αναφέρονται στις περιόδους λίγο πριν, κατά τη διάρκεια και λίγο μετά από μια κρίση, αντίστοιχα, και (δ) αναφέρεται στην περίοδο μεταξύ δύο κρίσεων. Με βάση την ανάλυση που παρουσιάζεται στο [62], η pre-ictal και η post-ictal περίοδοι ορίζονται, εδώ, ως μία ώρα πριν και μετά την κρίση, αντίστοιχα. Ο στόχος ενός συστήματος πρόβλεψης των επιληπτικών κρίσεων είναι να διακρίνει επιτυχώς τις pre-ictal από τις inter-ictal περιόδους. Στην βιβλιογραφία υπάρχουν πολλές αναφορές στις οποίες επιτυγχάνεται η πρόβλεψη των κρίσεων αυτών με πολύ υψηλή ακρίβεια, με την χρήση συστημάτων και μοντέλων μηχανικής μάθησης [63, 64, 62, 65, 66]. Αν αυτά τα μοντέλα πρόβλεψης μπορούσαν να παρακολουθούν διαρκώς τον ασθενή, αυτός θα μπορεί να ειδοποιηθεί έως και 1 ώρα νωρίτερα για τον επικείμενο κίνδυνο και να προετοιμαστεί κατάλληλα.

### 4.2.2 Επίδοση του Αναλογικού Ταξινομητή

Τα δεδομένα λαμβάνονται από τη βάση δεδομένων CHB-MIT Scalp EEG [67] και περιέχουν σήματα EEG από παιδιά με ανίατη επιληψία. Η κατάσταση των ασθενών κατά την διάρκεια των καταγραφών έχει επισημανθεί από ειδικούς γιατρούς. Για την εφαρμογή αυτή, το layout του κυκλώματος με 2 κλάσεις, 2 υποκατηγορίες ανά κλάση και εισόδους τεσσάρων διαστάσεων φαίνεται στο σχήμα 4.3. Στο ταξινομητή, έχει χρησιμοποιηθεί το αριστερά Bump του Σχήματος 2.7 και ένα Lazzaro WTA.



Σχήμα 4.3: Layout του GMM ταξινομητή, για την πρόβλεψη των επιληπτικών κρίσεων.

Έχοντας σαν στόχο την χαμηλή κατανάλωση χωρίς να μειωθεί η ακρίβεια του κυκλώματος, χρησιμοποιούνται μόνο 4 χαρακτηριστικά για τον ταξινομητή. Αυτά είναι η διακύμανση του σήματος και οι ενέργειες του στην άλφα και στο πρώτο και δεύτερο μισό της γάμμα ζώνης συχνοτήτων [68]. Αυτά τα χαρακτηριστικά μπορούν να προκύψουν απευθείας από τα ακατέργαστα σήματα EEG χρησιμοποιώντας τεχνικές εξαγωγής αναλογικών χαρακτηριστικών [69, 70]. Οι απαραίτητες παράμετροι του συστήματος παρέχονται μέσω εκπαίδευσης που βασίζεται σε λογισμικό, όπως έχει εξηγηθεί.

Προκειμένου να λειτουργεί ως κύκλωμα αφύπνισης ελάχιστης ισχύος, πρέπει να προβλέπει με επιτυχία όλες τις πιθανές κρίσεις διατηρώντας παράλληλα χαμηλό αριθμό ψευδώς θετικών ειδοποιήσεων. Το πρώτο υπολογίζεται με το sensitivity [71] του ταξινομητή και είναι απαραίτητος παράγοντας για την υγεία

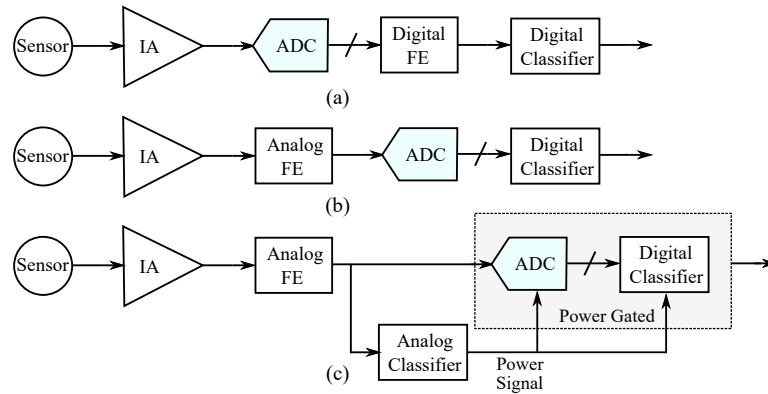
του ασθενούς:

$$\text{sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}. \quad (4.1)$$

Το δεύτερο ελαχιστοποιεί τον χρόνο κατά τον οποίο ενεργοποιείται το υψηλής ακρίβειας και κατανάλωσης, ψηφιακό back-end. Αυτό οδηγεί σε σημαντική μείωση της κατανάλωσης ενέργειας για ολόκληρο το σύστημα πρόβλεψης κρίσεων, όπως φαίνεται στο σχήμα 4.4 (γ). Ένα κατάλληλο μέτρο για την ποσοτικοποίηση αυτής της μείωσης είναι το specificity [71] του αναλογικού ταξινομητή, που δίνεται από:

$$\text{specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}. \quad (4.2)$$

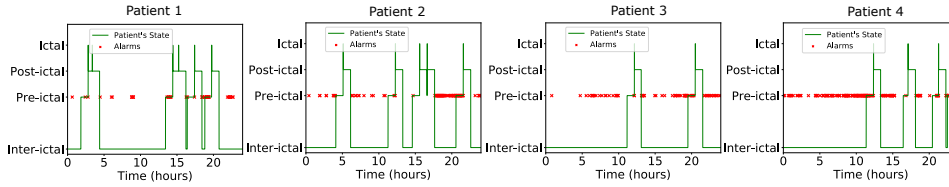
Στην πράξη, αυτή η μέτρηση είναι η αναλογία του χρόνου που το ψηφιακό back-end είναι αδρανές προς τη διάρκεια όλων των inter-ictal περιόδων (κατά τις οποίες δεν υπάρχει κίνδυνος κρίσης).



Σχήμα 4.4: Σύγκριση αρχιτεκτονικής. (α) Πλήρως ψηφιακό σύστημα. (β) Εξαγωγή αναλογικών χαρακτηριστικών. (γ) Αναλογικό wake-up κύκλωμα, όπου το ψηφιακό back-end ενεργοποιείται και απενεργοποιείται με βάση την έξοδο του αναλογικού ταξινομητή χαμηλής ισχύος.

Για να επισημανθεί η σωστή λειτουργία του GMM ταξινομητή, διεξάγονται τα δύο τεστ που εξηγήθηκαν και στην προηγούμενη ενότητα. Το πρώτο συγκρίνει την προτεινόμενη αρχιτεκτονική με μια βασισμένη σε λογισμικό ως προς το specificity, το οποίο αντιστοιχεί και στην μείωση κατανάλωσης του ψηφιακού κυκλώματος. Στον Πίνακα 4.2 συνοψίζονται τα αποτελέσματα για τις 20 επαναλήψεις. Το μέσο specificity της αναλογικής αρχιτεκτονικής είναι μόνο 2% χαμηλότερο από της υλοποίησης σε λογισμικό. Για λόγους επίδειξης, η κατάσταση τεσσάρων ασθενών μαζί με τις προβλέψεις του αναλογικού ταξινομητή

παρουσιάζονται στο Σχήμα 4.5. Για το δεύτερο τεστ, το ιστόγραμμα Monte-Carlo για  $N = 100$  σημεία, που φαίνεται στο σχήμα 4.6, έχει μέση τιμή  $\mu_M = 69.93\%$  και τυπική απόκλιση  $\sigma_M = 0.41\%$ . Και τα δύο τεστ επιβεβαιώνουν την απόδοση, την υψηλή ακρίβεια και την επιθυμητή ευαισθησία του προτεινόμενου ταξινομητή.



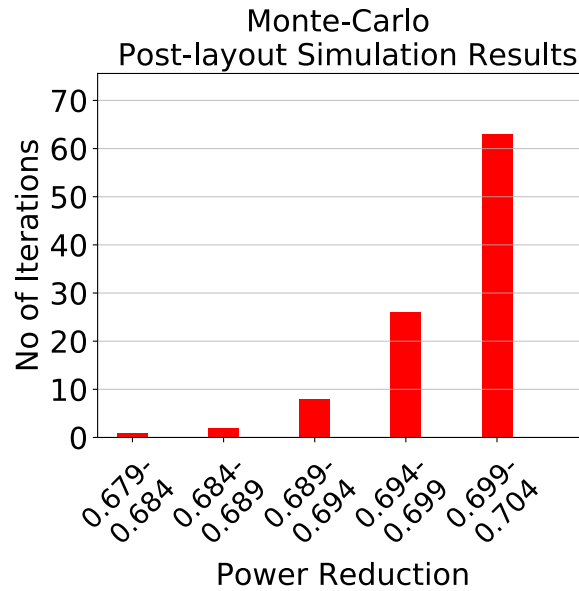
Σχήμα 4.5: Οι συναγερμοί που ενεργοποιούνται από τον ταξινομητή για τέσσερις ασθενείς σε ένα 24ωρο. Η ιδανική συμπεριφορά είναι η άνοδος τουλάχιστον ενός συναγερμού σε κάθε pre-ictal περίοδο, χωρίς να ακούγονται συναγερμοί κατά τη διάρκεια των inter-ictal. Οι ictal και post-ictal περιοχές είναι αδιάφορες για τον ταξινομητή.

Πίνακας 4.2: Αποτελέσματα μείωσης κατανάλωσης ισχύος (για 20 επαναλήψεις).

Μέθοδος	Καλύτερο	Χειρότερο	Μέση τιμή	Τυπική απόκλιση
Λογισμικό	71.30%	71.08%	71.27%	0.07%
Αναλογικά	70.65%	67.39%	69.07%	0.51%

### 4.3 Άλλες Εφαρμογές

Σε αυτή την ενότητα, δοκιμάζεται η αρχιτεκτονική και η προσαρμοστικότητα του GMM ταξινομητή, με το *Cascade Bump* και ένα *Cascaded WTA* σε ένα δοκιμαστικό και σε δύο πραγματικά σύνολα δεδομένων. Το δοκιμαστικό σύνολο δεδομένων έχει  $N_{cla} = 2$  κατηγορίες και  $N_d = 2$ -D δεδομένα εισόδου, ενώ τα άλλα δύο σύνολα δεδομένων έχουν  $N_{cla} = 4$  και  $N_{cla} = 3$  κατηγορίες και  $N_d = 13$ -D δεδομένα εισόδου. Ο αριθμός των υπο-κατηγοριών για τον ταξινομητή που στοχεύει το δοκιμαστικό σύνολο δεδομένων έχει οριστεί σε  $N_{clu} = 2$ , ενώ για τα άλλα δύο σύνολα δεδομένων είναι  $N_{clu} = 4$  και  $N_{clu} = 3$ . Για επίδειξη, οι παραπάνω πληροφορίες δεδομένων παρέχονται επίσης στον Πίνακα 4.3. Χρησιμοποιώντας τις τεχνικές που εξηγούνται στην Ενότητα 3.4.2, το layout του σχήματος 4.7 με 4 κλάσεις, 4 υπο-κατηγορίες ανά κλάση και 16-D εισόδους μπορεί να αντιμετωπίσει και τα τρία αναφερόμενα σύνολα δεδομένων.



Σχήμα 4.6: Αποτελέσματα προσομοίωσης ανάλυσης ευαισθησίας Μοντε-κάρλο μετά τη διάταξη.

Πίνακας 4.3: Ιδιότητες συνόλου δεδομένων

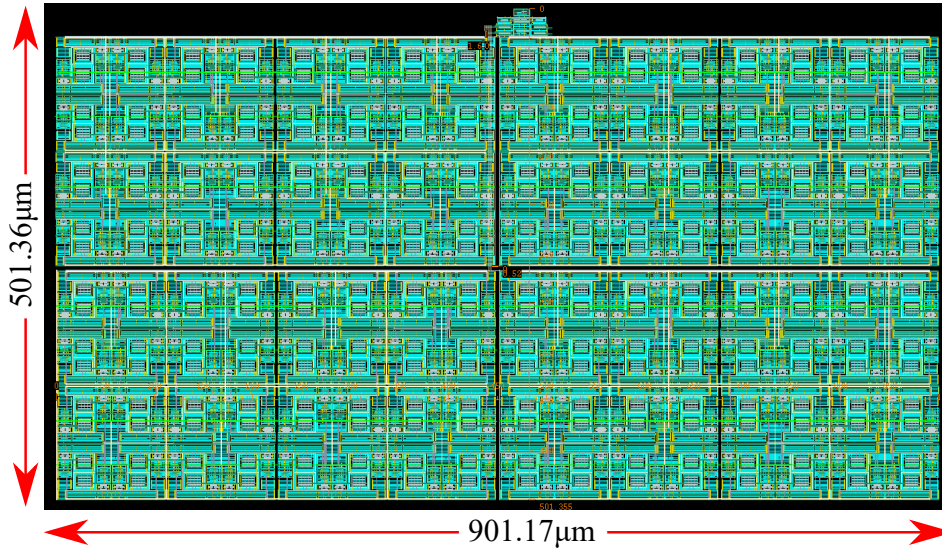
	Αρ. κλάσεων	Αρ. υπο- κατηγοριών	Αρ. διαστάσεων	Αρ. στοιχείων
2-D Toy	2	2	2	10000
CWRU	4	4	13	580
VSBD	3	3	13	354

Για σύγκριση, κατασκευάζονται δύο αρχιτεκτονικές GMM ταξινόμητή. Η πρώτη, που ονομάζεται Proposed I, χρησιμοποιεί το *Cascade Bump* και ένα *Cascaded WTA*. Το δεύτερο, το Proposed II, χρησιμοποιεί το Bump του [45] και ένα Lazzaro WTA. Και οι δύο ταξινόμητές συγκρίνονται και με μία υλοποίηση σε λογισμικό.

#### 4.3.1 Δοκιμαστικό Δισδιάστατο Σύνολο Δεδομένων

Το πρώτο πρόβλημα ταξινόμησης εξετάζει ένα δοκιμαστικό 2-D σύνολο δεδομένων. Τα όρια απόφασης αυτού του απλού προβλήματος μπορούν να απεικονιστούν σε μια εικόνα, επιτρέποντας επομένως την οπτική σύγκριση μεταξύ των



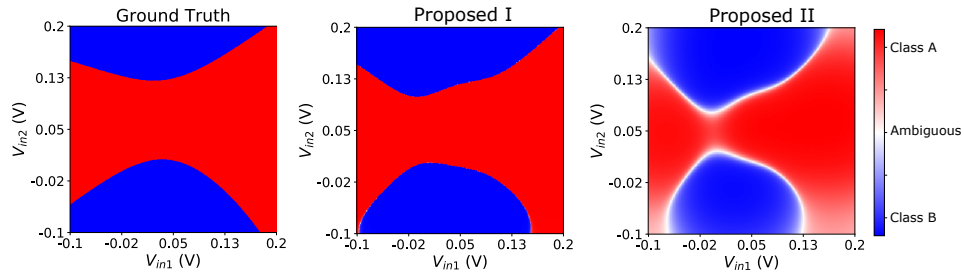


Σχήμα 4.7: Διάταξη μιας προτεινόμενης αρχιτεκτονικής GMM (Proposed I).

αποτελεσμάτων των εξεταζόμενων κυκλωμάτων. Χρησιμοποιώντας ένα σύνολο ομοιόμορφα καταναμημένων σημείων του χώρου εισόδου  $[-0.1, 0.2]^2$  ως είσοδο για τους ταξινομητές, οι προβλέψεις τους παρουσιάζονται στο Σχήμα 4.8. Σε αυτό το πρόβλημα, οι προβλέψεις του ταξινομητή φτιαγμένου στο λογισμικό θα θεωρηθούν ως σωστές (Ground Truth). Εφόσον τα όρια απόφασης αντιπροσωπεύουν την πρόβλεψη του συστήματος για οποιαδήποτε δεδομένη είσοδο, είναι λογικό να υποθέσουμε ότι ένα όριο απόφασης που μοιάζει με το Ground Truth έχει καλύτερη ακρίβεια. Σε αυτή την περίπτωση, η αρχιτεκτονική Proposed I έχει πιο περιορισμένες λευκές (διφορούμενες) περιοχές και μοιάζει περισσότερο με το Ground Truth, σε σύγκριση με την Proposed II. Αυτό επαληθεύεται στον Πίνακα 4.4, όπου παρουσιάζονται οι ακρίβειες των 2 αρχιτεκτονικών.

Πίνακας 4.4: Ακρίβεια ορίων απόφασης.

Ταξινομητής	Ακρίβεια
Proposed I	0.932
Proposed II	0.873



Σχήμα 4.8: Διαδιάστατα όρια απόφασης (αριστερά) Ground Truth (μέση) τα αποτελέσματα του ταξινομητή Proposed I (δεξιά) τα αποτελέσματα του ταξινομητή Proposed II

### 4.3.2 Πραγματικό Σύνολο Δεδομένων 1

Το σύνολο δεδομένων του Case Western Reserve University (CWRU) [72] περιέχει δεδομένα επιταχυνσιόμετρου κινητήρων που είτε λειτουργούν σωστά είτε έχουν υποστεί ζημιά σε 3 διαφορετικά σημεία. Αυτό το σύνολο δεδομένων χρησιμοποιείται για την ταξινόμηση των κινητήρων σε τέσσερις διαφορετικές συνθήκες λειτουργίας, δηλαδή να λειτουργούν σωστά ή ελαττωματικά με 3 πιθανά σημεία βλάβης. Σε κάθε κλάση, τα δεδομένα επιταχυνσιόμετρου που περιέχουν καταχωρήσεις 20-δευτερολέπτων. Οι καταχωρήσεις αυτές χωρίζονται σε 20 τμήματα ίσης διάρκειας (1 δευτερόλεπτο το καθένα) τα οποία υποβάλλονται σε επεξεργασία για την παραγωγή των 13 χαρακτηριστικών που εμφανίζονται στον Πίνακα 4.5. Αυτό έχει ως αποτέλεσμα 580 δεδομένα εισόδου 13-διαστάσεων που σχηματίζουν τα σετ εκπαίδευσης και δοκιμών χρησιμοποιώντας έναν διαχωρισμό 70% – 30%. Και τα δύο σετ εκπαίδευσης και δοκιμών θεωρούνται ισορροπημένα, αφού οι τέσσερις κατηγορίες έχουν 98, 112, 112, 84 και 42, 48, 48, 36 στοιχεία εκπαίδευσης και δοκιμής, αντίστοιχα.

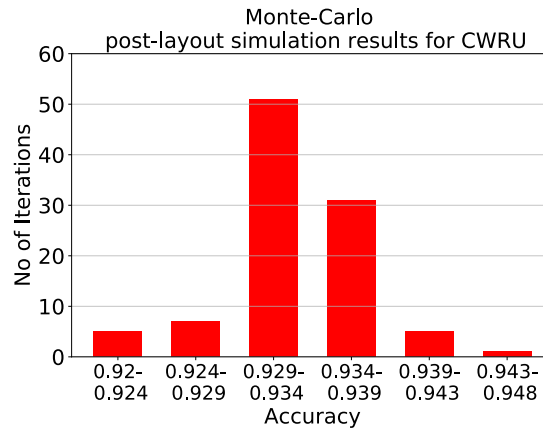
Για να επισημανθεί η σωστή λειτουργία του GMM ταξινομητή, διεξάγονται τα δύο τεστ που εξηγήθηκαν και στις προηγούμενες ενότητες. Στον Πίνακα 4.6 συνοψίζονται τα αποτελέσματα για τις 20 επαναλήψεις. Παρατηρείται ότι το Proposed I υπερτερεί του Proposed II και η ακρίβειά του είναι πολύ κοντά στην υλοποίηση που βασίζεται σε λογισμικό. Συγκεκριμένα, τα αποτελέσματα της αρχιτεκτονικής Proposed I έχουν 8 % αυξημένη μέση ακρίβεια και έχουν πολύ μικρότερη τυπική απόκλιση σε σύγκριση με το Proposed II. Για το δεύτερο τεστ, το ιστόγραμμα Monte-Carlo για  $N = 100$  σημεία, που φαίνεται στο σχήμα 4.9, έχει μέση τιμή  $\mu_M = 0.933$  και τυπική απόκλιση  $\sigma_M = 0.005$ . Τα δύο τεστ επιβεβαιώνουν τη σωστή απόδοση και ακρίβεια του ταξινομητή.

Πίνακας 4.5: Εξαγόμενα features [1]

Statistic	Equation	Statistic	Equation
Root mean square	$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	Crest factor	$CF = \frac{\max(x_i)}{RMS}$
Square root of amplitude	$SRA = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2$	Impulse factor	$IF = \frac{N \cdot \max(x_i)}{\sum_{i=1}^N  x_i }$
Kurtosis value	$KV = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu_x}{\sigma_x}\right)^4$	Margin factor	$MF = \frac{\max(x_i)}{SRA}$
Skewness value	$SV = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu_x}{\sigma_x}\right)^3$	Frequency center	$FC = \frac{1}{N} \sum_{i=1}^N f_i$
Peak-to-peak value	$PPV = \max(x_i) - \min(x_i)$	Root-mean-square frequency	$RMSF = \sqrt{\frac{1}{N} \sum_{i=1}^N f_i^2}$
Shape factor	$SF = \frac{\max(x_i)}{SV}$	Root variance frequency	$RVF = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - FC)^2}$
Kurtosis factor	$KF = \frac{KV}{RMS^4}$	-	-

Πίνακας 4.6: Αποτελέσματα ακρίβειας στο σύνολο δεδομένων CWRU (για 20 επαναλήψεις).

Μέθοδος	Καλύτερο	Χειρότερο	Μέση τιμή	Τυπική απόκλιση
Λογισμικό	0.9828	0.9310	0.9632	0.0112
Proposed I	0.9655	0.9023	0.9308	0.0164
Proposed II	0.9023	0.7816	0.8512	0.0340



Σχήμα 4.9: Αποτελέσματα προσομοίωσης Monte-Carlo (post-layout) της αρχιτεκτονικής Proposed I στο σύνολο δεδομένων CWRU.

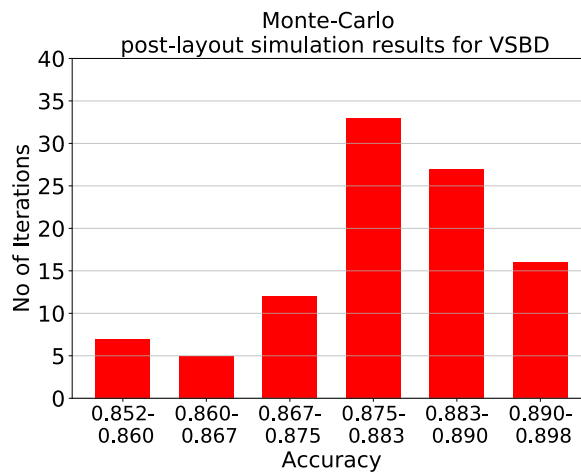
### 4.3.3 Πραγματικό Σύνολο Δεδομένων 2

Το σύνολο δεδομένων Bearing Vibration Data under Time-varying Rotational Speed Conditions (VSBD [73]) από το Mendeley Data [74] περιέχει σήματα δόνησης από ρουλεμάν που είτε λειτουργούν σωστά είτε έχουν υποστεί ζημιά σε 2 πιθανά σημεία. Το VSBD, ομοίως με το CWRU, χρησιμοποιείται για την ταξινόμηση των κινητήρων σε τέσσερις διαφορετικές συνθήκες λειτουργίας, δηλαδή να λειτουργούν σωστά ή ελαττωματικά με 2 πιθανά σημεία βλάβης. Η εξαγωγή χαρακτηριστικών είναι παρόμοια με το σύνολο δεδομένων CWRU με την εξαίρεση ότι εδώ τα δεδομένα είναι 10-δευτερολέπτων. Τόσο τα σετ εκπαίδευσης όσο και τα σετ δοκιμών είναι επίσης ισορροπημένα, αφού και οι τρεις τάξεις έχουν 82 διανύσματα εκπαίδευσης και 36 δοκιμής.

Και για αυτό το σετ δεδομένων, διεξάχθηκαν τα δύο τεστ που εξηγήθηκαν και στις προηγούμενες ενότητες. Τα αποτελέσματα για τις 20 επαναλήψεις συνοψίζονται στον Πίνακα 4.6. Σε αυτό το πρόβλημα, ο ταξινομητής Proposed I είναι σταθερά καλύτερος από τον Proposed II, έχοντας 2% καλύτερη μέση ακρίβεια. Η απόδοσή του είναι πολύ κοντά στην υλοποίηση σε λογισμικό. Στο σχήμα 4.10 φαίνεται το ιστόγραμμα Monte-Carlo για  $N = 100$  σημεία. Η μέση τιμή του είναι  $\mu_M = 0.881$  με τυπική απόκλιση  $\sigma_M = 0.013$ . Και τα δύο τεστ που πραγματοποιήθηκαν επιβεβαιώνουν τη σωστή λειτουργία του Proposed I και την υπεροχή του έναντι του Proposed II.

Πίνακας 4.7: Αποτελέσματα ακρίβειας στο σύνολο δεδομένων VSBD (για 20 επαναλήψεις).

Μέθοδος	Καλύτερο	Χειρότερο	Μέση τιμή	Τυπική απόκλιση
Λογισμικό	0.9537	0.8796	0.9241	0.0198
Proposed I	0.9259	0.8056	0.8704	0.0307
Proposed II	0.9074	0.7870	0.8519	0.0285



Σχήμα 4.10: Αποτελέσματα προσομοίωσης Monte-Carlo (post-layout) της αρχιτεκτονικής Proposed I στο σύνολο δεδομένων VSBD.

## 4.4 Σύγκριση

Σε αυτήν την ενότητα, παρέχεται μια σύνοψη των τεχνικών πτυχών άλλων πρόσφατων υλοποιήσεων αλγορίθμων μηχανικής μάθησης σε αναλογικό υλικό ή με μικρές υλοποιήσεις. Παράλληλα παρουσιάζονται και οι ίδιες μετρικές για τις αρχιτεκτονικές που παρουσιάστηκαν σε αυτήν την διπλωματική.

Αξίζει να σημειωθεί ότι δεν είναι δυνατή μια δίκαιη σύγκριση μεταξύ υλοποιήσεων μηχανικής μάθησης σε υλικό, καθώς σχεδιάζονται για συγκεκριμένες εφαρμογές και υπάρχουν πολλές πτυχές που πρέπει να ληφθούν υπόψη συνδυαστικά, όπως η εφαρμοσμένη τεχνολογία, οι προδιαγραφές ισχύος και επιφάνειας, η ταχύτητα υπολογισμού κ.λπ. Μια σύνοψη απόδοσης παρουσιάζεται στον Πίνακα 4.8. Αυτοί οι ταξινομητές βασίζονται σε αλγόριθμους ή μοντέλα ΜΛ τα οποία λειτουργούν καλύτερα σε συγκεκριμένες εφαρμογές. Αυτά ε-

ίνα Support Vector Machines (SVM) ή Support Vector Data descriptors (SVDD) [75, 76, 77, 78, 79], Radial Basis Function (RBF) Neural Networks (NNs) [80, 81], Object recognition ή Neuro-fuzzy επεξεργαστές [82, 83], Long Short-Term Memory (LSTM) δίκτυα [84] ή ο K-means αλγόριθμος [85].

Ο στόχος αυτής της διπλωματικής ήταν η υλοποίηση ενός χαμηλής κατανάλωσης και επιφάνειας αναλογικού ταξινομητή, εύκολα προσαρμόσιμου σε διαφορετικές εφαρμογές. Σε αυτήν την αρχιτεκτονική, το κύκλωμα Βυμπ έχει κυρίαρχο ρόλο για ολόκληρο τον ταξινομητή. Αυτό δικαιολογεί τη σημασία που δόθηκε στην μελέτη και βελτιστοποίηση του, για την αύξηση στην ακρίβεια του συνολικού ταξινομητή με ελάχιστη αύξηση στην περιοχή και στην κατανάλωση του κυκλώματος. Επιπλέον, ο πλήρως αναλογικός ταξινομητής επιτρέπει ένα πλήρως αναλογικό σύστημα, το οποίο είναι πιο αποδοτικό σε ισχύ και περιοχή από το ψηφιακό ή μεικτό, χωρίς να χρειάζονται μετατροπείς αναλογικού σε ψηφιακό σήμα (ADC).

Πίνακας 4.8: Σύνοψη αλγορίθμων μηχανική μάθησης.

	Τεχνολογία	Υλοποίηση	Ταξινομητής	Αριθμός διαστάσεων	Κατανάλωση	Ταχύτητα επεξεργασίας	Επιφάνεια
Παρούσα δουλειά	90nm	Αναλογική	Bayesian	5	365nW	170K $\frac{\text{classifications}}{\text{second}}$	0.030mm <sup>2</sup>
Παρούσα δουλειά	90nm	Αναλογική	GMM	4	180nW	166K $\frac{\text{classifications}}{\text{second}}$	0.026mm <sup>2</sup>
Παρούσα δουλειά	90nm	Αναλογική	GMM	16	12.0μW	125K $\frac{\text{classifications}}{\text{second}}$	0.451mm <sup>2</sup>
[75]	0.5μm	Μικτή	SVM	N/A	5.9mW	12.8M $\frac{\text{samples}}{\text{second}}$	9.000mm <sup>2</sup>
[76]	180nm	Μικτή	SVM	64	N/A	1M $\frac{\text{vectors}}{\text{second}}$	0.125mm <sup>2</sup>
[77]	180nm	Αναλογική	SVM	2	220.0μW	870K $\frac{\text{vectors}}{\text{second}}$	0.060mm <sup>2</sup>
[80]	0.5μm	Αναλογική	RBF NN	2	N/A	20K-40K $\frac{\text{classifications}}{\text{second}}$	2.250mm <sup>2</sup>
[78]	180nm	Αναλογική	SVDD	2	-	26.7M $\frac{\text{vectors}}{\text{second}}$	N/A
[81]	130nm	Μικτή	RBF NN	1280X720 pixels	2.2mW	N/A	0.140mm <sup>2</sup>
[82]	130nm	Μικτή	Object Recognition επεξεργαστής	640X480 pixels	496.0mW	N/A	49.000mm <sup>2</sup>
[83]	130nm	Μικτή	Neuro Fuzzy επεξεργαστής	N/A	57.0mW	N/A	13.500mm <sup>2</sup>
[84]	180nm	Μικτή	LSTM	16X16 matrix	460.3mW	N/A	9.990mm <sup>2</sup>
[79]	0.5μm	Αναλογική	SVM	14	840.0nW	40 $\frac{\text{classifications}}{\text{second}}$	9.000mm <sup>2</sup>
[85]	180nm	Αναλογική	K-means	164	N/A	10M $\frac{\text{vectors}}{\text{second}}$	N/A





## Κεφάλαιο 5

# Συμπεράσματα και Μελλοντική Δουλειά

Σε αυτή την εργασία, παρουσιάστηκε μια πλήρως αναλογική, εξαιρετικά χαμηλής κατανάλωσης ισχύος, μικρής επιφάνειας υλοποίηση ενός ταξινομητή βασισμένου στο GMM. Η αρχιτεκτονική αναλύθηκε σε επίπεδο συστήματος και διάφορες παραλλαγές στα βασικά δομικά στοιχεία της εξηγήθηκαν λεπτομερώς σε επίπεδο τρανζίστορ. Τα στοιχεία αυτά περιλαμβάνουν 4 διαφορετικές αρχιτεκτονικές Bump κυκλωμάτων και 2 αρχιτεκτονικές WTA κυκλωμάτων. Η προτεινόμενη αρχιτεκτονική εκτελεί αυτόνομα την ταξινόμηση με εντελώς αναλογικό τρόπο, εφόσον προηγουμένως έχει γίνει η εκπαίδευση του μοντέλου σε ένα εξωτερικό σύστημα. Τα πειραματικά αποτελέσματα αποδεικνύουν τη σωστή λειτουργία του σχεδιασμένου ταξινομητή και ως προς την ακρίβεια της ταξινόμησης (συγκριτικά με μια παραδοσιακή υλοποίηση λογισμικού) και ως προς την ευαισθησία του κυκλώματος.

Μελλοντικές εργασίες που σχετίζονται με αυτήν την αρχιτεκτονική θα μπορούσαν να περιλαμβάνουν την κατασκευή και την ρύθμιση ενός πραγματικού chip. Για να γίνει αυτό χρειάζεται ακόμα να σχεδιαστούν αναλογικά και χαμηλής ισχύος κυκλώματα για τη διασύνδεση μνήμης με την προτεινόμενη αρχιτεκτονική, με στόχο την αποθήκευση τιμών παραμέτρων. Επιπλέον, τα Bump κυκλώματα θα μπορούσαν να βελτιωθούν ώστε να λειτουργούν σε διανύσματα εισόδου υψηλότερων διαστάσεων από αυτών της εργασίας. Εναλλακτικά, εκτός από την υλοποίηση του GMM, τα βασικά δομικά στοιχεία αυτής της αρχιτεκτονικής θα μπορούσαν να τροποποιηθούν ώστε να χρησιμοποιηθούν σε άλλες υλοποιήσεις υλικού μηχανικής εκμάθησης, όπως SVM, K-means και άλλα.



## Δημοσιεύσεις

- Alimisis, V., Gourdouparis, M., Gennis, G., Dimas, C., & Sotiriadis, P. P. (2021). Analog Gaussian Function Circuit: Architectures, Operating Principles and Applications. *Electronics*, 10(20), 2530.
- Alimisis, V., Gennis, G., Dimas, C., & Sotiriadis, P. P. (2021, December). An Analog Bayesian Classifier Implementation, for Thyroid Disease Detection, based on a Low-Power, Current-Mode Gaussian Function Circuit. In 2021 33rd International Conference on Microelectronics (ICM) IEEE.
- Alimisis, V., Gennis, G., Dimas, C., & Sotiriadis, P. P. (2022). An Ultra Low Power Analog Integrated Radial Basis Function Classifier for Smart IoT Systems. *Analog Integrated Circuits and Signal Processing*. (under review)



# Bibliography

- [1] B. Panić, J. Klemenc, and M. Nagode, “Gaussian mixture model based classification revisited: Application to the bearing fault classification.,” *Strojnikski Vestnik/Journal of Mechanical Engineering*, vol. 66, no. 4, 2020.
- [2] A. Ishtiaq, M. U. Khan, S. Ali, K. Habib, S. Samer, and E. Hafeez, “A review of system on chip (soc) applications in internet of things (iot) and medical,” in *ICAME21, International Conference on Advances in Mechanical Engineering, Pakistan*, pp. 1–10, 2021.
- [3] A. Shehab, A. Ismail, L. Osman, M. Elhoseny, and I. M. El-Henawy, “Quantified self using iot wearable devices,” in *International conference on advanced intelligent systems and informatics*, pp. 820–831, Springer, 2017.
- [4] H. Jayakumar, A. Raha, Y. Kim, S. Sutar, W. S. Lee, and V. Raghunathan, “Energy-efficient system design for iot devices,” in *2016 21st Asia and South Pacific design automation conference (ASP-DAC)*, pp. 298–301, IEEE, 2016.
- [5] M. S. Aslanpour, A. N. Toosi, C. Cicconetti, B. Javadi, P. Sbarski, D. Taibi, M. Assuncao, S. S. Gill, R. Gaire, and S. Dustdar, “Serverless edge computing: vision and challenges,” in *2021 Australasian Computer Science Week Multiconference*, pp. 1–10, 2021.
- [6] M. Capra, B. Bussolino, A. Marchisio, M. Shafique, G. Masera, and M. Martina, “An updated survey of efficient hardware architectures for accelerating deep convolutional neural networks,” *Future Internet*, vol. 12, no. 7, p. 113, 2020.
- [7] A. X. M. Chang and E. Culurciello, “Hardware accelerators for recurrent neural networks on fpga,” in *2017 IEEE International symposium on circuits and systems (ISCAS)*, pp. 1–4, IEEE, 2017.

- [8] A. Sabne, P. Sakdhnagool, and R. Eigenmann, “Scaling large-data computations on multi-gpu accelerators,” in *Proceedings of the 27th international ACM conference on International conference on supercomputing*, pp. 443–454, 2013.
- [9] K. Iwai, N. Nishikawa, and T. Kurokawa, “Acceleration of aes encryption on cuda gpu,” *International Journal of Networking and Computing*, vol. 2, no. 1, pp. 131–145, 2012.
- [10] U. Rueckert, “Digital neural network accelerators,” in *NANO-CHIPS 2030*, pp. 181–202, Springer, 2020.
- [11] P. Mohan, W. Wang, B. Jungk, R. Niederhagen, J. Szefer, and K. Mai, “Asic accelerator in 28 nm for the post-quantum digital signature scheme xmss,” in *2020 IEEE 38th International Conference on Computer Design (ICCD)*, pp. 656–662, IEEE, 2020.
- [12] M. A. Talib, S. Majzoub, Q. Nasir, and D. Jamal, “A systematic literature review on hardware implementation of artificial intelligence algorithms,” *The Journal of Supercomputing*, vol. 77, pp. 1897–1938, 2021.
- [13] J. Lu, S. Young, I. Arel, and J. Holleman, “A 1 tops/w analog deep machine-learning engine with floating-gate storage in 0.13  $\mu\text{m}$  cmos,” *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 270–281, 2014.
- [14] S. Moon, K. Shin, and D. Jeon, “Enhancing reliability of analog neural network processors,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 6, pp. 1455–1459, 2019.
- [15] A. Wang, B. H. Calhoun, and A. P. Chandrakasan, *Sub-threshold design for ultra low-power systems*, vol. 95. Springer, 2006.
- [16] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbra, R. Douglas, *et al.*, *Analog VLSI: circuits and principles*. MIT press, 2002.
- [17] B. Gilbert, “Translinear circuits: A proposed classification,” *Electronics letters*, vol. 11, no. 1, pp. 14–16, 1975.
- [18] J. Mulder, W. A. Serdijn, A. C. van der Woerd, and A. van Roermund, *Dynamic translinear and log-domain circuits: analysis and synthesis*. Springer Science & Business Media, 1998.
- [19] T. Delbrueck and C. Mead, “Bump circuits,” in *Proceedings of International Joint Conference on Neural Networks*, vol. 1, pp. 475–479, 1993.

- [20] C. M. Bishop, “Pattern recognition,” *Machine learning*, vol. 128, no. 9, 2006.
- [21] S. Haykin, *Neural networks and learning machines, 3/E*. Pearson Education India, 2010.
- [22] M. Alber, A. B. Tepole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W. W. Lytton, P. Perdikaris, L. Petzold, *et al.*, “Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences,” *NPJ digital medicine*, vol. 2, no. 1, pp. 1–11, 2019.
- [23] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, “Speech recognition using deep neural networks: A systematic review,” *IEEE access*, vol. 7, pp. 19143–19165, 2019.
- [24] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, “Deep learning-based image recognition for autonomous driving,” *IATSS research*, vol. 43, no. 4, pp. 244–252, 2019.
- [25] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, “Forecasting stock market crisis events using deep and statistical machine learning techniques,” *Expert systems with applications*, vol. 112, pp. 353–371, 2018.
- [26] E. Strubell, A. Ganesh, and A. McCallum, “Energy and policy considerations for deep learning in nlp,” *arXiv preprint arXiv:1906.02243*, 2019.
- [27] P. Chi, S. Li, C. Xu, T. Zhang, J. Zhao, Y. Liu, Y. Wang, and Y. Xie, “Prime: A novel processing-in-memory architecture for neural network computation in rram-based main memory,” *ACM SIGARCH Computer Architecture News*, vol. 44, no. 3, pp. 27–39, 2016.
- [28] I. Akita, T. Okazawa, Y. Kurui, A. Fujimoto, and T. Asano, “A feed-forward noise reduction technique in capacitive mems accelerometer analog front-end for ultra-low-power iot applications,” *IEEE Journal of Solid-State Circuits*, vol. 55, no. 6, pp. 1599–1609, 2019.
- [29] E. Farella, M. Rusci, B. Milosevic, and A. L. Murphy, “Technologies for a thing-centric internet of things,” in *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 77–84, IEEE, 2017.

- [30] G. J. McLachlan and K. E. Basford, *Mixture models: Inference and applications to clustering*, vol. 38. M. Dekker New York, 1988.
- [31] J. Geweke and G. Amisano, “Hierarchical markov normal mixture models with applications to financial asset returns,” *Journal of Applied Econometrics*, vol. 26, no. 1, pp. 1–29, 2011.
- [32] J. Friedman and R. Posner, *What caused the financial crisis*. University of Pennsylvania Press, 2011.
- [33] X. Chen, X. Liu, and Y. Jia, “Discriminative structure selection method of gaussian mixture models with its application to handwritten digit recognition,” *Neurocomputing*, vol. 74, no. 6, pp. 954–961, 2011.
- [34] J. Richiardi and A. Drygajlo, “Gaussian mixture models for on-line signature verification,” in *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications*, pp. 115–122, 2003.
- [35] R. Dhineshkumar, A. B. Ganesh, and S. Sasikala, “Speaker identification system using gaussian mixture model and support vector machines (gmm-svm) under noisy conditions,” *Indian Journal of Science and Technology*, vol. 9, no. 19, pp. 1–6, 2016.
- [36] M. Genovese and E. Napoli, “Fpga-based architecture for real time segmentation and denoising of hd video,” *Journal of Real-Time Image Processing*, vol. 8, no. 4, pp. 389–401, 2013.
- [37] Z. Ji, Y. Xia, Q. Sun, Q. Chen, D. Xia, and D. D. Feng, “Fuzzy local gaussian mixture model for brain mr image segmentation,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 339–347, 2012.
- [38] V. Alimisis, M. Gourdouparis, G. Gennis, C. Dimas, and P. P. Sotiriadis, “Analog gaussian function circuit: Architectures, operating principles and applications,” *Electronics*, vol. 10, no. 20, p. 2530, 2021.
- [39] T. Delbrück, “bump,” in *Circuits for computing similarity and dissimilarity of analog voltages. In Proceedings of the International Joint Conference on Neural Networks*, 1991.
- [40] R. He and L. Zhang, “Evaluation of modern mosfet models for bulk-driven applications,” in *2008 51st Midwest Symposium on Circuits and Systems*, pp. 105–108, IEEE, 2008.



- [41] F. Khateb, D. Biolek, N. Khatib, and J. Vávra, "Utilizing the bulk-driven technique in analog circuit design," in *13th IEEE Symposium on Design and Diagnostics of Electronic Circuits and Systems*, pp. 16–19, IEEE, 2010.
- [42] P. Pavan, L. Larcher, and A. Marmiroli, *Floating gate devices: operation and compact modeling*. Springer Science & Business Media, 2007.
- [43] M. Burns, G. W. Roberts, *et al.*, *An introduction to mixed-signal IC test and measurement*, vol. 2001. IET, 2001.
- [44] V. Alimisis, M. Gourdouparis, C. Dimas, and P. P. Sotiriadis, "A 0.6 v, 3.3 nw, adjustable gaussian circuit for tunable kernel functions," in *2021 34th SBC/SBMicro/IEEE/ACM Symposium on Integrated Circuits and Systems Design (SBCCI)*, pp. 1–6, IEEE, 2021.
- [45] M. Gourdouparis, V. Alimisis, C. Dimas, and P. P. Sotiriadis, "An ultra-low power,  $\pm 0.3$  v supply, fully-tunable gaussian function circuit architecture for radial-basis functions analog hardware implementation," *AEU-International Journal of Electronics and Communications*, vol. 136, p. 153755, 2021.
- [46] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all networks of  $O(n)$  complexity," 1988.
- [47] V. Alimisis, M. Gourdouparis, C. Dimas, and P. P. Sotiriadis, "Ultra-low power, low-voltage, fully-tunable, bulk-controlled bump circuit," in *2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pp. 1–4, IEEE, 2021.
- [48] J. Choi and B. J. Sheu, "A high-precision vlsi winner-take-all circuit for self-organizing neural networks," *IEEE Journal of Solid-state circuits*, vol. 28, no. 5, pp. 576–584, 1993.
- [49] R. Sridharan, "Gaussian mixture models and the em algorithm," *Available in: <http://people.csail.mit.edu/rameshvs/content/gmm-em.pdf>*, 2014.
- [50] D. A. Reynolds, "Gaussian mixture models.," *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.
- [51] D. Yu and L. Deng, "Gaussian mixture models," in *Automatic Speech Recognition*, pp. 13–21, Springer, 2015.

- [52] M. Hock, A. Hartel, J. Schemmel, and K. Meier, "An analog dynamic memory array for neuromorphic hardware," in *2013 European Conference on Circuit Theory and Design (ECCTD)*, pp. 1–4, IEEE, 2013.
- [53] A. K. Sharma, M. Madhusudan, S. M. Burns, P. Mukherjee, S. Yaldiz, R. Harjani, and S. S. Sapatnekar, "Common-centroid layouts for analog circuits: Advantages and limitations," *Proc. DATE. IEEE, Piscataway, NJ*, 2021.
- [54] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, 2005.
- [55] P. R. Kinget, "Designing analog and rf circuits for ultra-low supply voltages," in *ESSDERC 2007-37th European Solid State Device Research Conference*, pp. 58–67, IEEE, 2007.
- [56] M. P. Vanderpump, "The epidemiology of thyroid disease.," *British medical bulletin*, vol. 99, no. 1, 2011.
- [57] D. J. Newman, "Uci repository of machine learning database," <http://www.ics.uci.edu/~mllearn/MLRepository.html>, 1998.
- [58] A. Jabbar, A. Pingitore, S. H. Pearce, A. Zaman, G. Iervasi, and S. Razvi, "Thyroid hormones and cardiovascular disease," *Nature Reviews Cardiology*, vol. 14, no. 1, pp. 39–55, 2017.
- [59] W. H. Organization, G. C. against Epilepsy, P. for Neurological Diseases, N. W. H. Organization), I. B. for Epilepsy, W. H. O. D. of Mental Health, S. Abuse, I. B. of Epilepsy, and I. L. against Epilepsy, *Atlas: epilepsy care in the world*. World Health Organization, 2005.
- [60] P. J. Karoly, V. R. Rao, N. M. Gregg, G. A. Worrell, C. Bernard, M. J. Cook, and M. O. Baud, "Cycles in epilepsy," *Nature Reviews Neurology*, vol. 17, no. 5, pp. 267–284, 2021.
- [61] J. R. Hughes, "A review of sudden unexpected death in epilepsy: prediction of patients at risk," *Epilepsy & Behavior*, vol. 14, no. 2, pp. 280–287, 2009.
- [62] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals," *Computers in biology and medicine*, vol. 99, pp. 24–37, 2018.

- [63] S.-K. Lin, L.-C. Wang, C.-Y. Lin, H. Chiueh, *et al.*, “An ultra-low power smart headband for real-time epileptic seizure detection,” *IEEE journal of translational engineering in health and medicine*, vol. 6, pp. 1–10, 2018.
- [64] A. M. Abdelhameed and M. Bayoumi, “An efficient deep learning system for epileptic seizure prediction,” in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, IEEE, 2021.
- [65] S. Iranmanesh, G. Raikos, S. A. Imtiaz, and E. Rodriguez-Villegas, “A seizure-based power reduction soc for wearable eeg in epilepsy,” *IEEE Access*, vol. 7, pp. 151682–151691, 2019.
- [66] S. Iranmanesh and E. Rodriguez-Villegas, “A 950 nw analog-based data reduction chip for wearable eeg systems in epilepsy,” *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2362–2373, 2017.
- [67] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals,” *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [68] R. Miller, “Theory of the normal waking eeg: from single neurones to waveforms in the alpha, beta and gamma frequency ranges,” *International journal of psychophysiology*, vol. 64, no. 1, pp. 18–23, 2007.
- [69] M. Chen, O. Boric-Lubecke, and V. M. Lubecke, “0.5- $\mu\text{m}$  cmos implementation of analog heart-rate extraction with a robust peak detector,” *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 4, pp. 690–698, 2008.
- [70] Y. Zhang, N. Mirchandani, M. Onabajo, and A. Shrivastava, “Rssi amplifier design for a feature extraction technique to detect seizures with analog computing,” in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, IEEE, 2020.
- [71] D. G. Altman and J. M. Bland, “Diagnostic tests. 1: Sensitivity and specificity,” *BMJ: British Medical Journal*, vol. 308, no. 6943, p. 1552, 1994.
- [72] K. A. Loparo, “Bearing data center,” *Case Western Reserve University*, vol. 338, 2013.

- [73] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data in brief*, vol. 21, pp. 1745–1749, 2018.
- [74]
- [75] R. Genov and G. Cauwenberghs, "Kerneltron: support vector machine" in silicon," *IEEE Transactions on Neural Networks*, vol. 14, no. 5, pp. 1426–1434, 2003.
- [76] R. Zhang and T. Shibata, "Fully parallel self-learning analog support vector machine employing compact gaussian generation circuits," *Japanese Journal of Applied Physics*, vol. 51, no. 4S, p. 04DE10, 2012.
- [77] K. Kang and T. Shibata, "An on-chip-trainable gaussian-kernel analog support vector machine," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1513–1524, 2009.
- [78] R. Zhang and T. Shibata, "A vlsi hardware implementation study of svdd algorithm using analog gaussian-cell array for on-chip learning," in *2012 13th International Workshop on Cellular Nanoscale Networks and their Applications*, pp. 1–6, IEEE, 2012.
- [79] S. Chakrabartty and G. Cauwenberghs, "Sub-microwatt analog vlsi trainable pattern classifier," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 1169–1179, 2007.
- [80] S.-Y. Peng, P. E. Hasler, and D. V. Anderson, "An analog programmable multidimensional radial basis function based classifier," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 10, pp. 2148–2158, 2007.
- [81] K. Lee, J. Park, and H.-J. Yoo, "A low-power, mixed-mode neural network classifier for robust scene classification," *Journal of Semiconductor Technology and Science*, vol. 19, no. 1, pp. 129–136, 2019.
- [82] J.-Y. Kim, M. Kim, S. Lee, J. Oh, K. Kim, and H.-J. Yoo, "A 201.4 gops 496 mw real-time multi-object recognition processor with bio-inspired neural perception engine," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 1, pp. 32–45, 2009.
- [83] J. Oh, G. Kim, B.-G. Nam, and H.-J. Yoo, "A 57 mw 12.5  $\mu$ j/epoch embedded mixed-mode neuro-fuzzy processor for mobile real-time ob-

- ject recognition,” *IEEE journal of solid-state circuits*, vol. 48, no. 11, pp. 2894–2907, 2013.
- [84] Z. Zhao, A. Srivastava, L. Peng, and Q. Chen, “Long short-term memory network design for analog computing,” *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 15, no. 1, pp. 1–27, 2019.
- [85] R. Zhang and T. Shibata, “An analog on-line-learning k-means processor employing fully parallel self-converging circuitry,” *Analog Integrated Circuits and Signal Processing*, vol. 75, no. 2, pp. 267–277, 2013.