



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΠΜΣ ΜΙΚΡΟΣΥΣΤΗΜΑΤΑ ΚΑΙ ΝΑΝΟΔΙΑΤΑΞΕΙΣ

ΠΡΟΣΟΜΟΙΩΣΗ ΝΕΥΡΟΜΟΡΦΙΚΟΥ ΔΙΚΤΥΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΔΙΚΤΥΩΜΑΤΟΣ ΑΠΟ ΜΝΗΜΕΣ ΜΕΤΑΒΛΗΤΗΣ ΑΝΤΙΣΤΑΣΗΣ

ΛΥΜΠΕΡΟΠΟΥΛΟΣ ΣΤΑΥΡΟΣ

ΑΚΑΔΗΜΑΪΚΟΣ ΕΠΙΒΛΕΠΩΝ:

ΔΡ. ΔΗΜΗΤΡΙΟΣ ΤΣΟΥΚΑΛΑΣ

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2021

Σελίδα σκόπμα κενή

ΕΥΧΑΡΙΣΤΙΕΣ

Πρωτίστως, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Δρ. Δ. Τσουκαλά που με καθοδήγησε και με ευγένεια μου έδωσε αυτή την υπέροχη ευκαιρία να ασχοληθώ με το αντικείμενο της παρούσας πτυχιακής. Είμαι ιδιαίτερα ευγνώμων για τη καθοδήγηση, τη στήριξη και τις γόνιμες συζητήσεις με τον μεταδιδακτορικό ερευνητή Δρ. Π. Μπούσουλα, τον υποψήφιο διδάκτορα Σ. Κίτσιο και τον συμφοιτητή μου Χ. Τσιούστα.

Δεν θα μπορούσα να μην αναφερθώ και στη στήριξη και το ενδιαφέρον του καθηγητή μου και διευθυντή του μεταπτυχιακού προγράμματος, Δρ. Ι. Ράπτη, που είναι ένας από τους λόγους που ολοκληρώνω το πτυχίο αυτό.

Σας ευχαριστώ όλους ολόψυχα για όλα!

ΠΕΡΙΛΗΨΗ

Ζούμε σε μια εποχή που χαρακτηρίζεται από την εκρηκτική δημιουργία τεράστιων όγκων δεδομένων και τη προσπάθεια επεξεργασίας τους. Αυτό το γεγονός έχει τρομακτική επίπτωση σε όλες τις πτυχές της σύγχρονης ζωής με τις εξελίξεις στη διαχείριση και επεξεργασία των δεδομένων και τη χρήση της τεχνητής νοημοσύνης. Η παρούσα πτυχιακή εργασία αναλύει το πολλά υποσχόμενο δομικό στοιχείο των τεχνητών νευρωνικών δικτύων στο επίπεδο του υλισμικού (hardware), το οποίο είναι η μνημαντίσταση (memristor). Η μνημαντίσταση, θεμελιωμένη θεωρητικά το 1971 από τον Leon Chua και υλοποιημένη σε εργαστήριο το 2008 από τον Dmitri Strukov, έχει έντονο ερευνητικό ενδιαφέρον επειδή τα χαρακτηριστικά της είναι όμοια με τις βιολογικές συνάψεις. Τα πρώτα δυο κεφάλαια της πτυχιακής εργασίας χτίζουν το θεωρητικό υπόβαθρο στα νευρομορφικά συστήματα ενώ το τρίτο κεφάλαιο προχωρά σε μια προσομοίωση ενός δικτύωματος μνημών μεταβλητής αντίστασης (memristor crossbar) σε λειτουργία τεχνητού νευρωνικού δικτύου.

Συγκεκριμένα, το πρώτο κεφάλαιο περιγράφει τη θεωρία στην οποία βασίζεται η μνήμη μεταβλητής αντίστασης και τις διάφορες αρχές λειτουργίας της με βάση το υλικό κατασκευής της. Ακολουθεί μια ανάλυση των πιο διαδεδομένων γραμμικών και μη-γραμμικών μοντέλων που περιγράφουν τη λειτουργία των μνημών μεταβλητής αντίστασης και η σύγκρισή τους.

Το δεύτερο κεφάλαιο επικεντρώνεται στα τεχνητά νευρωνικά δίκτυα και ξεκινάει με ανάλυση των συνάψεων στη βιολογία και της μνήμης μεταβλητής αντίστασης, που είναι η αντίστοιχη σύναψη στην ηλεκτρονική. Στη συνέχεια συναπτικοί τύποι μονής και διπλής μνήμης μεταβλητής αντίστασης, σύνθετες διατάξεις καθώς και μνήμη μεταβλητής αντίστασης με τη χρήση επιλογέα αναλύονται στα ηλεκτρονικά κυκλώματα. Το κεφάλαιο καταλήγει με τη περιγραφή τεχνητών νευρωνικών δικτύων, τις αρχές λειτουργίας τους, τα χαρακτηριστικά τους και ο τρόπος υλοποίησής τους στα ηλεκτρονικά κυκλώματα με τη χρήση δικτυωμάτων.

Το τρίτο και τελευταίο κεφάλαιο συναρμολογεί όλα τα παραπάνω σε μια προσομοίωση δικτύωματος μνημών μεταβλητής αντίστασης. Η προσομοίωση γίνεται για ένα τεχνητό νευρωνικό δίκτυο ενός επιπέδου που κατηγοριοποιεί σε τέσσερις κατηγορίες ασπρόμαυρες εικόνες διαστάσεων 4x4 εικονοστοιχείων (pixel). Το τεχνητό νευρωνικό δίκτυο υλοποιήθηκε σε ένα δίκτυωμα διαστάσεων 5x4 μνημών μεταβλητής αντίστασης με τη χρήση μιας μνήμης μεταβλητής αντίστασης ως σύναψη. Ο αλγόριθμος εκπαίδευσης που χρησιμοποιήθηκε ήταν της οπισθοδιάδοσης (backpropagation) και σε κάθε κύκλο εκπαίδευσης (epoch) χρησιμοποιήθηκε μια εικόνα εισόδου. Η προσομοίωση έδειξε ότι το δίκτυωμα μνημών μεταβλητής αντίστασης ήταν ικανό να εκπαιδευτεί και να κατηγοριοποιήσει σωστά τις εικόνες μετά από δεκαέξι κύκλους εκπαίδευσης και να επιδείξει 100% ακρίβεια στα δεδομένα ελέγχου. Το παραπάνω αποδεικνύει την ικανότητα των δικτυωμάτων μνημών μεταβλητής μνήμης να λειτουργήσουν ως τεχνητά νευρωνικά δίκτυα.

ABSTRACT

We are living in times of explosive growth of data creation and processing. This fact has an immense impact on all aspects of modern living through the advancements of data manipulation and the use of artificial intelligence. This dissertation analyzes a promising building block of artificial neural networks at the hardware level, the memristor. The memristor, first conceptually founded in 1971 by Leon Chua and realized in a lab in 2008 by Dmitri Strukov, has intense research due to its characteristics that are very similar to biological synapses. The first two chapters build a theoretical foundation on neuromorphic computing while the third and final chapter simulates a memristor crossbar functioning as an artificial neural network.

Specifically, the first chapter describes the theory behind the memristor and its physical functioning mechanisms based on the materials used to develop it. Following there is an analysis of the most widely used memristor models, both linear and non-linear and a comparison between them.

The second chapter emphasizes on artificial neural networks and starts off with an analysis on synapses in biology and their equivalent in electronics that is the memristor. Synaptic types like single memristor, double memristors, complex memristor configurations and memristors with selectors in electronic systems are described. The chapter continues with descriptions of artificial neural networks, their functioning logic and characteristics and implementation in hardware by memristor crossbars.

The third and final chapter puts all the above in action through a memristor crossbar simulation. The simulation is for a single layer artificial neural network that classifies 4x4 pixel black and white images into 4 categories. The artificial neural network is implemented as a 5x4 memristor crossbar using as synapses a single memristor. The training algorithm used was backpropagation on a single sample of input images. It was shown that the memristor crossbar was able to be trained and capable of correctly classifying images after 16 epochs and achieved 100% classification accuracy on the testing dataset. The above proves the capability of memristor crossbars to function as artificial neural networks.

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1: Η μνήμη μεταβλητής αντίστασης (μνημαντίσταση).....	10
1.1 Εισαγωγή.....	10
1.2 Η θεωρία της μνήμης μεταβλητής αντίστασης	10
1.3 Μηχανισμοί των μνημών μεταβλητής αντίστασης	12
1.4 Μοντέλα των μνημών μεταβλητής αντίστασης	16
1.4.1 Γραμμικά & μη-γραμμικά μοντέλα.....	16
1.4.2 Εξελιγμένα μη-γραμμικά μοντέλα	17
1.4.3 Σχέσεις/συναρτήσεις μοντέλων.....	18
1.5 Αναφορές 1 ^{ου} κεφαλαίου	20
Κεφάλαιο 2: Τεχνητή νοημοσύνη και οι μνήμες μεταβλητής αντίστασης.....	22
2.1 Συνάψεις στη βιολογία και την ηλεκτρονική.....	22
2.2 Τεχνητά νευρωνικά δίκτυα (ΤΝΔ)	25
2.3 Δικτυώματα μνημών μεταβλητής αντίστασης	27
2.4 Αναφορές 2 ^{ου} κεφαλαίου	30
Κεφάλαιο 3: Προσομοίωση νευρωνικού δικτύου μέσω δικτυώματος μνημών μεταβλητής αντίστασης. 32	
3.1 Γενική περιγραφή του προβλήματος και δομικά στοιχεία της προσομοίωσης.....	32
3.2 Το LTSprice	32
3.3 Χρησιμοποιούμενο μοντέλο μνήμης μεταβλητής αντίστασης και τα χαρακτηριστικά του	34
3.4 Ανάλυση του συνόλου δεδομένων (dataset)	41
3.5 Αρχιτεκτονική του δικτυώματος μνημών μεταβλητής αντίστασης.....	46
3.6 Επαναληπτική μέθοδος εκπαίδευσης του δικτυώματος μνημών μεταβλητής αντίστασης	48
3.7 Αποτελέσματα της προσομοίωσης.....	52
3.7 Δεύτερο μοντέλο μνήμης μεταβλητής αντίστασης για προσομοίωση δικτυώματος.....	58
3.8 Αναφορές 3 ^{ου} κεφαλαίου	62
Κεφάλαιο 4: Συμπεράσματα και μελλοντική έρευνα.....	63
Παράρτημα 1 ^ο : Sprice κώδικας της μνήμης μεταβλητής αντίστασης	65
Παράρτημα 2 ^ο : Ρεύματα εξόδου νευρωνικού δικτυώματος ανά κύκλο εκπαίδευσης.....	66

ΣΧΗΜΑΤΑ

Σχήμα 1: Τα βασικά ηλεκτρονικά στοιχεία και η μνήμη μεταβλητής αντίστασης του L. O. Chua [5].....	11
Σχήμα 2: Βρόγχος υπερτονισμένης υστέρησης μνήμη μεταβλητής αντίστασης [5].....	12
Σχήμα 3: Μονοπολικοί και διπολικοί μηχανισμοί εναλλαγής αντίστασης [7].....	13
Σχήμα 4: Ταξινόμηση μνημών μεταβλητής αντίστασης με βάση την αρχή λειτουργίας τους [8].....	14
Σχήμα 5: Μηχανισμός μεταγωγής σε απλή MIM δομή (single layer) [13].....	15
Σχήμα 6: Νευρώνες και συνάψεις στους βιολογικούς οργανισμούς [1].....	22
Σχήμα 7: Η μνήμη μεταβλητής αντίστασης ως σύναψη [4].....	23
Σχήμα 8: Είδη συνάψεων - (Α) 1M , (Β) 2M [6], (Γ) 2M1R [7], (Δ) 1T1M [8], (Ε) 2T1M [9], (ΣΤ) 4M [10], 5M [11].....	23
Σχήμα 9: Απομόνωση στηλών δικτύωμα μνημών μεταβλητής αντίστασης με συνάψεις 1T1M [12].....	24
Σχήμα 10: Αναπαράσταση ΤΝΔ.....	25
Σχήμα 11: Μικρό δείγμα από τις διάφορες αρχιτεκτονικές ΤΝΔ [13].....	27
Σχήμα 12: Δικτύωμα με μνήμες μεταβλητής αντίστασης σε χρήση μνήμης με ρεύμα ερπυσμού [14]....	28
Σχήμα 13: I-V χαρακτηριστικές - συμπεριφορές για αντιμετώπιση ρευμάτων ερπυσμού [14].....	29
Σχήμα 14: Δομές 1T1M και 1S1M ως επιλογείς για την αντιμετώπιση ρευμάτων ερπυσμού [14].....	29
Σχήμα 15: LTSpice δικτύωμα μνημών μεταβλητής αντίστασης και τα μέρη του.....	33
Σχήμα 16: Ισοδύναμα ηλεκτρονικά κυκλώματα για το Yakopcic μοντέλο [1].....	36
Σχήμα 17: LTSpice κύκλωμα για χαρακτηριστική I-V της μνήμης μεταβλητής αντίστασης.....	38
Σχήμα 18: : Τάση εισόδου της μνήμης μεταβλητής αντίστασης για παραγωγή I-V.....	38
Σχήμα 19: Χαρακτηριστική I-V της μνήμης μεταβλητής αντίστασης της προσομοίωσης ($\chi_0=0.001$).....	39
Σχήμα 21: Μέτρηση αγωγιμότητας για $\chi_0=0.5$ (περιοχή στο κόκκινο πλαίσιο).....	40
Σχήμα 22: Δυαδική κωδικοποίηση των εικόνων εισόδου [3].....	41
Σχήμα 23: Παράδειγμα δυαδικής κωδικοποίησης τάσεων εισόδου.....	42
Σχήμα 24: Σύνολο δεδομένων εκπαίδευσης νευρωνικού δικτύου.....	43
Σχήμα 25: Σήματα εισόδου μέσα στο LTSpice.....	44
Σχήμα 26: Δεδομένα ελέγχου του νευρωνικού δικτύου.....	45
Σχήμα 27: Δεδομένα ελέγχου μέσα από το LTSpice.....	46
Σχήμα 28: : Το LTSpice δικτύωμα μνημών μεταβλητής αντίστασης της προσομοίωσης.....	47
Σχήμα 29: Τυπική αναπαράσταση του ΤΝΔ της προσομοίωσης.....	48
Σχήμα 30: Εικόνες εισόδου και έξοδοι νευρώνων [3].....	50
Σχήμα 31: Διαδικασία προσομοίωσης και εκπαίδευσης του δικτύωματος μνημών μεταβλητής αντίστασης.....	52
Σχήμα 32: Αρχική έξοδος του δικτύωματος με αρχικές μεταβλητές καταστάσεων.....	52
Σχήμα 33: Έξοδος του δικτύωματος μετά από 8 κύκλους εκπαίδευσης.....	53
Σχήμα 34: Έξοδος του δικτύωματος μετά από 16 κύκλους εκπαίδευσης.....	53
Σχήμα 35: Έξοδος του δικτύωματος μετά από 24 κύκλους εκπαίδευσης.....	54
Σχήμα 36: Απόλυτη ακρίβεια κατηγοριοποίησης ανά κύκλο εκπαίδευσης.....	55
Σχήμα 37: Μεταβολές αγωγιμότητας ανά μνήμη μεταβλητής αντίστασης (U1 - U8) ανά κύκλο εκπαίδευσης.....	56
Σχήμα 38: Μεταβολές αγωγιμότητας ανά μνήμη μεταβλητής αντίστασης (U9 - U16) ανά κύκλο εκπαίδευσης.....	57
Σχήμα 39: Μεταβολές αγωγιμότητας ανά μνήμη μεταβλητής αντίστασης (U17 - U20) ανά κύκλο εκπαίδευσης.....	58

Σχήμα 40: Τάση εισόδου 2 ^{ης} μνήμης μεταβλητής αντίστασης.....	59
Σχήμα 41: I-V γράφημα 2 ^{ης} μνήμης μεταβλητής αντίστασης ($\chi_0=0.01$).....	60
Σχήμα 42: Σύγκριση αγωγιμότητας στην κατάσταση υψηλής αντίστασης των δυο μνημών μεταβλητής αντίστασης.....	61
Σχήμα 43: Κώδικας Spice για το μοντέλο Yakorcic	65
Σχήμα 44: Αρχική έξοδος του δικτυώματος με αρχικές μεταβλητές καταστάσεων	66
Σχήμα 45: Έξοδος του δικτυώματος μετά από 1 κύκλο εκπαίδευσης	66
Σχήμα 46: Έξοδος του δικτυώματος μετά από 2 κύκλους εκπαίδευσης.....	67
Σχήμα 47: Έξοδος του δικτυώματος μετά από 3 κύκλους εκπαίδευσης.....	67
Σχήμα 48: Έξοδος του δικτυώματος μετά από 4 κύκλους εκπαίδευσης.....	68
Σχήμα 49: Έξοδος του δικτυώματος μετά από 5 κύκλους εκπαίδευσης.....	68
Σχήμα 50: Έξοδος του δικτυώματος μετά από 6 κύκλους εκπαίδευσης.....	69
Σχήμα 51: Έξοδος του δικτυώματος μετά από 7 κύκλους εκπαίδευσης.....	69
Σχήμα 52: Έξοδος του δικτυώματος μετά από 8 κύκλους εκπαίδευσης.....	70
Σχήμα 53: Έξοδος του δικτυώματος μετά από 9 κύκλους εκπαίδευσης.....	70
Σχήμα 54: Έξοδος του δικτυώματος μετά από 10 κύκλους εκπαίδευσης.....	71
Σχήμα 55: Έξοδος του δικτυώματος μετά από 11 κύκλους εκπαίδευσης.....	71
Σχήμα 56: Έξοδος του δικτυώματος μετά από 12 κύκλους εκπαίδευσης.....	72
Σχήμα 57: Έξοδος του δικτυώματος μετά από 13 κύκλους εκπαίδευσης.....	72
Σχήμα 58: Έξοδος του δικτυώματος μετά από 14 κύκλους εκπαίδευσης.....	73
Σχήμα 59: Έξοδος του δικτυώματος μετά από 15 κύκλους εκπαίδευσης.....	73
Σχήμα 60: Έξοδος του δικτυώματος μετά από 16 κύκλους εκπαίδευσης.....	74
Σχήμα 61: Έξοδος του δικτυώματος μετά από 17 κύκλους εκπαίδευσης.....	74
Σχήμα 62: Έξοδος του δικτυώματος μετά από 18 κύκλους εκπαίδευσης.....	75
Σχήμα 63: Έξοδος του δικτυώματος μετά από 19 κύκλους εκπαίδευσης.....	75
Σχήμα 64: Έξοδος του δικτυώματος μετά από 20 κύκλους εκπαίδευσης.....	76
Σχήμα 65: Έξοδος του δικτυώματος μετά από 21 κύκλους εκπαίδευσης.....	76
Σχήμα 66: Έξοδος του δικτυώματος μετά από 22 κύκλους εκπαίδευσης.....	77
Σχήμα 67: Έξοδος του δικτυώματος μετά από 23 κύκλους εκπαίδευσης.....	77
Σχήμα 68: Έξοδος του δικτυώματος μετά από 24 κύκλους εκπαίδευσης.....	78

ΠΙΝΑΚΕΣ

Πίνακας 1: Οι I-V σχέσεις των διάφορων μοντέλων μνημών μεταβλητής αντίστασης [23].....	18
Πίνακας 2: Οι μεταβλητές καταστάσεων του κάθε μοντέλου μνήμης μεταβλητής αντίστασης [23]	18
Πίνακας 3: Σύγκριση των μοντέλων μνημών μεταβλητής αντίστασης [21].....	19
Πίνακας 4: Σύγκριση των διάφορων συναρτήσεων παραθύρου [21].....	19
Πίνακας 5: Παράμετροι μοντέλου και η σχέση τους με τη φυσική συμπεριφορά των υλικών [1]	36
Πίνακας 6: Παράμετροι μοντέλου μνήμης μεταβλητής αντίστασης	37
Πίνακας 7: Αγωγιμότητα - μεταβλητή κατάσταση μνήμης μεταβλητής αντίστασης.....	40
Πίνακας 8: Γραμμή παλινδρόμησης για μετατροπή αγωγιμότητας σε μεταβλητή κατάσταση	41
Πίνακας 9: Γραμμή παλινδρόμησης για μετατροπή της μεταβλητής κατάστασης σε αγωγιμότητα	41
Πίνακας 10: Πίνακας δεδομένων εκπαίδευσης σε volt.....	43
Πίνακας 11: Πίνακας δεδομένων ελέγχου	45
Πίνακας 12: Αρχικές τιμές μεταβλητής κατάστασης των μνημών μεταβλητής αντίστασης.....	49
Πίνακας 13: Παράμετροι μοντέλου Yakorcić για τη 1η και 2η μνήμη μεταβλητής αντίστασης	59
Πίνακας 14: Συντελεστές μετατροπής αγωγιμότητας σε μεταβλητή κατάσταση.....	60
Πίνακας 15: Οι αγωγιμότητες των δυο μνημών μεταβλητής αντίστασης και η σχέση τους.....	61

Κεφάλαιο 1: Η μνήμη μεταβλητής αντίστασης (μνημαντίσταση)

1.1 Εισαγωγή

Τα τελευταία χρόνια ο χώρος της πληροφορικής και των τεχνολογιών χαρακτηρίζεται από την ταχεία καθημερινή δημιουργία τεράστιου όγκου δεδομένων, μια πραγματικότητα που θα έχει όλο και αυξανόμενη ένταση. Η κλασική αρχιτεκτονική von Neumann των υπολογιστικών συστημάτων φτάνει στα όριά της και χρειάζονται νέες λύσεις και αρχιτεκτονικές ώστε να αξιοποιηθεί ο αυξανόμενος όγκος δεδομένων καθώς και να ικανοποιηθούν νέες εφαρμογές όπως επεξεργασία σε πραγματικό χρόνο (real time actions) και επεξεργασία στις “άκρες” (edge) των δικτύων. Για την αξιοποίηση των δεδομένων, όλοι οι κλάδοι γύρω από την αποθήκευση μέχρι την αξιοποίηση μέσω τεχνητής νοημοσύνης γνωρίζουν εκρηκτική άνθηση. Μέρος της κάλυψης των νέων υπολογιστικών αναγκών θα το συμπληρώσουν τα νευρομορφικά (neuromorphic) συστήματα, τα οποία αποτελούνται από ένα ευρύ φάσμα διατάξεων, κυκλωμάτων, αλγορίθμων και τον συνδυασμό των άνω ως συστήματα που εμπνέονται από τα νευρωνικά συστήματα των έμβιων όντων.

Ένα δομικό στοιχείο των νευρομορφικών συστημάτων είναι η μνημαντίσταση, η οποία θα αναφέρεται από εδώ και στο εξής ως μνήμη μεταβλητής αντίστασης. Η μνήμη μεταβλητής αντίστασης είναι μια ηλεκτρονική δομή δυο ακροδεκτών της οποίας η μη πτητική δυναμική αντίσταση μπορεί να ελεγχθεί από μια εξωτερική τάση ή ρεύμα και η λειτουργία θυμίζει τις δυναμικές ιδιότητες (μνήμη και προσαρμογή) μιας βιολογικής σύναψης. Η αποθηκευμένη αντίσταση χωρίς τη χρήση εξωτερικής τροφοδοσίας, η δυνατότητα χρήσης ως σύναψη σε μεγάλες δομές παράλληλης επεξεργασίας δεδομένων, η πολύ μικρή διάστασή τους (τάξη των νανομέτρων) καθώς και η εξέλιξη της νανοτεχνολογίας δίνουν την δυνατότητα να δημιουργηθούν νέες μνήμες και συστήματα τεχνητής νοημοσύνης, χαμηλής ισχύος [1].

1.2 Η θεωρία της μνήμης μεταβλητής αντίστασης

Η ιδέα της κατασκευής γνωστικά προσαρμοζόμενων συσκευών στερεάς κατάστασης, εμπνευσμένων από την βιολογία, υπάρχει εδώ και πολλές δεκαετίες[2-3]. Το 1971 όμως ο Leon Chua παρουσίασε θεωρητικά την ύπαρξη της μνήμης μεταβλητής αντίστασης σαν το τέταρτο στοιχείο των ηλεκτρονικών κυκλωμάτων μαζί με την αντίσταση, τον πυκνωτή και το πηνίο [4]. Ο συλλογισμός του Chua ήταν ότι οι τέσσερις βασικοί παράμετροι των ηλεκτρονικών κυκλωμάτων, η τάση V , το ρεύμα I , το φορτίο q και η μαγνητική ροή ϕ μπορούν να συνδυαστούν ανά δυο με έξι διαφορετικούς τρόπους:

$$\text{Αντίσταση: } dV = R * dI \quad (1)$$

$$\text{Πυκνωτής: } dq = C * dV \quad (2)$$

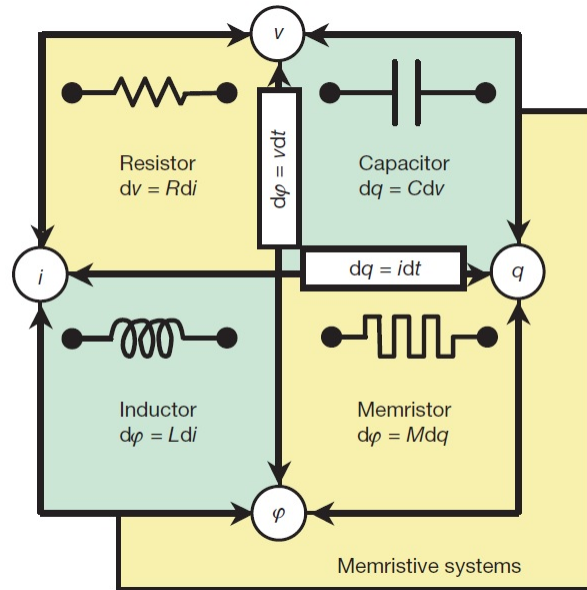
$$\text{Πηνίο: } d\phi = L * dI \quad (3)$$

$$\text{ρεύμα: } dq = I * dt \quad (4)$$

$$\text{Νόμος του Farady: } d\phi = V * dt \quad (5)$$

$$\text{μνήμη μεταβλητής αντίστασης: } d\phi = M(q) * dq \quad (6)$$

όπου R, C, L, t και M η αντίσταση, η χωρητικότητα, η επαγωγή, ο χρόνος και η μνημαντίσταση (memristance) αντίστοιχα. Από τα παραπάνω φαίνεται ότι η μνήμη μεταβλητής αντίστασης συσχετίζει τη μαγνητική ροή με το φορτίο. Η μνημαντίσταση, που χαρακτηρίζει τη μνήμη μεταβλητής αντίστασης, έχει μονάδες αντίστασης – Ohm (Ω).



Σχήμα 1: Τα βασικά ηλεκτρονικά στοιχεία και η μνήμη μεταβλητής αντίστασης του L. O. Chua [5]

Μόλις το 2008 ανακαλύφθηκε η μνήμη μεταβλητής αντίστασης πειραματικά από τους D. B. Strukon, G. S. Snider, D. R. Stewart και R. S. Williams [5], οι οποίοι έδειξαν ότι μια ημιτονοειδής τάση παράγει μια χαρακτηριστική τάσης-ρεύματος (I-V) με τη μορφή βρόγχου υπερτονισμένης υστέρησης (Pinched Hysteresis Loop) σε μια νανοδομή Pt/TiO₂/Pt.

Ο βρόγχος υπερτονισμένης υστέρησης είναι από τα βασικά χαρακτηριστικά μιας μνήμης μεταβλητής αντίστασης και η ύπαρξη του δείχνει τη δυνατότητα χρήσης ως μνήμη (σχήμα 2). Επιπρόσθετα, η υστέρηση εξαρτάται από τη συχνότητα του σήματος εισόδου. Για χαμηλές συχνότητες, η μνήμη μεταβλητής αντίστασης παρουσιάζει την υστέρηση αλλά όσο αυξάνεται η συχνότητα του σήματος εισόδου, τόσο η μνήμη μεταβλητής αντίστασης συμπεριφέρεται σαν μια τυπική γραμμική αντίσταση. Επίσης στη μνήμη μεταβλητής αντίστασης πρέπει η τάση και το ρεύμα να έχουν το ίδιο χρονικό σημείο μηδενισμού ώστε να εξασφαλίζεται η μη αποθήκευση ενέργειας.

Οι εξισώσεις που χαρακτηρίζουν μια ιδανική μνήμη μεταβλητής αντίστασης, αν ελέγχεται από ρεύμα ή τάση αντίστοιχα, είναι [6]:

$$V(t) = M(q(t)) * I(t) = M * \left[\int_{-\infty}^t dt' * I(t') \right] * I(t) \quad (7)$$

$$V(t) = M * \left[\int_{-\infty}^t dt' * I(t') \right] * I(t) \quad (8)$$

όπου $M(q(t))$ είναι μια συνάρτηση αναλογικότητας με διαστάσεις αντίστασης και $q(t)$ είναι το φορτίο που διατρέχει τη διάταξη. Στη σχέση (7), αν η $M(q(t))$ είναι χρονοανεξάρτητη, τότε η σχέση αντιπροσωπεύει τον νόμο του Ohm.

Οι μνήμες μεταβλητής αντίστασης όμως ανήκουν σε πιο περίπλοκα δυναμικά συστήματα όπου μια μεταβλητή κατάσταση (state variable) μπορεί να καθορίζει την κατάσταση του συστήματός τους σε μια χρονική στιγμή αντί του φορτίου τους. Αν x είναι ένα σύνολο από n πιθανές μεταβλητές καταστάσεις, τότε η χρονική εξέλιξη του x μπορεί να οριστεί ως:

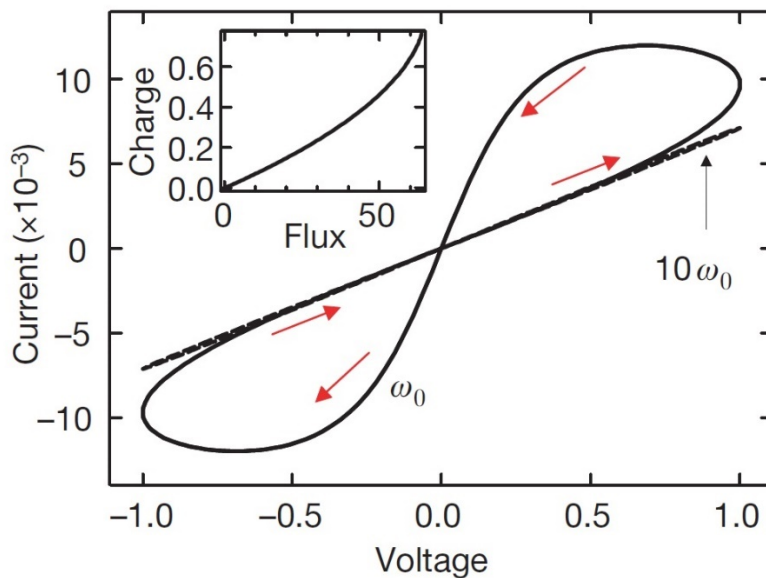
$$\frac{dx}{dt} = f(x, I, t) \quad (9)$$

Όπου η $f(x, I, t)$ είναι μια συνεχής n -διαστάσεων διανυσματική συνάρτηση.

Σε αυτές τις μνήμες μεταβλητής αντίστασης, όπου η μεταβλητή κατάσταση τα χαρακτηρίζει, οι συναρτήσεις I-V για ελεγχόμενα από ρεύμα και από τάση αντίστοιχα είναι:

$$V(t) = M(x, I, t) * I(t) \quad (10)$$

$$V(t) = M(x, V, t) * I(t) \quad (11)$$



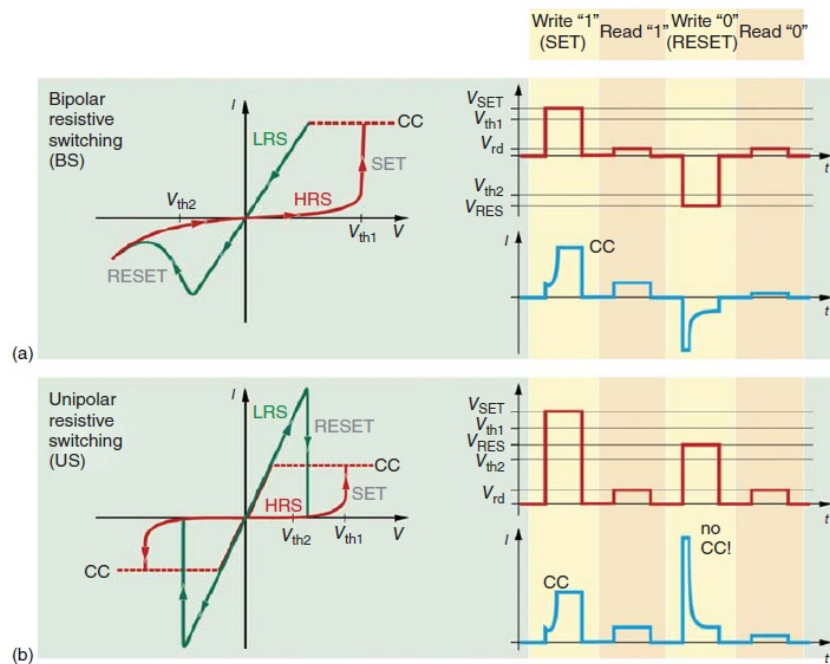
Σχήμα 2: Βρόγχος υπερτονισμένης υστέρησης μνήμη μεταβλητής αντίστασης [5]

1.3 Μηχανισμοί των μνημών μεταβλητής αντίστασης

Οι μνήμες μεταβλητής αντίστασης είναι διατάξεις που θα παίξουν μεγάλο ρόλο στις τωρινές και μελλοντικές ανάγκες της πληροφορικής στα πλαίσια της μνήμης καθώς και δομικού στοιχείου για νευρομορφικές λύσεις. Το στοιχείο που δίνει στις μνήμες μεταβλητής αντίστασης το χαρακτηριστικό της μνήμης είναι η εμφάνιση του βρόγχου υπερτονισμένης υστέρησης και κατ' επέκταση ο μηχανισμός

εναλλαγής αντίστασης (Resistive Switching – RS). Η εναλλαγή αντίστασης είναι η αντιστρεπτή αλλαγή της αντίστασης, με μια μη-μεταβαλλόμενη συμπεριφορά, σε διατάξεις με δυο ακροδέκτες όταν εφαρμοστεί μια ηλεκτρική διέγερση, όπως δηλαδή συμπεριφέρονται οι μνήμες μεταβλητής αντίστασης. Η μη-μεταβολή της κατάστασης σημαίνει ότι η αντίσταση της διάταξης παραμένει σταθερή για ένα μεγάλο χρονικό διάστημα αφότου αφαιρεθεί η ηλεκτρική διέγερση.

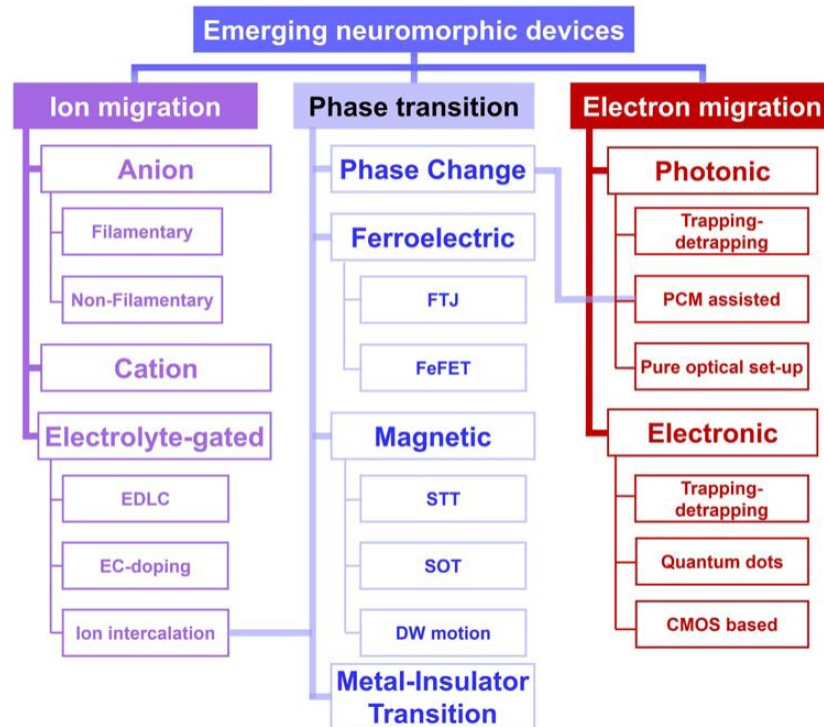
Για τις εφαρμογές ως μνήμη, η μνήμη μεταβλητής αντίστασης αποθηκεύει πληροφορίες εναλλάσσοντας καταστάσεις μέσω της εναλλαγής αντίστασης μεταξύ μιας κατάστασης χαμηλής αντίστασης (low resistive state – LRS; R_{on}) και μιας κατάσταση υψηλής αντίστασης (high resistive state – HRS; R_{off}) κατά την εφαρμογή μιας εξωτερικής τάσης ή ρεύματος. Η εναλλαγή από κατάσταση υψηλής αντίστασης σε κατάσταση χαμηλής αντίστασης ονομάζεται SET ή εγγραφή ενώ το αντίθετο ονομάζεται RESET ή διαγραφή. Όταν η πορεία της εναλλαγής είναι ανεξάρτητη από την πολικότητα του εξωτερικού σήματος αλλά εξαρτάται μόνο από το πλάτος, τότε η εναλλαγή ονομάζεται μονοπολική (unipolar). Σε αυτή τη περίπτωση η εγγραφή και η διαγραφή μπορούν να πραγματοποιηθούν με σήμα εισόδου της ίδιας πολικότητας (σχήμα 3a). Στην αντίθετη περίπτωση, όταν η εναλλαγή εξαρτάται από της πολικότητα του σήματος εισόδου, τότε μιλάμε για διπολική (bipolar) λειτουργία και μια εγγραφή μπορεί να γίνει με αντίστροφης πολικότητας σήματος εισόδου της διαδικασίας διαγραφής (σχήμα 3b).



Σχήμα 3: Μονοπολικοί και διπολικοί μηχανισμοί εναλλαγής αντίστασης [7]

Το σχήμα 3 δείχνει τους δυο τρόπους λειτουργίας της μνήμης μεταβλητής αντίστασης με βάση τη συμπεριφορά της πολικότητας του σήματος εισόδου. Τα αριστερά γραφήματα στο σχήμα 3 δείχνουν την σχέση I-V με περιοδικές τάσεις σάρωσης (voltage sweeps) καθώς οι δεξιές εικόνες δείχνουν τη διέγερση και απόκριση της μνήμης μεταβλητής αντίστασης με παλμούς τάσης. Επίσης στο σχήμα 3 φαίνεται και η τάση κατωφλίου, την οποία πρέπει να περάσει η εξωτερική διέγερση ώστε να αρχίσει να εναλλάσσει καταστάσεις η μνήμη μεταβλητής αντίστασης.

Πέρα από τις δυο κατηγορίες μνημών μεταβλητής αντίστασης που προκύπτουν από τη συμπεριφορά εγγραφής και διαγραφής ανάλογα με την πολικότητα του σήματος εισόδου, οι μνήμες μεταβλητής αντίστασης κατηγοριοποιούνται περαιτέρω με βάση την αρχή λειτουργίας τους που οδηγείται από το υλικό και την αρχιτεκτονική τους. Το σχήμα 4 δείχνει την πληθώρα των κατηγοριών των μνημών μεταβλητής αντίστασης με βάση την αρχή λειτουργίας τους.

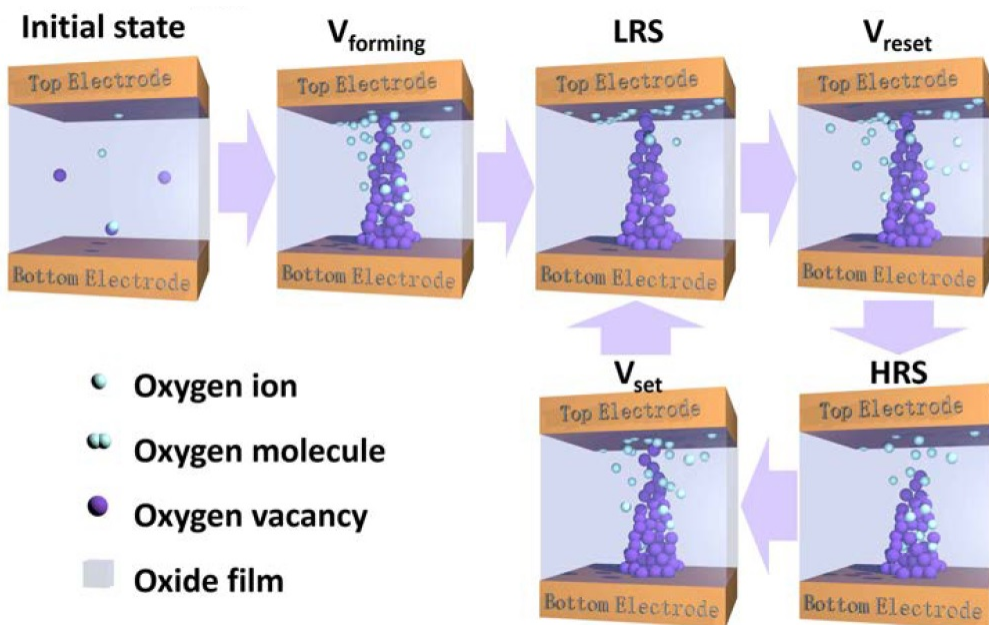


Σχήμα 4: Ταξινόμηση μνημών μεταβλητής αντίστασης με βάση την αρχή λειτουργίας τους [8]

Η κατηγοριοποίηση του σχήματος 4 δείχνει τρεις βασικές κατηγορίες, της μετανάστευσης ιόντων (ion migration), της μετάβασης κατάστασης (phase transition) και της μετανάστευσης ηλεκτρονίων (electron migration) και τις υπό-κατηγορίες αυτών. Στη παρούσα πτυχιακή, θα παρουσιαστεί η υποκατηγορία της μετανάστευσης ιόντων και ειδικά του αγώγιμου δρόμου (Conductive Filamentary - CF) στην οποία βασίζεται το μοντέλο της μνήμης μεταβλητής αντίστασης που χρησιμοποιείται στη προσομοίωση του κεφαλαίου 3. Το σχήμα 4 δίνει μια πολύ πρόσφατη εικόνα του χώρου και ο αναγνώστης μπορεί να εμβαθύνει στη κατηγορία ενδιαφέροντός του.

Η πιο απλή νανοδομή που παρουσιάζει συμπεριφορά εναλλαγής αντίστασης και ιδιότητες μνήμης μεταβλητής αντίστασης με βάση τον αγώγιμο δρόμο είναι η διάταξη Μετάλλου-Μονωτή-Μετάλλου (Metal-Insulator-Metal, MIM). Μια MIM δομή είναι ένα μονωτικό υλικό (οξειδίο) ανάμεσα σε ηλεκτροχημικά ενεργά ή/και αδρανή μεταλλικά ηλεκτρόδια σε δομή σαν πυκνωτή. Οι διατάξεις με αυτή τη δομή ονομάζονται Resistive Random Access Memories (ReRAM) και έχουν πολύ καλές ιδιότητες όπως ταχύτητα εγγραφής κάτω του 1ns [9], μεγάλη αντοχή με περισσότερους από 10^{12} κύκλους εγγραφής-διαγραφής [10], σταθερότητα περισσότερη από 10 χρόνια στους 85°C [11] και πολύ μικρό μέγεθος (κάτω

από 10nm) [12]. Οι MIM δομές είναι οι μόνες που υποστηρίζουν και τους δυο τρόπους λειτουργίας πολικότητας, μονοπολικά και διπολικά, αλλά με διαφορετικούς εσωτερικούς μηχανισμούς. Στη πλειοψηφία της βιβλιογραφίας καθώς και από πειραματικές μετρήσεις/παρατηρήσεις, πιστεύεται ότι στη διαδικασία του αγώγιμου δρόμου δημιουργείται και καταστρέφεται ένας αγώγιμος αγωγός στο οξειδίο, πάχους μερικών nm (αγώγιμο νήμα), το οποίο πραγματοποιεί της εναλλαγή αντίστασης. Τα αγώγιμα νήματα σχηματίζονται από κενές θέσεις οξυγόνου ή/και μεταλλικά ιζήματα.



Σχήμα 5: Μηχανισμός μεταγωγής σε απλή MIM δομή (single layer) [13]

Η λειτουργία για τη δημιουργία του αγώγιμου δρόμου φαίνεται στο σχήμα 5. Κάποια υλικά χρειάζονται σε πρώτη φάση έναν ηλεκτροσχηματισμό (electroforming) πριν μπορέσουν να αρχίσουν να κάνουν εναλλαγή αντίστασης. Ο ηλεκτροσχηματισμός είναι μια ηλεκτροχημική διαδικασία κατά την οποία ιόντα οξυγόνου απομακρύνονται (ή ιόντα μετάλλου εγχέονται, ανάλογα με τα υλικά που χρησιμοποιούνται) από το μονωτικό υλικό της δομής MIM και προκαλείται μια ημι-μόνιμη μετατροπή στη δομή του υλικού η οποία ονομάζεται μαλακή διάτρηση του διηλεκτρικού (soft dielectric breakdown). Γενικά ο ηλεκτροσχηματισμός χρειάζεται μεγαλύτερη τάση από τη τάση που προκαλεί την εγγραφή στη μνήμη μεταβλητής αντίστασης. Αφότου γίνει ο ηλεκτροσχηματισμός, η δομή βρίσκεται σε κατάσταση χαμηλής αντίστασης και με την κατάλληλη τάση γίνεται διαγραφή και η δομή φτάνει στη κατάσταση υψηλής αντίστασης. Από αυτό το σημείο η δομή μπορεί να κάνει εναλλαγή αντίστασης για πολλούς κύκλους.

Αναλυτικά, στη φάση του ηλεκτροσχηματισμού, μετακινούνται ιόντα οξυγόνου προς την διεπαφή της ανόδου. Ανάλογα με το τι υλικό είναι η άνοδος, τα ιόντα είτε αποφορτίζονται αν η άνοδος είναι ευγενές μέταλλο είτε αντιδρούν με την έσω επιφάνεια της ανόδου σχηματίζοντας ένα στρώμα οξειδίου. Παράλληλα, μέσα στον μονωτή οι κενές θέσεις οξυγόνου και τα μεταλλικά ιζήματα σχηματίζουν νήματα και δομές, με το σύνηθες ένα νάνο-νήμα να είναι αυτό που υλοποιεί την αγωγιμότητα (dominant path). Στη συνέχεια από τη κατάσταση χαμηλής αντίστασης και με διέγερση σήματος διαγραφής, το αγώγιμο νήμα σπάει αφού τα ιόντα οξυγόνου επιστρέφουν στον μονωτή στη θέση των κενών οξυγόνων ή/και

οξειδώνουν τα μεταλλικά ιζήματα. Όπως φαίνεται στο σχήμα 5, συνήθως το αγώγιμο νήμα σπάει στο σημείο ένωσης με την άνοδο και κρατάει την δομή του στην κάθοδο. Όπως αναφέρθηκε στην προηγούμενη παράγραφο, στο επόμενο κύκλο εγγραφής – διαγραφής, η διέγερση είναι μικρότερη από αυτή του ηλεκτροσχηματισμού και η εναλλαγή αντίστασης ακολουθεί την ίδια λογική.

Η αρχική δημιουργία του αγώγιμου δρόμου από κατάσταση υψηλής αντίστασης σε κατάσταση χαμηλής αντίστασης είναι ένα φαινόμενο υποκινούμενο από το ηλεκτρικό πεδίο, το οποίο καθορίζει και τη φορά της δημιουργίας του αγώγιμου νήματος. Το φαινόμενο της εναλλαγής αντίστασης μέσω αγώγιμου δρόμου σχετίζεται με την κίνηση των ιόντων του οξυγόνου, εμπλέκοντας τη μεταφορά και τις ηλεκτροχημικές αντιδράσεις ανιόντων ή/και κατιόντων. Η κίνηση των ιόντων μπορεί να γίνεται λόγω μιας μεταβολής του χημικού δυναμικού (diffusion), μιας μεταβολής στο ηλεκτρικό δυναμικό (drift) ή μιας μεταβολής στη θερμοκρασία (thermodifusion) [14]. Ακόμη πιο αναλυτικά, οι Lim και Ismail [15] ανέλυσαν τους περισσότερους μηχανισμούς αγωγιμότητας, οι οποίοι είναι:

- Προερχόμενοι από τους ακροδέκτες/ηλεκτρόδια: (i) Schottky emission, (ii) Fowler-Nordheim (F-N) tunneling και (iii) direct tunneling
- Προερχόμενοι από τον μονωτή: (i) Poole-Frenkel (P-F) emission, (ii) Ohmic conduction, (iii) SCLC, (iv) ionic conduction, (v) hopping conduction και (vi) trap-assisted tunneling (TAT)

1.4 Μοντέλα των μνημών μεταβλητής αντίστασης

Η μοντελοποίηση των μνημών μεταβλητής αντίστασης είναι απαραίτητη για τη σχεδίαση μνημών, νευρομορφικών κυκλωμάτων και συστημάτων γενικότερα. Γενικά, οι περισσότερες προσεγγίσεις για την μοντελοποίηση μνημών μεταβλητής αντίστασης, συμπεριλαμβανομένου των ReRAM, ακολουθούν τον φορμαλισμό του L. Chua [4], ο οποίος λέει ότι μια διάταξη μνήμης μεταβλητής αντίστασης μπορεί να χαρακτηριστεί από δυο συζευμένες διαφορικές εξισώσεις:

$$I(t) = G(x, V, t) * V(t) \quad (12)$$

$$\dot{x} = F(x, V, t) \quad (13)$$

όπου $V(t)$ είναι η χρονοεξαρτώμενη τάση εισόδου, $I(t)$ το χρονοεξαρτώμενο ρεύμα που διαπερνά την μνήμη μεταβλητής αντίστασης και x είναι ένα διάνυσμα που περιγράφει τη μεταβλητή κατάσταση της μνήμης μεταβλητής αντίστασης και τη χαρακτηρίζει σε οποιαδήποτε χρονική στιγμή. Όπως θα δει ο αναγνώστης στο κεφάλαιο της προσομοίωσης και γίνεται συχνά στη μοντελοποίηση των μνημών μεταβλητής αντίστασης, το x χρησιμοποιείται ως βαθμωτή τιμή για λόγους απλοποίησης. Η σχέση 12 δίνει τη στατική σχέση τάσης-ρεύματος ενώ η σχέση 13 δίνει τη δυναμική της διάταξης μέσω χρονικών βαθμίδων ενός συνόλου της μεταβλητής κατάστασης x .

1.4.1 Γραμμικά & μη-γραμμικά μοντέλα

Τα μοντέλα χωρίζονται σε δύο βασικές κατηγορίες, τα γραμμικά και τα μη-γραμμικά.

Τα αρχικά μοντέλα βασίζονταν στην αρχική κατασκευή μνημών μεταβλητής αντίστασης από την HP, οι οποίες αποτελούνταν από δομές MIM. Τα μοντέλα αυτά θεωρούν ότι η μνήμη μεταβλητής αντίστασης αποτελείται από δυο περιοχές του μονωτή με διαφορετικές αγωγιμότητες και μοντελοποιείται με δυο αντιστάσεις σε σειρά. Η μεταβλητή κατάσταση είναι ο σχετικός λόγος του πάχους της περιοχής με την υψηλότερη αγωγιμότητα προς το πάχος της περιοχής με την μικρότερη αγωγιμότητα [5]. Η θεώρηση αυτή αν και δημιουργεί ένα απλό και λειτουργικό μοντέλο δεν αποτυπώνει τη συμπεριφορά της μνήμης μεταβλητής αντίστασης όταν η μεταβλητή κατάσταση φτάσει στα όρια της διάταξης. Σε αυτά τα μοντέλα, η μεταβλητή κατάσταση διατρέχει όλο το υλικό κάτι το οποίο δεν ισχύει στη πραγματικότητα και δίνει υποδεέστερα αποτελέσματα σε σύγκριση με τα εξελιγμένα μη-γραμμικά μοντέλα. Τα μη-γραμμικά μοντέλα, για να αντιμετωπίσουν τη μη-γραμμική συμπεριφορά της μεταβλητής κατάστασης στα όρια της διάταξης (του μονωτικού υλικού) χρησιμοποιούν μια συνάρτηση παραθύρου (windowing function) ώστε να περιορίσουν το εύρος και να προσομοιώσουν σωστά τη συμπεριφορά της μεταβλητής κατάστασης. Το αρχικό γραμμικό μοντέλο εξελίχθηκε σε μη-γραμμικό με τη χρήση δύο συναρτήσεων παραθύρου, του Joblekar [16] και του Biolek [17]. Στη συνέχεια δημιουργήθηκαν και άλλες βελτιώσεις/παραλλαγές συναρτήσεων παραθύρου, όπως αυτές του Προδρομάκη [26] και το ThrEshold Adaptive Memristor (TEAM) [21].

1.4.2 Εξελιγμένα μη-γραμμικά μοντέλα

Πέρα από τα προηγούμενα γραμμικά και μη-γραμμικά μοντέλα, δημιουργήθηκαν πιο εξελιγμένα μοντέλα που αποτυπώνουν καλύτερα τη συμπεριφορά των μνημών μεταβλητής αντίστασης.

Το πρώτο μοντέλο είναι το **Simmons tunnel barrier model** [18-19], το οποίο αντί να θεωρήσει δυο αντιστάσεις σε σειρά, θεωρεί μια αντίσταση σε σειρά με ένα φραγμό σήραγγος ηλεκτρονίων (electron tunnel barrier). Το πάχος του φραγμού είναι η μεταβλητή κατάσταση του συστήματος. Το μοντέλο αυτό έχει ισχυρή βάση στα φυσικά φαινόμενα που διαδραματίζονται κατά τη λειτουργία μιας μνήμης μεταβλητής αντίστασης και αποτυπώνουν εξαιρετικά τις διατάξεις με βάση τα οξειδία του Τιτανίου (TiO_2) και Τανταλίου (Ta_2O_5) αλλά έχουν μεγάλο υπολογιστικό κόστος λόγω των ένθετων εκθετικών που χρησιμοποιούν οι εξισώσεις του μοντέλου (σε παρακάτω πίνακα).

Το μοντέλο του **Yakovpic** [20], το οποίο είναι το μοντέλο που χρησιμοποιείται στη προσομοίωση του τρίτου κεφαλαίου όπου και γίνεται αναλυτική περιγραφή, χρησιμοποιείται ώστε να περιγράψει τη στατική συμπεριφορά διάφορων μνημών μεταβλητής αντίστασης, προσαρμόζοντας τις εξισώσεις του σε πειραματικές μετρήσεις. Έτσι μπορεί κάποιος να προσομοιώσει νευρομορφικά συστήματα με διάφορες διατάξεις. Η συνάρτηση παραθύρου που χρησιμοποιεί το μοντέλο είναι μια απλή μορφή του φραγμού σήραγγος ηλεκτρονίων. Επιπρόσθετα, το μοντέλο υποστηρίζει τάσεις κατωφλίου. Είναι πολύ πιο γρήγορο υπολογιστικά από το Simmons tunnel model.

Το **TEAM** [21] είναι μια αριθμητικά απλοποιημένη μορφή του Simmons tunnel model και επιτυγχάνει μεγάλο υπολογιστικό κέρδος σε σχέση με αυτό. Το TEAM μοντέλο υποστηρίζει, όπως του Yakovpic, τάσεις (και ρεύματα) κατωφλίου. Αρχικά το μοντέλο σχεδιάστηκε για μνήμες μεταβλητής αντίστασης ελεγχόμενα από ρεύμα, αλλά υπήρξε επέκταση που υποστηρίζει έλεγχο από τάση (VTEAM [22]). Η συνάρτηση παραθύρου που χρησιμοποιείται είναι και πάλι με βάση της φυσικής λογικής του φραγμού σήραγγος ηλεκτρονίων.

1.4.3 Σχέσεις/συναρτήσεις μοντέλων

Οι σχέσεις που ορίζουν τα παραπάνω μοντέλα φαίνονται συγκεντρωτικά στους παρακάτω πίνακες. Οι πίνακες περιλαμβάνουν τις σχέσεις τάσης-ρεύματος του κάθε μοντέλου, τις σχέσεις της μεταβλητής κατάστασης που χρησιμοποιεί το κάθε μοντέλο, συγκριτικά στοιχεία ανάμεσα στα μοντέλα καθώς και σύγκριση των διάφορων συναρτήσεων παραθύρου.

Model	I or V	$k(x)$	\times	$h(I$ or $V)$
Linear ion drift [5]	$V(t) =$	$\frac{R_{on}x(t)}{D} + R_{off} \left(1 - \frac{x(t)}{D}\right)$	\times	$I(t)$
Nonlinear ion drift [24]	$I(t) =$	$x(t)^n \times \beta$	\times	$\sinh(\alpha * V(t)) + \chi \times (\exp(\gamma V(t)) - 1)$
Simmons tunneling barrier [25]	$I(t) =$	$\frac{j_0 A}{\Delta w^2(x, V)} \left\{ \phi_I(x, V) e^{-B\sqrt{\phi_I(x, V)}} - (\phi_I(x, V) + e v_g(V)) e^{-B\sqrt{\phi_I(x, V) + e v_g(V)}} \right\}$		
TEAM [21]	$V(t) =$	$e^{\left(\frac{\lambda}{x_{off}} - x_{on}\right)(x - x_{on})}$	\times	$I(t) \times R_{on}$
Yakopcic [20]	$I(t) =$	$x(t)$	\times	$\begin{cases} a_1 \sinh(b_1 V(t)), & V(t) \geq 0 \\ a_2 \sinh(b_2 V(t)), & V(t) < 0 \end{cases}$

Πίνακας 1: Οι I-V σχέσεις των διάφορων μοντέλων μνήμων μεταβλητής αντίστασης [23]

Model	$\frac{dx}{dt} =$	$f(x)$	\times	$g(I$ or $V)$
Linear ion drift [5]	$\frac{dx}{dt} =$	1	\times	$\frac{\mu_n R_{on}}{D} \times I$
Nonlinear ion drift [16]	$\frac{dx}{dt} =$	$1 - (2x - 1)^{2p}$	\times	$a \times V^m$
Nonlinear ion drift [17]	$\frac{dx}{dt} =$	$\begin{cases} 1 - (x)^{2p}, & I \geq 0 \\ 1 - (x - 1)^{2p}, & I < 0 \end{cases}$	\times	$a \times V^m$
Simmons tunneling barrier [25]	$\frac{dx}{dt} =$	$\begin{cases} c_{off} \sinh\left(\frac{I}{i_{off}}\right) \exp\left[-\exp\left(\frac{x - a_{off}}{w_c} - \frac{ I }{b}\right) - \frac{x}{w_c}\right], & I \geq 0 \\ c_{on} \sinh\left(\frac{I}{i_{on}}\right) \exp\left[-\exp\left(\frac{x - a_{on}}{w_c} - \frac{ I }{b}\right) - \frac{x}{w_c}\right], & I < 0 \end{cases}$		
TEAM [21]	$\frac{dx}{dt} =$	$\begin{cases} \exp\left[-\exp\left(\frac{x - a_{off}}{w_c}\right)\right], & I \geq 0 \\ \exp\left[-\exp\left(-\frac{x - a_{on}}{w_c}\right)\right], & I < 0 \end{cases}$	\times	$\begin{cases} k_{off} \left(\frac{I(t)}{i_{off}} - 1\right)^{a_{off}}, & 0 < i_{off} < I \\ 0, & i_{on} < I < i_{off} \\ k_{on} \left(\frac{I(t)}{i_{on}} - 1\right)^{a_{on}}, & I < i_{on} < 0 \end{cases}$
Yakopcic [20]	$\frac{dx}{dt} =$	$\begin{cases} e^{-\alpha_p(x-x_p)} \frac{x_p - x}{1 - x_p} + 1, & x \geq x_p, V \geq 0 \\ 1, & x < x_p \\ e^{\alpha_n(x-x_p)} \frac{x}{1 - x_n}, & x \leq 1 - x_n, V < 0 \\ 1, & x > 1 - x_n \end{cases}$	\times	$\begin{cases} A_p (e^{V(t)} - e^{V_p}), & V(t) > V_p \\ -A_n (e^{V(t)} - e^{V_n}), & V(t) < -V_n \\ 0, & -V_n \leq V(t) \leq V_p \end{cases}$

Πίνακας 2: Οι μεταβλητές καταστάσεων του κάθε μοντέλου μνήμης μεταβλητής αντίστασης [23]

Model	Linear ion drift [5]	Nonlinear ion drift [27]	Simmons tunneling barrier [18]	Yakopcic <i>et al</i> [20]	TEAM [21]
State variable	$0 \leq w \leq D$ Doped region physical width	$0 \leq w \leq 1$ Doped region normalized width	$a_{off} \leq x \leq a_{on}$ Undoped region width	$0 \leq x \leq 1$ No physical explanation	$x_{on} \leq x \leq x_{off}$ Undoped region width
Control mechanism	Current controlled	Voltage controlled	Current controlled	Voltage controlled	Current controlled
Current-voltage relationship and memristance deduction	Explicit	I-V relationship – explicit Memristance deduction - ambiguous	Ambiguous	Ambiguous	Explicit
Matching memristive system definition	Yes	No	No	No	Yes
Generic	No	No	No	Moderate	Yes
Accuracy comparing practical memristive devices	Lowest accuracy	Low accuracy	Highest accuracy	Moderate accuracy	Sufficient accuracy
Threshold exists	No	No	Practically exists	Yes	Yes

Πίνακας 3: Σύγκριση των μοντέλων μνημών μεταβλητής αντίστασης [21]

Function	Joglekar[16]	Biolek[17]	Prodromakis[26]	TEAM [21]
$f(x)/f(w)$	$f(w) = 1 - (2w/D - 1)^{2p}$	$f(w) = 1 - (w/D - \text{stp}(-i))^{2p}$	$f(w) = j(1 - [(w - 0.5)^2 + 0.75]^p)$	$f_{on,off} = \exp[-\exp(x - x_{on,off} /w_c)]$
Symmetric	Yes	Yes	Yes	Not necessarily
Resolve boundary conditions	No	Yes	Practically yes	Practically yes
Impose nonlinear drift	Partially	Partially	Partially	Yes
Scale factor $f_{max} < 1$	No	No	Yes	No
Fits memristive device model	Linear/nonlinear ion drift/TEAM	Linear/nonlinear ion drift/TEAM	Linear/nonlinear ion drift/TEAM	TEAM for Simmons tunneling barrier fitting

Πίνακας 4: Σύγκριση των διάφορων συναρτήσεων παραθύρου [21]

1.5 Αναφορές 1^{ου} κεφαλαίου

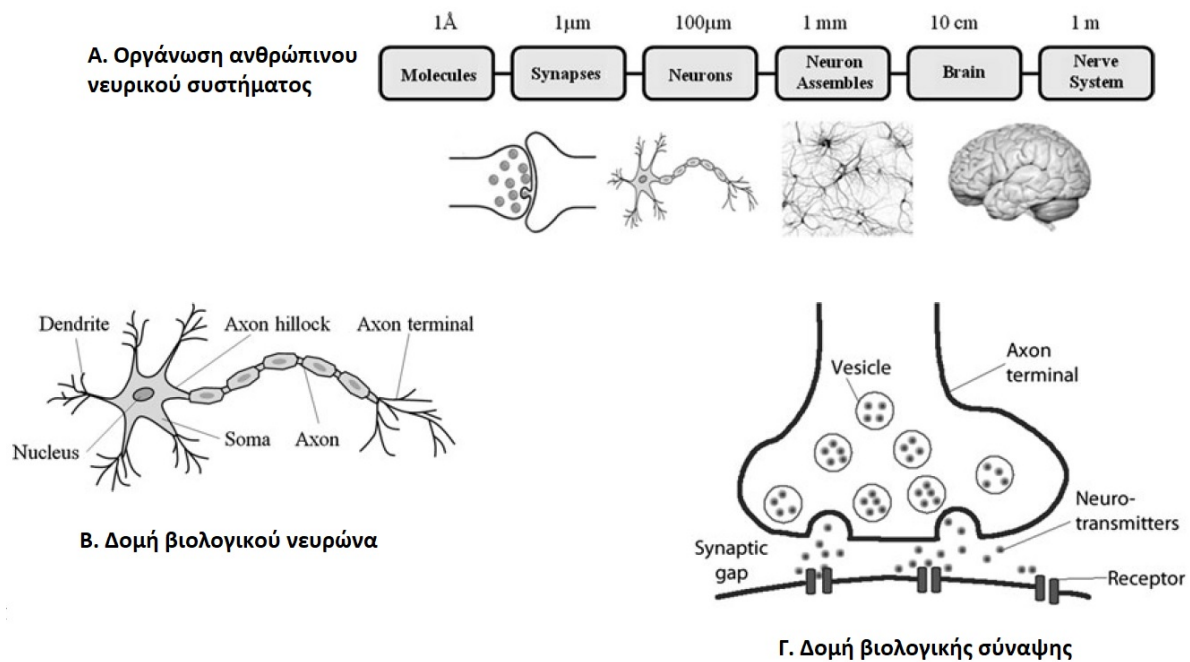
- [1] R. Kozma, R. E. Pino, G. E. Paziienza, “Advances in Neuromorphic Memristor Science and Applications”, Springer Netherlands, Dordrecht, 2012.
- [2] K. Steinbuch, Ettingen, Baden, and P. Muller, “Learning Matrix for Analog Signals”, Patented Nov. 15, 1966, <https://www.freepatentsonline.com/3286238.pdf> .
- [3] R. Stanford, Iris M. Ovshinsky, “Analog Models for Information Storage and Transmission in Physiological Systmes”, Mat. Res. Bull. Vol. 5, pp. 681-690, Pergamon Press , Inc., 1970.
- [4] L. Chua, “Memristor - The missing circuit element”, IEEE Transactions on Circuit Theory, Volume: 18, Issue: 5, 1971.
- [5] Dmitri B. Strukov, Gregory S. Snider, Duncan R. Stewart, and R. Stanley Williams, “The missing memristor found”, Nature, Vol 453, 2008.
- [6] Y. V. Pershin, and M. Di Ventra, “Memory effects in complex materials and nanoscale systems”, Advances in Physics, 2011.
- [7] R. Waser, D. Ielmini, H. Akinaga, H. Shima, H.-S. P. Wong, J. J. Yang, and S. Yu, “Introduction to Nanoionic Elements for Information Technology”, Wiley-VCH Verlag GmbH & Co. KGaA, Chapter 1, 2016.
- [8] J. Zhu, T. Zhang, Y. Yang, and R. Huang, “A comprehensive review on emerging artificial neuromorphic devices”, Applied Physics Reviews, Rev. 7, 011312, 2020.
- [9] B. J. Choi, A. C. Torrezan, J. P. Strachan, P. G. Kotula, A. J. Lohn, M. J. Marinella, Z. Li, R. S. Williams, and J. J. Yang, “Advanced Functional Materials”, High-Speed and Low-Energy Nitride Memristors, 2016.
- [10] J. J. Yang, D. B. Strukov, and D. R. Stewart, “Memristive devices for computing”, Nature Nanotechnology, 2012.
- [11] T. Ninomiya, S. Muraoka, Z. Wei, R. Yasuhara, K. Katayama, and T. Takagi, “Improvement of Data Retention During Long-Term Use by Suppressing Conductive Filament Expansion in TaOx Bipolar-eRAM”, IEEE Electron Device Letters, Volume: 34, Issue: 6, 2013.
- [12] S. Pi, C. Li, H. Jiang, W. Xia, H. Xin, J. J. Yang, and Q. Xia, “Memristor crossbar arrays with 6-nm half-pitch and 2-nm critical dimension”, Nature Nanotechnology, volume 14, pages35–39, 2019.
- [13] J. Wu, J. Cao, W. Han, A. Janotti, and H. Kim, “Functional Metal Oxide Nanostructures”, Springer, 2012.
- [14] R. A. De Souza, “Ion Transport in Metal Oxides, in: Resistive Switching”, Wiley-VCH Verlag GmbH & Co., Chapter 5, 2016.
- [15] E. W. Lim, and R. Ismail, “Conduction Mechanism of Valence Change Resistive Switching Memory: A Survey”, Electronics, ISSN 2079-9292, 2015.
- [16] Y. N. Joglekar, and S. J Wol, “The elusive memristor: properties of basic electrical circuits”, IOP Publishing, Eur. J. Phys. 30 (2009) 661–675.

- [17] Z. Biolek, D. Biolek, and V. Biolkova, "SPICE model of memristor with nonlinear dopant drift", *Radioengineering*, Vol. 18, No. 2, 2009.
- [18] M. D. Pickett, D. B. Strukov, J. L. Borghetti, J. J. Yang, and G. S. Snider, "Switching dynamics in titanium dioxide memristive devices", *Journal of Applied Physics*, 106, 074508, 2009
- [19] J. G. Simmons, "Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film", *Journal of Applied Physics*, 34(6), 1793–1803, 1963.
- [20] C. Yakopcic, T. M. Taha, G. Subramanyam, R. E. Pino, and S. Rogers, "A Memristor Device Model", *IEEE Electron Device Letters*, Vol. 32, No. 10, 2011.
- [21] S. Kvatinsky, E. G. Friedman, and A. Kolodny, "TEAM: ThrEshold Adaptive Memristor Model", *IEEE Transactions on Circuits and Systems, Regular Papers* Vol. 60 No.1, 2013.
- [22] S. Kvatinsky, M. Ramadan, E. G. Friedman, and A. Kolodny, "VTEAM: A General Model for Voltage-Controlled Memristors", *IEEE Transactions on Circuits and Systems*, Volume: 62 Issue: 8, 2015.
- [23] J. Niroula, S. Agarwal, R. Jacobs-Gedrim, R. L. Schiek, D. Hughart, A. Hsia, C. D. James, and M. J. Marinella, "Piecewise empirical model (PEM) of resistive memory for pulsed analog and neuromorphic applications", *Springer Science, J Comput Electron*, 2017.
- [24] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, and R. Stanley Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices", *Nature Nanotechnology* volume 3, pages 429–433, 2008.
- [25] John Paul Strachan; Antonio C. Torrezan; Feng Miao; Matthew D. Pickett; J. Joshua Yang; Wei Yi, and G. Medeiros-Ribeiro, "State Dynamics and Modeling of Tantalum Oxide Memristors", *IEEE Transactions on Electron Devices*, Volume: 60, Issue: 7, 2013.
- [26] T. Prodromakis, B. P. Peh, C. Papavassiliou, and C. Toumazou, "A Versatile Memristor Model With Nonlinear Dopant Kinetics", *IEEE Transactions on Electron Devices*, Volume: 58, Issue: 9, 2011.
- [27] E. Lehtonen and M. Laiho, "CNN using memristors for neighborhood connections", *IEEE, 2010 12th International Workshop on Cellular Nanoscale Networks and their Applications, CNNA 2010*.

Κεφάλαιο 2: Τεχνητή νοημοσύνη και οι μνήμες μεταβλητής αντίστασης

2.1 Συνάψεις στη βιολογία και την ηλεκτρονική

Ο ανθρώπινος εγκέφαλος έχει δυο θεμελιώδη στοιχεία για τη μάθηση και τη μνήμη, τις συνάψεις και τους νευρώνες. Οι νευρώνες είναι κύτταρα ηλεκτρικά ευερέθιστα τα οποία μπορούν να ανταποκριθούν σε ερεθίσματα, να παράξουν παλμούς και να επικοινωνήσουν με άλλους νευρώνες. Οι συνάψεις είναι χασμοσύνδεσμοι (synaptic gap junctions) της τάξεως μερικών δεκάδων nm μεταξύ των νευρώνων και υλοποιούν την επικοινωνία μεταξύ των νευρώνων μέσω ταχείας μετάδοσης ηλεκτρικών ή χημικών σημάτων [1].

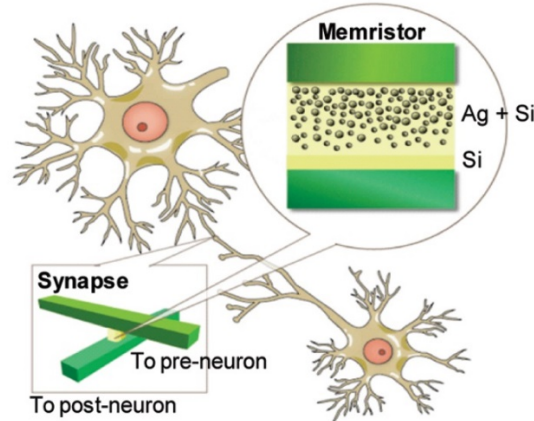


Σχήμα 6: Νευρώνες και συνάψεις στους βιολογικούς οργανισμούς [1]

Οι συνάψεις είναι υπεύθυνες να παρέχουν στους νευρώνες ένα τροποποιημένο σήμα εισόδου από άλλους νευρώνες που είναι ανάλογο με τη σημασία αυτού του σήματος. Όταν ένα παραγόμενο δυναμικό ενέργειας από ένα νευρώνα φτάσει σε ένα προ-συναπτικό ακροδέκτη, μια διαδοχή γεγονότων οδηγεί στην απελευθέρωση νευροδιαβιβαστών, οι οποίοι στη συνέχεια αυξάνουν μια ροή ιοντικών ρευμάτων προς ή από τον μετά-συναπτικό νευρώνα.

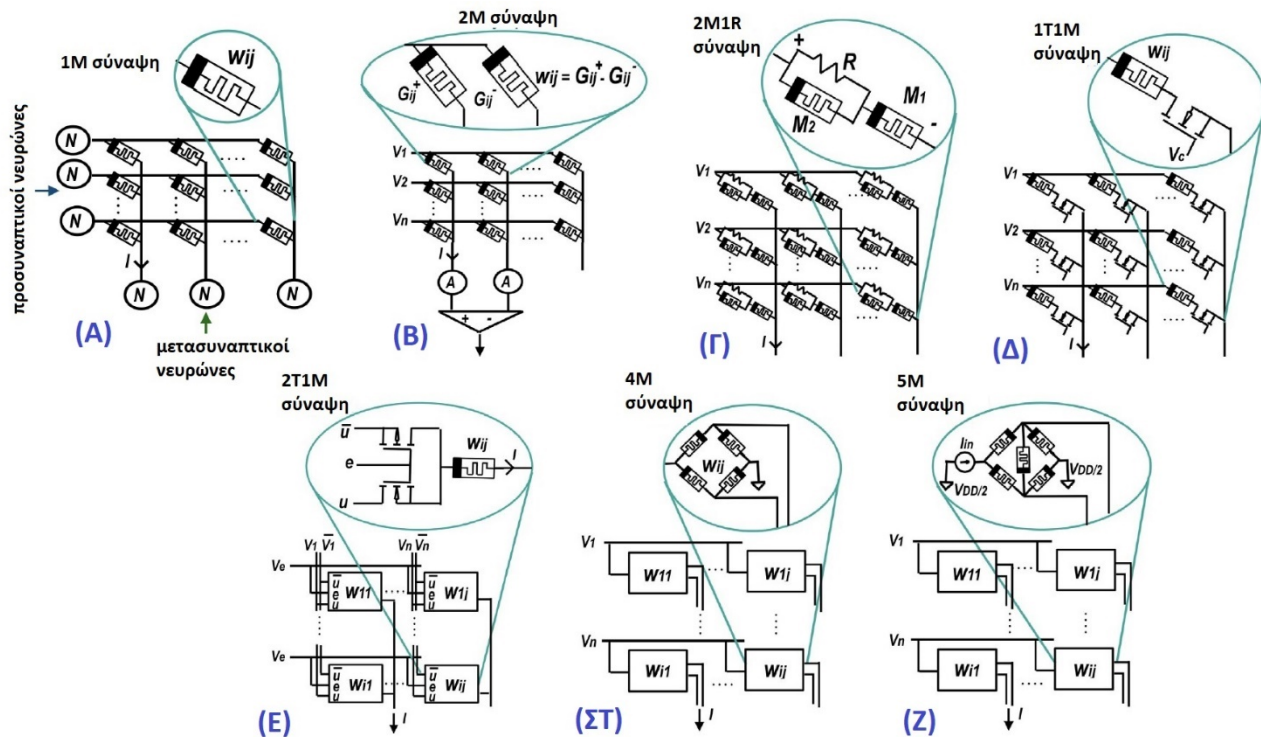
Η μάθηση και η μνήμη στους βιολογικούς εγκέφαλους είναι η ικανότητά τους να προσλαμβάνουν νέες πληροφορίες και να τις αποθηκεύουν ώστε να μπορούν να τις ανακαλούν. Στην βιολογία, είναι πλέον αποδεκτό ότι η πληροφορία αποθηκεύεται στο συναπτικό "βάρος" και η μάθηση πραγματοποιείται με τη τροποποίηση του συναπτικού βάρους, είτε αυξάνοντάς είτε μειώνοντάς το [2]. Η πλαστικότητα των βαρών των συνάψεων δίνει τη δυνατότητα για την αποθήκευση πληροφοριών αλλά και στην απόκριση σε νέα σήματα με βάση προηγούμενη γνώση [3]. Από τα παραπάνω, γίνεται προφανές ότι το ρόλο της

σύναψης στο πεδίο των υλικών μπορεί να τον αναλάβει η μνήμη μεταβλητής αντίστασης, η οποία με τους δυο ακροδέκτες της μπορεί να παρεμβάλλεται μεταξύ δυο νευρώνων και να μεταβάλλει και να διατηρεί την αγωγιμότητά της σαν συναπτικό βάρος (σχήμα 7).



Σχήμα 7: Η μνήμη μεταβλητής αντίστασης ως σύναψη [4]

Η δομή της σύναψης μπορεί να είναι απλά μια μνήμη μεταβλητής αντίστασης αλλά και διάφοροι άλλοι συνδυασμοί και συνδεσμολογίες μνημών μεταβλητής αντίστασης και άλλων ηλεκτρονικών στοιχείων όπως τρανζίστορ και αντιστάσεις (σχήμα 8).



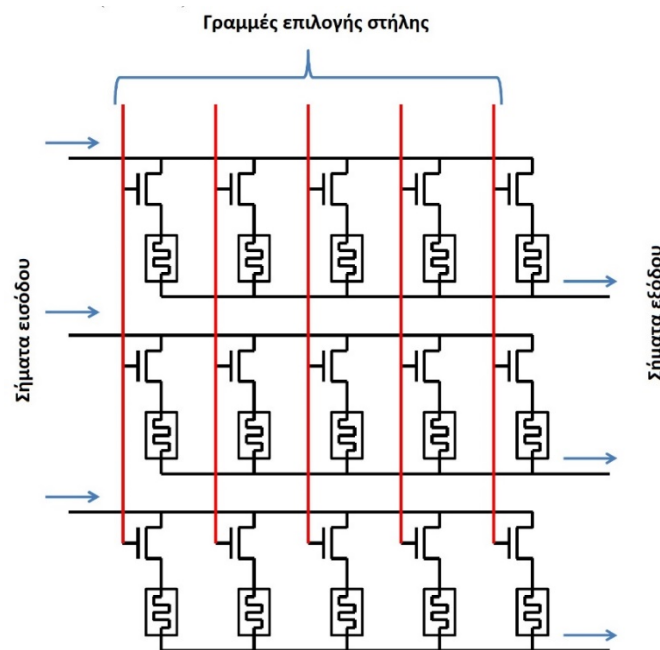
Σχήμα 8: Είδη συνάψεων - (A) 1M , (B) 2M [6], (Γ) 2M1R [7], (Δ) 1T1M [8], (E) 2T1M [9], (ΣΤ) 4M [10], 5M [11]

Η χρήση μιας μνήμης μεταβλητής αντίστασης (1M) ως σύναψη σε TND έχει σαν προτέρημα την επίτευξη της μεγαλύτερης δυνατής (όρια λιθογραφικής μεθόδου) πυκνότητας διατάξεων στην επιφάνεια ενός προς κατασκευή κυκλώματος καθώς και την μικρότερη πολυπλοκότητα κατασκευής σε σχέση με τις άλλες διατάξεις συνάψεων. Το σημείο που υστερεί η 1M διάταξη είναι ότι επιτρέπει διαδρομές ρευμάτων ερπυσμού (sneak path currents) μέσα στο κύκλωμα, ένα πρόβλημα που περιορίζει το μέγεθος των TND σε διατάξεις δικτυωμάτων με μνήμες μεταβλητής αντίστασης (crossbars) (ανάλυση σε επόμενη παράγραφο). Επιπρόσθετα, η 1M διάταξη επιτρέπει μόνο θετικά βάρη κατά τη λειτουργία και εκπαίδευση του TND.

Ο κάθε συνδυασμός μνημών μεταβλητής αντίστασης επιτυγχάνει να βελτιώσει τους περιορισμούς της 1M σύναψης σε διαφορά επίπεδα. Οι δυο πιο διαδεδομένοι συνδυασμοί στη βιβλιογραφία είναι η χρήση διπλών μνημών μεταβλητής αντίστασης (2M) ανά σύναψη και η χρήση τρανζίστορ στη σειρά με τη μνήμη μεταβλητής αντίστασης.

Η διάταξη 2M χρησιμοποιείται για την επίτευξη θετικών και αρνητικών βαρών στα TND καθώς κάθε σύναψη αποτελείται από δυο μνήμες μεταβλητής αντίστασης των οποίων το αποτέλεσμα μέσα στο TND είναι η αφαίρεση του ενός από το άλλο. Στο σχήμα 8 (B) φαίνεται η υλοποίηση της 2M λογικής όπου τα σήματα εισόδου εισέρχονται σε ένα δικτύωμα μνημών μεταβλητής αντίστασης και η έξοδος δυο στηλών οδηγεί ένα διαφορικό ενισχυτή. Η προσέγγιση αυτή οδηγεί σε καλύτερη εκπαίδευση του TND αλλά δεν λύνει το πρόβλημα των ρευμάτων ερπυσμού.

Η διάταξη 1T1M χρησιμοποιεί ένα τρανζίστορ σε σειρά με μια μνήμη μεταβλητής αντίστασης. Το τρανζίστορ έχει τον ρόλο του επιλογέα μέσα στο TND και τα δικτυώματα μνημών μεταβλητής αντίστασης με τη χρήση 1T1M μπορούν να απομονώσουν στήλες από μνήμες μεταβλητής αντίστασης και έτσι να μην επηρεάζονται από ρεύματα ερπυσμού κατά την εκπαίδευση, προγραμματισμό ή λειτουργία τους. Η λύση αυτή βέβαια χρειάζεται κατάλληλο χρονικό προγραμματισμό των σημάτων εισόδου και της τροφοδοσίας των πυλών των τρανζίστορ.

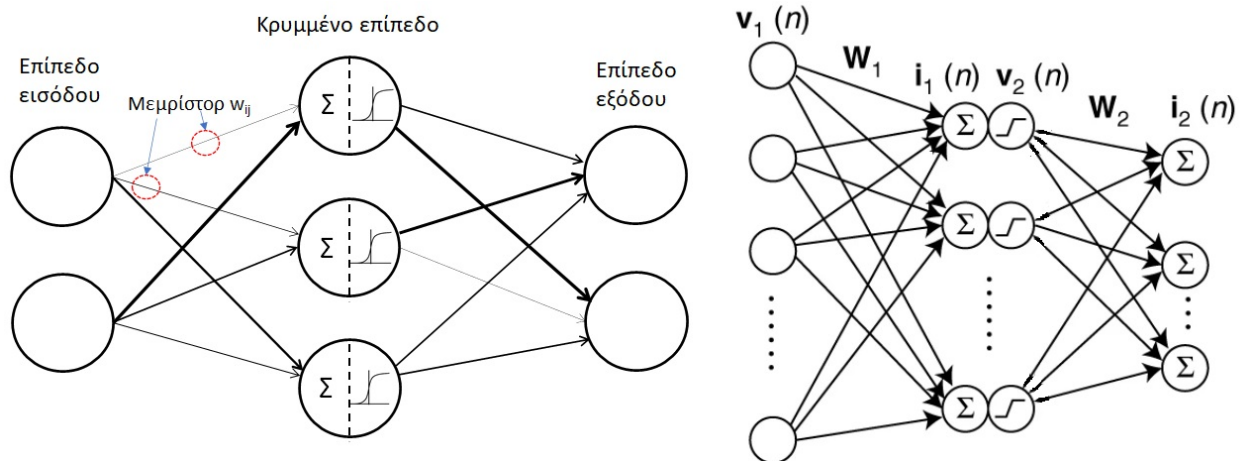


Σχήμα 9: Απομόνωση στηλών δικτύωμα μνημών μεταβλητής αντίστασης με συνάψεις 1T1M [12]

2.2 Τεχνητά νευρωνικά δίκτυα (ΤΝΔ)

Η πληροφορική εμπνευσμένη από τη λειτουργία των βιολογικών εγκεφάλων αποτελεί ένα μέρος της τεχνητής νοημοσύνης και αποσκοπεί στο να λύνει προβλήματα με οδηγό τις αρχές λειτουργίας που χρησιμοποιούν οι εγκεφαλοι. Οι εγκεφαλοι μαθαίνουν, αποθηκεύουν πληροφορίες, λύνουν προβλήματα και επεξεργάζονται απευθείας αναλογικά σήματα. Ο εγκέφαλος επεξεργάζεται τις πληροφορίες στο σημείο που είναι αποθηκευμένες χωρίς να χωρίζει τις λειτουργίες της επεξεργασίας και της μνήμης, όπως γίνεται στις αρχιτεκτονικές von Neumann.

Ο χώρος της τεχνητής νοημοσύνης και συγκεκριμένα των ΤΝΔ στο επίπεδο των υλικών και των μνημών μεταβλητής αντίστασης χωρίζεται σε δυο βασικές κατηγορίες. Η πρώτη είναι τα ΤΝΔ Αιχμών (spiking neural networks – SNNs) όπου η πληροφορία κωδικοποιείται και μεταφέρεται βασισμένα στη χρονική σχέση μεταξύ των διάφορων παλμών. Οι υπολογισμοί των SNN ενεργοποιούνται όταν υπάρχει χρονική συσχέτιση μεταξύ των σημάτων πριν και μετά τη σύναψη. Τα μεγάλα οφέλη των SNN είναι η μικρότερη κατανάλωση ενέργειας σε σχέση με τα άλλα είδη ΤΔΝ καθώς και η προσέγγιση στον πραγματικό τρόπο επεξεργαστικής λειτουργίας των έμβιων νευρωνικών δικτύων στον χρόνο επεξεργασίας, ο οποίος προσεγγίζει τον πραγματικό χρόνο (real time). Η δεύτερη βασική κατηγορία ΤΝΔ είναι αυτή που δεν κωδικοποιεί τα σήματα με βάση τον χρονισμό και δεν έχει αιχμές. Αυτά τα ΤΝΔ αποτελούνται από επίπεδα (layers) τεχνητών νευρώνων που διαδοχικά δέχονται, επεξεργάζονται και προωθούν σήματα και από συνάψεις που συνδέουν τους νευρώνες με μεταβαλλόμενο βάρος. Σε αυτή τη κατηγορία των ΤΝΔ ανήκει η προσομοίωση του επόμενου κεφαλαίου. Στο παρακάτω σχήμα φαίνεται ένα ενδεικτικό ΤΝΔ σε κλασική αναπαράσταση (αριστερά) καθώς και μια αναπαράσταση με ηλεκτρονική λογική [5] (δεξιά).



Σχήμα 10: Αναπαράσταση ΤΝΔ

Οι κύκλοι στα παραπάνω σχήματα αντιπροσωπεύουν τους νευρώνες. Στο επίπεδο εισόδου τα ΤΝΔ δέχονται τα σήματα εισόδου. Τα σήματα εισόδου προωθούνται στις συνάψεις που συνδέουν το επίπεδο εισόδου με το κρυμμένο επίπεδο. Το άθροισμα του γινομένου των βαρών με τα σήματα εισόδου σε κάθε νευρώνα του κρυμμένου επιπέδου περνάει σε μια μη-γραμμική συνάρτηση που ονομάζεται συνάρτηση ενεργοποίησης (activation functions - στο σχήμα φαίνεται η σιγμοειδής συνάρτηση) της οποίας η έξοδος είναι το σήμα μεταφοράς προς το επίπεδο εξόδου (ή επόμενων κρυφών επιπέδων) μέσω των συνάψεων

που συνδέουν τα δυο επίπεδα αυτά. Οι πιο ευρέως χρησιμοποιούμενες συναρτήσεις ενεργοποίησης είναι η σιγμοειδής, η υπερβολική εφάπτομένη και η ανορθωμένη γραμμική συνάρτηση ράμπας (Rectified Linear Unit – ReLU).

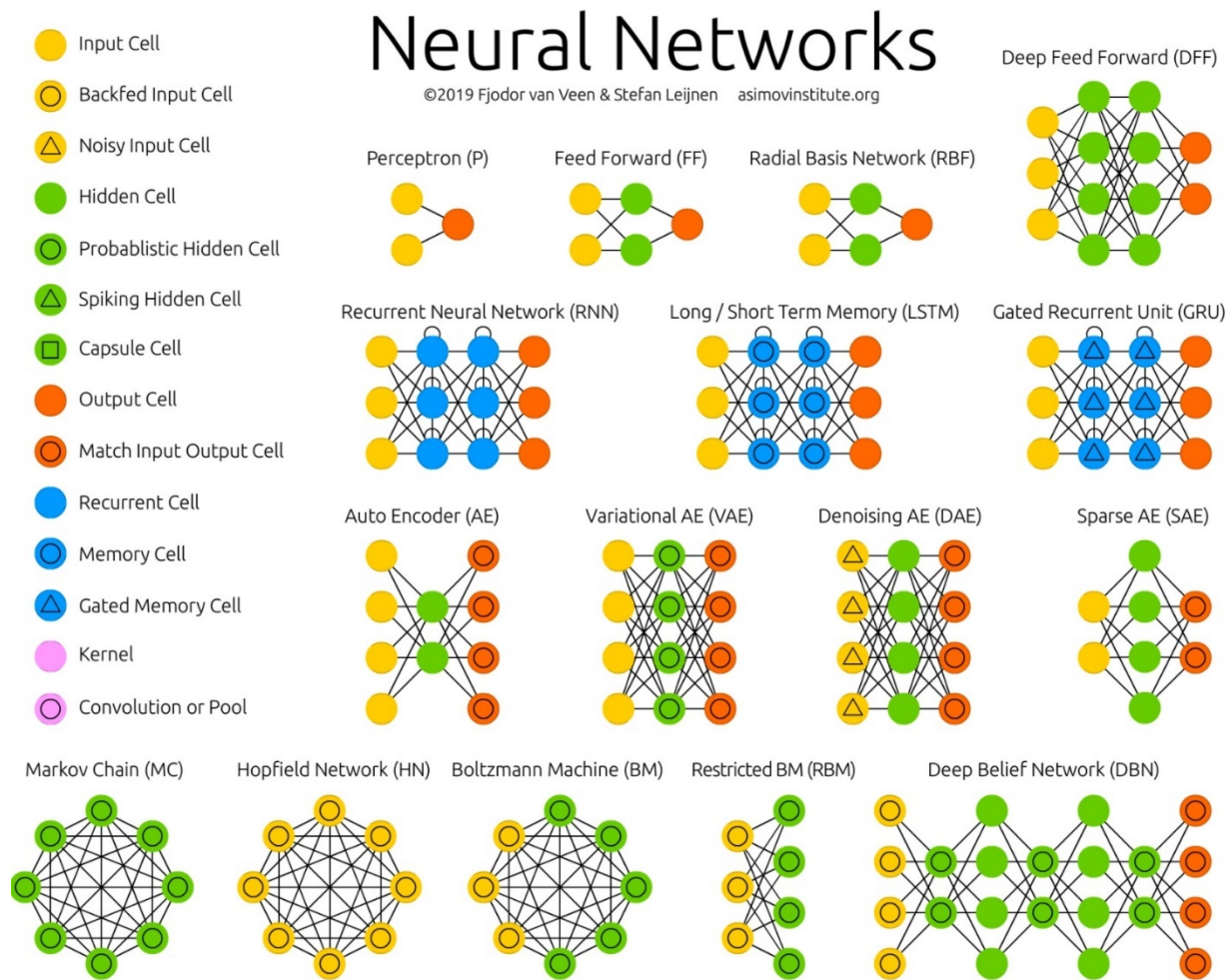
Οι συνάψεις έχουν διαφορετικά βάρη κάτι το οποίο αντικατοπτρίζεται σχηματικά από το διαφορετικό πάχος του κάθε βέλους (αριστερό δικτύωμα στο σχήμα 10). Τα συναπτικά βάρη θα αναπροσαρμοστούν κατά τη διάρκεια της εκπαίδευσης του ΤΝΔ. Τα ΤΝΔ χαρακτηρίζονται από τον αριθμό των επιπέδων τους, τον αριθμό των νευρώνων ανά επίπεδο (μπορεί να διαφέρει ανά επίπεδο) και τις συναρτήσεις ενεργοποίησης που εκτελούν οι νευρώνες του (μπορεί να είναι διαφορετικές ανά επίπεδο). Αν το ΤΝΔ έχει πάνω από ένα κρυμμένο επίπεδο, τότε η ορολογία του κλάδου το ορίζει ως βαθύ νευρωνικό δίκτυο (Deep Neural Network – DNN).

Για να εκτελέσει το σκοπό του το ΤΝΔ θα πρέπει τα βάρη των συνάψεών του να τροποποιηθούν με τέτοιο τρόπο ώστε να ελαχιστοποιηθεί μια συνάρτηση σφάλματος, η οποία μετράει το σφάλμα μεταξύ της εξόδου του ΤΝΔ και γνωστών στόχων (targets), μια διαδικασία που ονομάζεται μάθηση του ΤΝΔ. Η κατηγορία της μάθησης που μόλις περιγράφηκε είναι της ελεγχόμενης μάθησης (supervised learning), η οποία θα χρησιμοποιηθεί στη προσομοίωση της παρούσας πτυχιακής. Στον χώρο των υλικών, η μάθηση των βαρών και ο προγραμματισμός μπορεί να γίνει στην ίδια διάταξη, μια προσέγγιση που θέλει εξωτερικά κυκλώματα ελέγχου και μνήμες και ονομάζεται απευθείας εκπαίδευση (on-line training ή in situ training). Η αντίθετη περίπτωση είναι να γίνει η εκπαίδευση των βαρών σε άλλο σύστημα (λογισμικό) και να μεταφέρονται στο κύκλωμα. Η προσέγγιση εκπαίδευσης αυτή ονομάζεται εκτός κυκλώματος (offline training ή ex situ training).

Υπάρχουν άλλες δυο κατηγορίες μάθησης, η μη-ελεγχόμενη μάθηση (unsupervised learning) και η ενισχυτική μάθηση (reinforcement learning). Στην μη-ελεγχόμενη μάθηση, τα δεδομένα εισόδου δεν έχουν κάποιο επιθυμητό αποτέλεσμα και το ΤΝΔ καλείται να αυτό-οργανώσει τα δεδομένα σε κατηγορίες με βάση τα χαρακτηριστικά των δεδομένων και κανόνες που δημιουργεί το ΤΝΔ από μόνο του. Στην ενισχυτική μάθηση, το ΤΝΔ μαθαίνει μια στρατηγική ενεργειών για μια δεδομένη παρατήρηση. Χρησιμοποιείται σε καταστάσεις όπου υπάρχει ένα πράκτορας (agent) που λειτουργεί σε ένα περιβάλλον. Ένα χαρακτηριστικό παράδειγμα είναι ένα ΤΝΔ που παίζει το παιχνίδι Go ή ένα αυτόνομο όχημα.

Ο πιο διαδεδομένος αλγόριθμος βελτιστοποίησης στην ελεγχόμενη μάθηση είναι αυτός της κλίσης καθόδου (gradient descent), ο οποίος παίρνει τις παραγώγους της συνάρτησης σφάλματος σε συνδυασμό τις συναρτήσεις ενεργοποίησης, μετράει την αλλαγή στο σφάλμα που προκαλείται από αλλαγές στα βάρη των συνάψεων ανά επίπεδο και επιλέγει τη βέλτιστη μεταβολή στα βάρη ώστε να ελαχιστοποιηθεί το ολικό σφάλμα. Η όλη διαδικασία της κλίσης καθόδου ξεκινάει από το επίπεδο εξόδου και προχωράει σε αντίθετη φορά από τη φορά των σημάτων κατά την εκπαίδευση. Επειδή η φορά είναι αντίθετη, η ονομασία αυτής της διαδικασίας ονομάζεται οπισθοδιάδοσης (back propagation).

Ο χώρος των ΤΝΔ αναπτύσσεται ραγδαία και υπάρχει μια πληθώρα αρχιτεκτονικών που είναι αδύνατο να περιγράψουν στην ολότητά τους. Μια πολύ κατατοπιστική παρουσίαση των ΤΝΔ και των αρχιτεκτονικών τους γίνεται από το Asimov Ινστιτούτου [13].

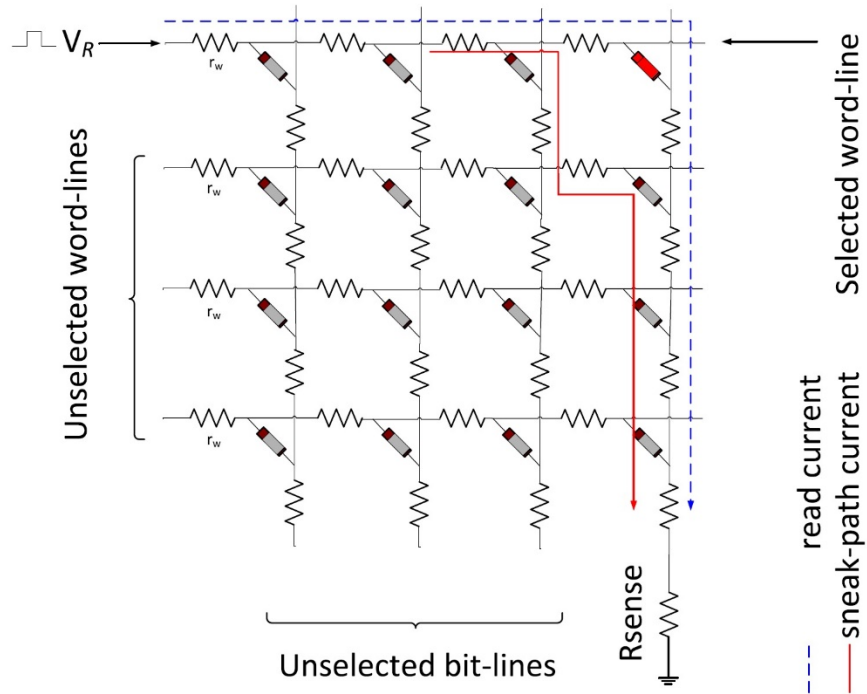


Σχήμα 11: Μικρό δείγμα από τις διάφορες αρχιτεκτονικές ΤΝΔ [13]

2.3 Δικτυώματα μνημών μεταβλητής αντίστασης

Τα δικτυώματα μνημών μεταβλητής αντίστασης είναι τα κυκλώματα που υλοποιούν τις μνήμες αλλά και τις νευρομορφικές εφαρμογές. Η διαφορά στη χρήση ανάμεσα στις μνήμες και των νευρομορφικών κυκλωμάτων είναι στη περιοχή λειτουργίας των μνημών μεταβλητής αντίστασης που χρησιμοποιούνται. Οι μνήμες χρησιμοποιούν τις μνήμες μεταβλητής αντίστασης και στις δυο καταστάσεις τους ώστε να αποθηκεύουν τις πληροφορίες δυαδικά (0 - κατάσταση υψηλής αντίστασης, 1 - κατάσταση χαμηλής αντίστασης). Για τις νευρομορφικές εφαρμογές όμως οι μνήμες μεταβλητής αντίστασης χρησιμοποιούνται αποκλειστικά στη κατάσταση υψηλής αντίστασης, όπου η μεταβολή της αγωγιμότητάς τους είναι σχετικά γραμμική. Η προβλέψιμη αυτή γραμμική συμπεριφορά της αγωγιμότητάς επιτρέπει την εκπαίδευση μέσω κλίσης καθόδου και οπισθοδιάδοσης. Στις νευρομορφικές εφαρμογές το δικτύωμα των μνημών μεταβλητής αντίστασης εκτελεί το εσωτερικό γινόμενο (dot product) των σημάτων εισόδου με τα βάρη των συνάψεων σε πραγματικό χρόνο με αναλογικά σήματα.

Το δικτύωμα μνημών μεταβλητής αντίστασης αποτελείται από ένα σύνολο από παράλληλα κάτω ηλεκτρόδια και ένα σύνολο από άνω παράλληλα ηλεκτρόδια που είναι κάθετα με το κάτω επίπεδο ηλεκτροδίων. Τα δυο επίπεδα ενώνονται στα κάθετα σημεία τομής τους με υλικά που είναι μνήμες μεταβλητής αντίστασης, οι οποίες λειτουργούν σαν συνάψεις.

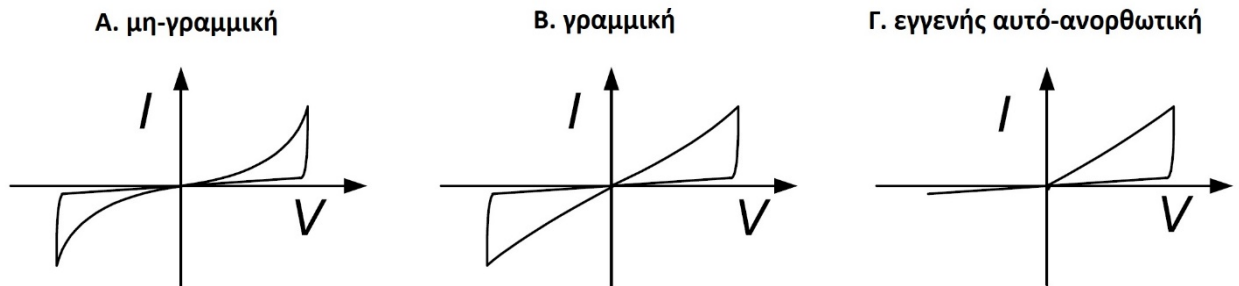


Σχήμα 12: Δικτύωμα με μνήμες μεταβλητής αντίστασης σε χρήση μνήμης με ρεύμα ερπυσμού [14]

Το εγγενές πρόβλημα των δικτυωμάτων με μνήμες μεταβλητής αντίστασης είναι τα ρεύματα ερπυσμού, τα οποία σε λειτουργία μνήμης επηρεάζουν την σωστή ανάγνωση των bit-γραμμών καθώς στη λειτουργία νευρομορφικού δικτύου, επηρεάζουν τη βέλτιστη μέτρηση εξόδου (ρεύματος ή τάση) και έτσι καθυστερούν την εκπαίδευση του ΤΝΔ ή και ακόμη δεν επιτρέπουν καν τη σύγκλιση (convergence). Το ρεύμα ερπυσμού, όπως φαίνεται και στο σχήμα 12, είναι ρεύμα που περνάει από άλλον κάθετο κλάδο στο κλάδο που θέλουμε να μετρήσουμε, κάτι το οποίο είναι λογικό, αφού δεν απομονώνονται ηλεκτρικά οι κάθετοι κλάδοι. Στο δικτύωμα του σχήματος 12, όπως έχει αναφερθεί και στους διάφορους τύπους συνάψεων (σχήμα 8) έχει χρησιμοποιηθεί σύναψη τύπου 1M, δηλαδή μια μνήμη μεταβλητής αντίστασης χωρίς κάποιον επιλογέα. Για μικρές διαστάσεις δικτυωμάτων, όπως 16×16 , τα ρεύματα ερπυσμού δεν δημιουργούν μεγάλο πρόβλημα αλλά όσο μεγαλώνουν οι διαστάσεις του δικτυώματος, χρειάζονται λύσεις απομόνωσης. Επειδή η προσομοίωση του τρίτου κεφαλαίου γίνεται σε δικτύωμα με διαστάσεις 5×4 , τα ρεύματα ερπυσμού δεν επηρεάζουν τα αποτελέσματα του νευρομορφικού δικτύου και χρησιμοποιήθηκε σύναψη του τύπου 1M.

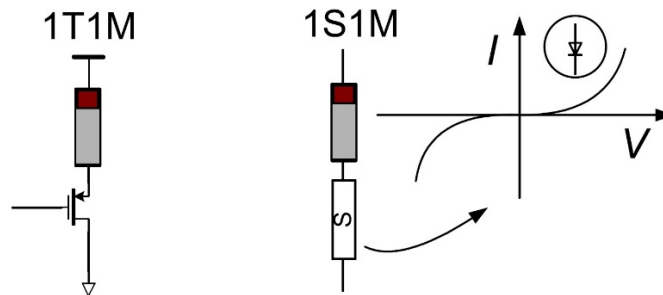
Οι λύσεις για να αντιμετωπιστεί το πρόβλημα του ρεύματος ερπυσμού ακολουθούν δυο κατηγορίες. Η πρώτη κατηγορία αντιμετωπίζει τα ρεύματα ερπυσμού μέσω εγγενών χαρακτηριστικών ειδικών μνημών μεταβλητής αντίστασης. Τα χαρακτηριστικά αυτά είναι τα μη γραμμικά χαρακτηριστικά εναλλαγής (nonlinear switching characteristics) και η αυτό-ανορθωτική (self-rectifying) συμπεριφορά. Στο σχήμα

που ακολουθεί, φαίνονται οι χαρακτηριστικές I-V για τις δυο παραπάνω συμπεριφορές καθώς και η συνηθισμένη γραμμική I-V συμπεριφορά των μνημών μεταβλητής αντίστασης.



Σχήμα 13: I-V χαρακτηριστικές - συμπεριφορές για αντιμετώπιση ρευμάτων ερπυσμού [14]

Η δεύτερη κατηγορία χρησιμοποιεί κάποιο στοιχείο σε σειρά με τη μνήμη μεταβλητής αντίστασης που λειτουργεί σαν επιλογέας και απομονώνει τη μνήμη μεταβλητής αντίστασης ή έναν κάθετο κλάδο του δικτύματος. Το στοιχείο αυτό μπορεί να είναι τρανζίστορ (1T1M) ή δίοδος (1S1M) (σχήμα 14).



Σχήμα 14: Δομές 1T1M και 1S1M ως επιλογείς για την αντιμετώπιση ρευμάτων ερπυσμού [14]

Για μεγάλα δικτυώματα με μνήμες μεταβλητής αντίστασης σε χρήση νευρομορφικών εφαρμογών, τα ρεύματα ερπυσμού αντιμετωπίζονται κυρίως με τη χρήση τρανζίστορ. Για τη χρήση ως μνήμη, χρησιμοποιείται και πάλι επιλογέας τρανζίστορ για την αντιμετώπιση των ρευμάτων ερπυσμού. Τέτοιες μνήμες έχουν παραχθεί και πωλούνται στην αγορά (Dialog Semiconductors, Panasonic) [14].

Η μη χρήση επιλογέα σε δικτυώματα σε χρήση ως μνήμη απαιτεί διάφορες εξειδικευμένες τεχνικές ανάγνωσης με τάσεις ανάγνωσης (V_{read}) $V_{read}/3$ ή $V_{read}/2$ σε συνδυασμό με μνήμες μεταβλητής αντίστασης με εγγενή χαρακτηριστικά που αντιμετωπίζουν τα ρεύματα ερπυσμού.

2.4 Αναφορές 2^{ου} κεφαλαίου

- [1] L.P. Shi · K.J. Yi · K. Ramanathan · R. Zhao · N. Ning ·D. Ding · T.C. Chong, “Artificial cognitive memory—changing from density driven to functionality driven”, Applied Physics A, Material Science & Processing, 102: 865-875, 2011.
- [2] Z. Q. Wang, H. Y. Xu, X. H. Li, H. Yu, Y. C. Liu, and X. J. Zhu, “”, Advanced Functional Materials, Vol. 22, Issue 13, 2759-2765, 2012.
- [3] S.-J. Choi, G.-B. Kim, K. Lee, K.-H. Kim, W.-Y. Yang, S. Cho, H.-J. Bae, D.-S. Seo, S.-I. Kim, and K.-J. Lee, “Synaptic behaviors of a single metal–oxide–metal resistive device”, Springer, Applied Physics A: Materials Science & Processing, Vol. 102, Issue 4, 1019-1025, 2011.
- [4] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, “Nanoscale Memristor Device as Synapse in Neuromorphic Systems”, American Chemical Society, Nano Letters, 10, 1297-1301, 2010.
- [5] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, “Efficient and self-adaptive in-situ learning in multilayer memristor neural networks”, Nature Communications, 9:2385, 2018
- [6] M. Prezioso, F. Merrih-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, “Training and operation of an integrated neuromorphic network based on metal-oxide memristors”, Nature, 521, 61-64, 2015.
- [7] M. Hu, Y. Chen, J. J. Yang, Y. Wang, and H. H. Li, “A Compact Memristor-Based Dynamic Synapse for Spiking Neural Networks”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol 36, Issue 8, 2016.
- [8] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S. P. Wong, and H. Qian, “Face classification using electronic synapses”, Nature Communications, Art. Num. 15199, 2017.
- [9] D. Soudry, D. D. Castro. A. Gal, A. Kolodny, and S. Kvatinsky, “Memristor-Based Multilayer Neural Networks with Online Gradient Descent Training”, IEEE Transactions on Neural Networks and Learning Systems, Vol 26, Issue 10, 2015.
- [10] H. Kim, M. P. Sah, C. Yang, T. Roska, and L. O. Chua, “Memristor Bridge Synapses”, Proceedings of the IEEE, Vol 100, Issue 6, 2012.
- [11] Y.-S. Kim, and K.-S. Min, “Synaptic weighting circuits for Cellular Neural Networks”, 2012 13th International Workshop on Cellular Nanoscale Networks and their Applications, IEEE, 2012.
- [12] C. Yakopcic, and T. M. Taha, “Energy efficient perceptron pattern recognition using segmented memristor crossbar arrays”, The 2013 International Joint Conference on Neural Networks (IJCNN), IEEE, 2013.
- [13] The neural network zoo, Asimov Institute website, 25/2/21, <https://www.asimovinstitute.org/neural-network-zoo/>

[14] Y. Gao, O. Kavehei, S. F. Al-Sarawia, D. C. Ranasinghe, and D. Abbott, "Read operation performance of large selectorless cross-point array with self-rectifying memristive device", Elsevier, Integration the VLSI journal, Vol 54, pages 56-64, 2016.

Κεφάλαιο 3: Προσομοίωση νευρωνικού δικτύου μέσω δικτυώματος μνημών μεταβλητής αντίστασης

3.1 Γενική περιγραφή του προβλήματος και δομικά στοιχεία της προσομοίωσης

Τα ΤΝΔ χρησιμοποιούνται για διάφορες εφαρμογές όπως αναγνώριση εικόνων ή/και μοτίβων καθώς και πρόβλεψη τιμών σε αριθμητικά προβλήματα όπως επενδύσεις και πωλήσεις. Στη παρούσα πτυχιακή εργασία θα χρησιμοποιηθεί ένα ρηχό (shallow) νευρωνικό δίκτυο (χωρίς κρυμμένο επίπεδο) για την αναγνώριση και κατηγοριοποίηση συγκεκριμένων μοτίβων σε εικόνες των 16 pixels.

Η προσομοίωση για την εκπαίδευση του δικτυώματος μνημών μεταβλητής αντίστασης ως κυκλώματος κατηγοριοποίησης εικόνων πραγματοποιήθηκε μέσω της εφαρμογής LTSpice της Analog Devices. Το μοντέλο της μνήμης μεταβλητής αντίστασης που χρησιμοποιήθηκε είναι του Yakorcic [1], το οποίο παραμετροποιήθηκε ώστε να προσεγγίζει τη συμπεριφορά μιας μνήμης μεταβλητής αντίστασης διοξειδίου του Τιτανίου (TiO_2). Το δικτύωμα μνημών μεταβλητής αντίστασης που χρησιμοποιήθηκε έχει διαστάσεις 5×4 , δηλαδή 5 εισόδους (τέσσερις των δεδομένων, μια το bias) και τέσσερις εξόδους, όπου η κάθε έξοδος αντιπροσωπεύει μια κατηγορία. Οι 20 διαφορετικές μνήμες μεταβλητής αντίστασης παίζουν τον ρόλο των συνάψεων στο ΤΝΔ και η αγωγιμότητά τους το βάρος της κάθε σύναψης. Τα δεδομένα απαρτίζονται από 8 εικόνες των 16 εικονοστοιχείων, με 4 κατηγορίες (δυο εικόνες ανά κατηγορία). Η εκπαίδευση του δικτύου έγινε μέσω της οπισθοδιάδοσης. Στις επόμενες παραγράφους γίνεται αναλυτική περιγραφή του κάθε σημείου της προσομοίωσης.

3.2 Το LTSpice

Το LTSpice είναι μια εφαρμογή για υψηλής απόδοσης SPICE προσομοιώσεις, σχεδιασμού κυκλωμάτων και δημιουργίας κυματομορφών με πολλές επιλογές επιπρόσθετων δυνατοτήτων και μοντέλων για τη προσομοίωση αναλογικών κυκλωμάτων [2].

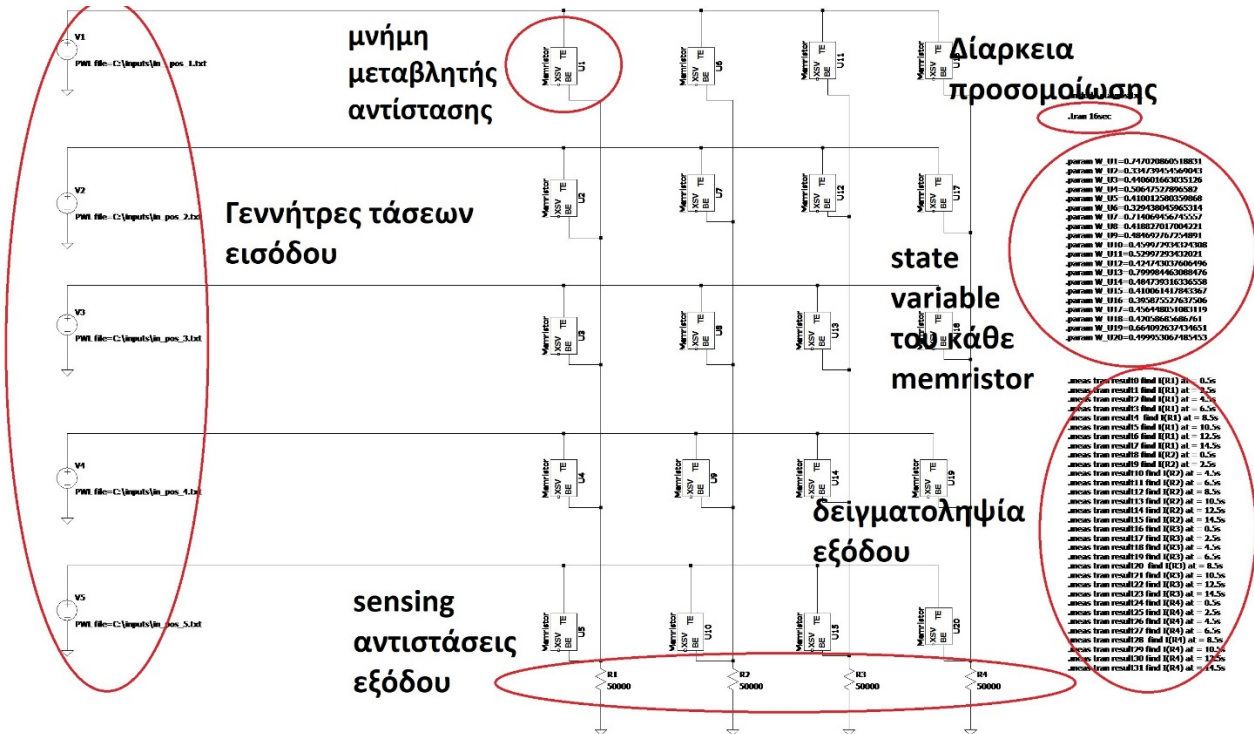
Στο LTSpice, δημιουργήθηκε ένα σχηματικό (αρχείο με επέκταση .asc) στο οποίο τοποθετήθηκαν οι γεννήτριες των τάσεων εισόδου (κωδικοποιημένες εικόνες σε τάσεις), οι μνήμες μεταβλητής αντίστασης, οι αντιστάσεις εξόδου για τις μετρήσεις των σημάτων εξόδου καθώς και εντολές της εφαρμογής (directives) (εικόνα 1).

Οι LTSpice εντολές που χρειάστηκαν για τη προσομοίωση ήταν οι παρακάτω τρεις:

- .tran 16s, οδηγία για τη χρονική διάρκεια της προσομοίωσης
- .param W_U1 μέχρι .param W_U20, παράμετροι ώστε να προγραμματιστούν οι μεταβλητές καταστάσεων της κάθε μνήμης μεταβλητής αντίστασης, αρχικά καθώς και οι μετέπειτα τιμές που προέκυπταν κατά την εκπαίδευση του κυκλώματος
- .means tran result0 find I(R1) at = 0.5s μέχρι .means tran result7 find I(R1) at = 14.5s για τη καταγραφή του ρεύματος εξόδου του πρώτου νευρώνα ανά 2 δευτερόλεπτα για όλα τα σήματα εισόδου. Τα δεδομένα αυτά τα εξάγει το LTSpice σε log αρχείο

.means tran result8 find I(R2) at = 0.5s μέχρι .means tran result15 find I(R2) at = 14.5s. Το ίδιο με παραπάνω αλλά για το ρεύμα εξόδου του δεύτερου νευρώνα εξόδου.
 .means tran result16 find I(R3) at = 0.5s μέχρι .means tran result23 find I(R3) at = 14.5s. Το ίδιο με παραπάνω αλλά για το ρεύμα εξόδου του τρίτου νευρώνα εξόδου.
 .means tran result24 find I(R4) at = 0.5s μέχρι .means tran result31 find I(R4) at = 14.5s. Το ίδιο με παραπάνω αλλά για το ρεύμα εξόδου του τέταρτου νευρώνα εξόδου.

Οι εντολές αυτές φαίνονται στη δεξιά πλευρά του παρακάτω σχήματος.



Σχήμα 15: LTSpice δικτύωμα μνημών μεταβλητής αντίστασης και τα μέρη του

Το χρησιμοποιούμενο μοντέλο της μνήμης μεταβλητής αντίστασης εισήχθη στο LTSpice ως υπό-κύκλωμα σαν αρχείο με επέκταση .sub. Μέσα στον κώδικα του υπό-κυκλώματος (αναλυτικά στην επόμενη παράγραφο) μπήκε σαν εξωτερική παράμετρος η μεταβλητή κατάσταση ώστε να μπορεί να παραμετροποιηθεί η αγωγιμότητα της κάθε μνήμης μεταβλητής αντίστασης με αρχικές τιμές καθώς και με τις τιμές από την εκπαίδευση οπισθοδιάδοσης μετά από κάθε κύκλο εκπαίδευσης (epoch). Στο σχηματικό, τα στοιχεία (components) των μνημών μεταβλητής αντίστασης τοποθετήθηκαν αφού πρώτα δημιουργήθηκε το κατάλληλο σύμβολο (αρχείο με επέκταση .asy).

3.3 Χρησιμοποιούμενο μοντέλο μνήμης μεταβλητής αντίστασης και τα χαρακτηριστικά του

Το μοντέλο της μνήμης μεταβλητής αντίστασης που επιλέχθηκε να χρησιμοποιηθεί στην προσομοίωση είναι αυτό του Yakorčic [1], το οποίο βασίζεται στα τρία παρακάτω χαρακτηριστικά των μνημών μεταβλητής αντίστασης:

1. ενός κόμβου μετάλλου-μονωτή-μετάλλου (MIM junction)
2. ενός κατωφλίου τάσεως για την κίνηση της μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης
3. μη γραμμική ταχύτητα των ιόντων στα φυσικά όρια της μνήμης μεταβλητής αντίστασης

Η πρώτη εξίσωση, που περιγράφει την σχέση ρεύματος-τάσης (I-V) φαίνεται παρακάτω:

$$I(t) = \begin{cases} \alpha_1 * x(t) * \sinh(b * V(t)), & V(t) \geq 0 \\ \alpha_2 * x(t) * \sinh(b * V(t)), & V(t) < 0 \end{cases} \quad (14)$$

Η εξίσωση (14) βασίζεται σε υπερβολικά ημίτονα ώστε να αποτυπωθεί η MIM συμπεριφορά της μνήμης μεταβλητής αντίστασης. Οι τρεις παράμετροι, α_1 , α_2 και b χρησιμοποιούνται για τη προσαρμογή της σχέσης (14) σε πραγματικά δεδομένα μιας μνήμης μεταβλητής αντίστασης. Οι διαφορετικοί παράμετροι α_1 και α_2 χρησιμοποιούνται για την αποτύπωση διαφορετικής αγωγιμότητας στις διαφορετικές πολικότητες. Η παράμετρος b χρησιμοποιείται για να ελεγχθεί η ένταση της σχέσης κατωφλίου μεταξύ του μεγέθους της τάσης εισόδου και της αγωγιμότητας της μνήμης μεταβλητής αντίστασης. Επιπρόσθετα, η σχέση I-V (14) εξαρτάται και από τη μεταβλητή κατάσταση ($x(t)$), η οποία αποτυπώνει την αλλαγή της αγωγιμότητας με βάση την δυναμική του κάθε υλικού/αρχιτεκτονικής της μνήμης μεταβλητής αντίστασης. Το εύρος τιμών της μεταβλητής κατάστασης στο μοντέλο αυτό είναι από μηδέν μέχρι ένα.

Η συμπεριφορά της μεταβλητής κατάστασης της προς μοντελοποίησης μνήμης μεταβλητής αντίστασης περιγράφεται από δυο σχέσεις, τις $g(V(t))$ και $f(x(t))$.

Η σχέση $g(V(t))$ αποδίδει τη συμπεριφορά τάσης κατωφλίου της μνήμης μεταβλητής αντίστασης, όπου αν υπερβεί τη τάση αυτή η τάση στους ακροδέκτες της διάταξης, υπάρχει μεταβολή στη μεταβλητή κατάσταση και κατ' επέκταση της αγωγιμότητας/αντίστασης της μνήμης μεταβλητής αντίστασης.

$$g(V(t)) = \begin{cases} A_p * (e^{V(t)} - e^{V_p}), & V(t) > V_p \\ -A_n * (e^{-V(t)} - e^{V_n}), & V(t) < -V_n \\ 0, & V_n \leq V(t) \leq V_p \end{cases} \quad (15)$$

Η παραπάνω σχέση δίνει τη δυνατότητα να αποτυπωθεί η διαφορετική τάση κατωφλίου στις δυο διαφορετικές πολικότητες της τάσης εισόδου. Τα μεγέθη V_p και V_n είναι αντίστοιχα οι τάσεις κατωφλίου για θετικές και αρνητικές τάσεις. Τα εκθετικά τους στη παραπάνω σχέση εξασφαλίζουν ότι η τιμή του $g(V(t))$ είναι μηδέν στο σημείο που ξεπερνιέται η τιμή του κατωφλίου. Τα μεγέθη A_p και A_n καθορίζουν το μέγεθος των εκθετικών και στη φυσική διάσταση το πόσο γρήγορα η μεταβλητή κατάσταση αλλάζει μόλις ξεπεραστεί η τάση κατωφλίου.

Η σχέση $f(x(t))$ είναι η δεύτερη σχέση που μοντελοποιεί τη συμπεριφορά της μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης και περιγράφει τη μη γραμμική κίνηση των ιόντων (nonlinear dopant drift) καθώς γίνεται όλο και πιο δύσκολη η αλλαγή της μεταβλητής κατάστασης κοντά στα φυσικά όρια της διάταξης.

$$f(x(t)) = \begin{cases} e^{-a_p * (x-x_p)} * w_p(x, x_p), & x \geq x_p \\ 1, & x < x_p \end{cases} \quad (16)$$

$$f(x(t)) = \begin{cases} e^{a_n * (x+x_n-1)} * w_n(x, x_n), & x \leq 1 - x_n \\ 1, & x > 1 - x_n \end{cases} \quad (17)$$

Και σε αυτή τη περίπτωση, το μοντέλο δίνει την ευελιξία για διαφορετική συμπεριφορά ανάλογα με τη πολικότητα της τάσης εισόδου.

Η παράμετρος “ η ” (δεν φαίνεται στις παραπάνω σχέσεις αλλά παρακάτω) του μοντέλου περιγράφει τη κατεύθυνση της μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης σε σχέση με τη πολικότητα της τάσης εισόδου. Αν $\eta=1$, τότε μια θετική τάση εισόδου πάνω από V_p θα αυξήσει τη μεταβλητή κατάσταση. Αντίστοιχη λογική για αρνητικές τάσεις εισόδου. Αν $\eta * V(t) > 0$, τότε η πρώτη σχέση $f(x(t))$ περιγράφει τη κίνηση της μεταβλητής κατάστασης. Σε κάθε άλλη περίπτωση, η δεύτερη σχέση $f(x(t))$ περιγράφει τη κίνηση της μεταβλητής κατάστασης.

Η σχέση $f(x(t))$ χωρίζει την κίνηση της μεταβλητής κατάστασης σε δυο περιοχές ανάλογα με τη κατάσταση της μνήμης μεταβλητής αντίστασης. Οι μεταβλητές x_p και x_n καθορίζουν τις τάσεις κατωφλίου, θετική και αρνητική αντίστοιχα, όπου η μεταβολή της μεταβλητής κατάστασης είναι σταθερή μέχρι εκεί. Πάνω από τις δυο παραπάνω μεταβλητές, η μεταβολή της μεταβλητής κατάστασης περιορίζεται από μια εκθετική σχέση, που μειώνεται με ρυθμό είτε a_p είτε a_n (θετικές τάσεις και αρνητικές αντίστοιχα).

Επιπρόσθετα χρησιμοποιούνται και δυο συναρτήσεις παραθύρου, για να εξασφαλιστούν οι συνθήκες $\eta f(x)$ να ισούται με μηδέν όταν $x(t)=1$ (μέσω της w_p) και $x(t)$ να μην γίνει αρνητική (μέσω της w_n) όταν η ροή του ρεύματος αντιστραφεί.

$$w_p(x, x_p) = \frac{x_p - x}{1 - x_p} + 1 \quad (18)$$

$$w_n(x, x_n) = \frac{x}{1 - x_n} \quad (19)$$

Η τελευταία εξίσωση, που γενικεύει τη μοντελοποίηση της μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης είναι:

$$\frac{dx}{dt} = \eta * g(V(t)) * f(x(t)) \quad (20)$$

Όπου “ η ” η ίδια μεταβλητή που περιγράφηκε πριν, η οποία καθορίζει τη κατεύθυνση της μεταβολής της μεταβλητής κατάστασης.

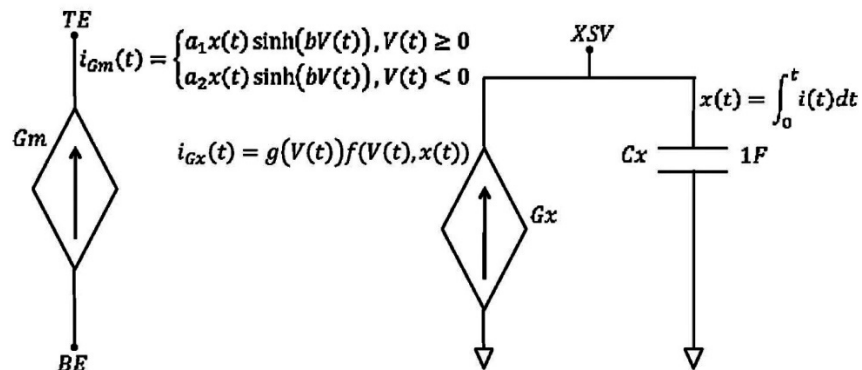
Από τα παραπάνω, το μοντέλο της μνήμης μεταβλητής αντίστασης του Yakovlevic έχει δώδεκα παραμέτρους για τη προσαρμογή των σχέσεων στα πραγματικά δεδομένα μιας διάταξης.

Ο πίνακας που ακολουθεί παραθέτει τις παραμέτρους του μοντέλου και τη σχέση τους με τη φυσική συμπεριφορά των υλικών.

Παράμετρος μοντέλου	Σχέση με φυσική συμπεριφορά
α_1 και α_2	Αυτές οι παράμετροι σχετίζονται στενά με το πάχος του διηλεκτρικού στη διάταξη της μνήμης μεταβλητής αντίστασης. Περισσότερα ηλεκτρόνια μπορούν κινηθούν μέσα από ένα λεπτότερο φραγμό κάτι που οδηγεί σε αύξηση της αγωγιμότητας
b	Η παράμετρος b καθορίζει πόση καμπυλότητα θα υπάρχει στη χαρακτηριστική καμπύλη I-V σε σχέση με την εφαρμοσμένη τάση. Αυτό σχετίζεται με το πόση αγωγιμότητα οφείλεται στην ωμική αντίσταση και πόση στο φραγμό σήραγγας ηλεκτρονίων
A_p και A_n	Οι παράμετροι αυτοί ελέγχουν την ταχύτητα της κίνησης των ιόντων και κατ' επέκταση του αγωγίμου δρόμου. Αυτό μπορεί να σχετίζεται με το χρησιμοποιούμενο υλικό στο διηλεκτρικό αφού τα κενά οξυγόνων έχουν διαφορετική κινητικότητα που εξαρτάται σε ποιο οξείδιο μετάλλου εμπεριέχονται
V_p και V_n	Οι παράμετροι αυτοί είναι οι τάσεις κατωφλίου. Αυτές οι τάσεις μπορεί να σχετίζονται με τον αριθμό των κενών θέσεων οξυγόνου στις μνήμες μεταβλητής αντίστασης στην αρχική τους κατάσταση. Μια μνήμη μεταβλητής αντίστασης με περισσότερες κενές θέσεις οξυγόνου θα πρέπει να έχει και μεγαλύτερη άντληση ρεύματος που μπορεί να οδηγήσει σε χαμηλότερο κατώφλι μεταγωγής αν υποθεθεί ότι η μεταγωγή βασίζεται στο ολικό εφαρμοζόμενο φορτίο.
α_p, α_n, x_p και x_n	Αυτές οι παράμετροι καθορίζουν πότε η κίνηση της μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης δεν είναι πλέον γραμμική και καθορίζουν τον βαθμό που η κίνηση αποσβένει. Αυτό μπορεί να σχετίζεται με το μέταλλο των ηλεκτροδίων εκατέρωθεν του διηλεκτρικού φιλμ αφού τα επιλεγμένα μέταλλα μπορούν να αντιδράσουν με τις κενές θέσεις οξυγόνων διαφορετικά.

Πίνακας 5: Παράμετροι μοντέλου και η σχέση τους με τη φυσική συμπεριφορά των υλικών [1]

Η μεταφορά των παραπάνω σχέσεων σε ανάλογα κυκλώματα και κατ' επέκταση σε κώδικα Spice έγινε με τα ακόλουθα κυκλώματα:



Σχήμα 16: Ισοδύναμα ηλεκτρονικά κυκλώματα για το Yakorcic μοντέλο [1]

Η πηγή ρεύματος G_m υλοποιεί τις σχέσεις ρεύματος τάσης (14). Οι ακροδέκτες TE και BE αντιπροσωπεύουν τους πάνω και κάτω ακροδέκτες αντίστοιχα και θα χρησιμοποιηθούν στο σχηματικό του LTSpice για τη διασύνδεση της μνήμης μεταβλητής κατάστασης μέσα στο κύκλωμα.

Για τον καθορισμό της τιμής της μεταβλητής κατάστασης της μνήμης μεταβλητής κατάστασης χρησιμοποιείται άλλη μια πηγή ρεύματος, G_x , της οποίας το ρεύμα καθορίζεται από το γινόμενο των εξισώσεων $g(V(t))$ και $f(x(t))$. Η χρήση του πυκνωτή γίνεται για την ολοκλήρωση του ρεύματος της πηγής G_x ώστε να βρεθεί η τιμή της μεταβλητής κατάστασης $x(t)$. Η μεταβλητή κατάσταση παίρνει τιμές από 0 μέχρι 1. Επιπρόσθετα, το κύκλωμα έχει έναν ακροδέκτη XSV, ο οποίος επιτρέπει την απευθείας μέτρηση και παρακολούθηση της μεταβλητής κατάστασης.

Ο Spice κώδικας που χρησιμοποιήθηκε στη παρούσα πτυχιακή εργασία είναι από τη δημοσίευση [1] και ακολουθεί όλη την παραπάνω λογική. Ο αναλυτικός κώδικας φαίνεται στο παράρτημα 1.

Στον παραπάνω κώδικα φαίνονται οι παράμετροι προσαρμογής που χρησιμοποιήθηκαν για την προσομοίωση της παρούσας εργασίας.

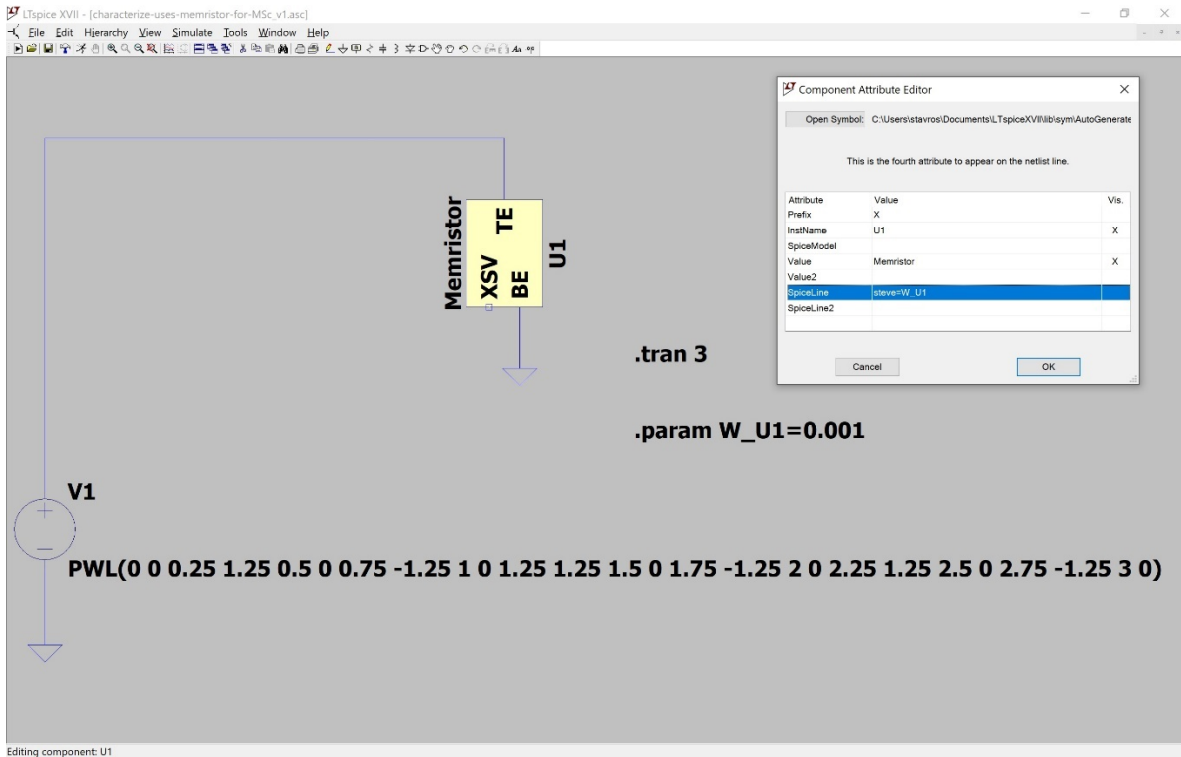
Παράμετρος	Τιμή
α_1	0.17
α_2	0.17
b	0.05
V_p	0.6
V_n	0.5
A_p	4000
A_n	4000
X_p	0.5
X_n	0.5
a_p	1
a_n	5
h	1
x_0	steve (μεταβλητή από χρήστη)

Πίνακας 6: Παράμετροι μοντέλου μνήμης μεταβλητής αντίστασης

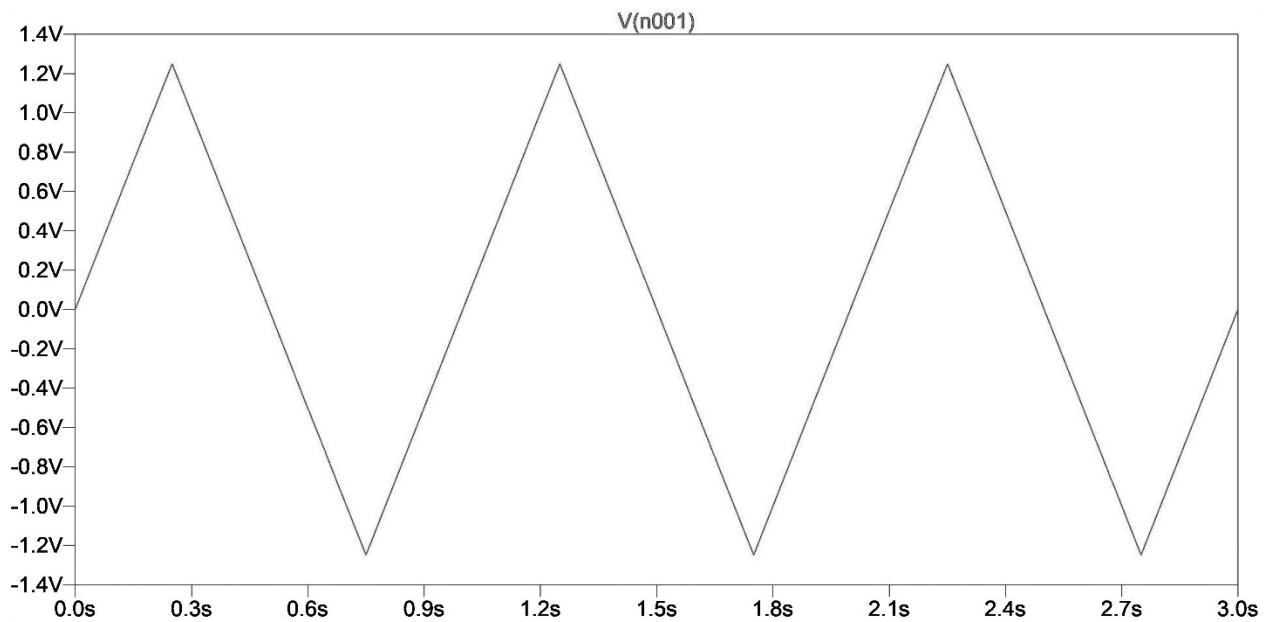
Πρέπει να σημειωθεί ότι η τιμή της τάσεως κατωφλίου στη θετική πόλωση είναι 0.6V δηλαδή όποια θετική τάση εφαρμοστεί στον ακροδέκτη TE μικρότερη του V_p , η μνήμη μεταβλητής αντίστασης θα παραμένει στην ίδια κατάσταση αποκοπής.

Επίσης σημαντική σημείωση είναι η παραμετροποίηση της μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης με τη χρήση της εντολής .params $x_0=\{steve\}$, όπου steve μια μεταβλητή που θα καθορίζει την τιμή της ο χρήστης της εφαρμογής μέσα από SpiceLine τιμή μέσα στο στοιχείο της κάθε μνήμης μεταβλητής αντίστασης. Έτσι θα μπορεί ο χρήστης της προσομοίωσης να παραμετροποιεί μέσω αρχείων παραμετροποίησης (config files) ή LTSpice εντολές τις τιμές των μεταβλητών παραμέτρων όλων των μνημών μεταβλητής αντίστασης στο κύκλωμα.

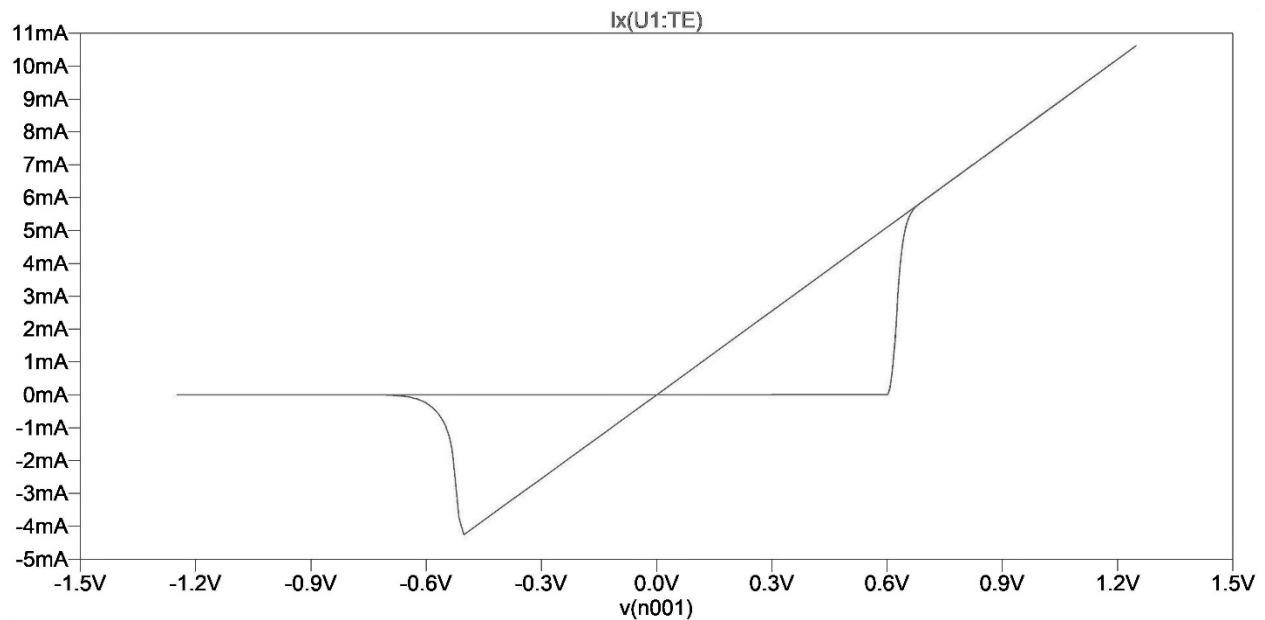
Στα παρακάτω σχήματα φαίνονται το κύκλωμα που χρησιμοποιήθηκε στο LTSpice για να προσδιοριστεί η εξίσωση αγωγιμότητας – μεταβλητής κατάστασης της μνήμης μεταβλητής αντίστασης στην κατάσταση υψηλής αντίστασης, η τάση εισόδου και η χαρακτηριστική I-V της παραπάνω μνήμης.



Σχήμα 17: LTSpice κύκλωμα για χαρακτηριστική I-V της μνήμης μεταβλητής αντίστασης



Σχήμα 18: : Τάση εισόδου της μνήμης μεταβλητής αντίστασης για παραγωγή I-V



Σχήμα 19: Χαρακτηριστική I-V της μνήμης μεταβλητής αντίστασης της προσομοίωσης ($x_0=0.001$)

Στο σχήμα 17 φαίνεται το απλό κύκλωμα που δημιουργήθηκε για να παραχθεί η χαρακτηριστική I-V της μνήμης μεταβλητής αντίστασης για μεταβλητές κατάστασης x_0 από μηδέν μέχρι ένα. Επίσης φαίνονται και τα χαρακτηριστικά του στοιχείου της μνήμης μεταβλητής αντίστασης (memristor component properties) και πως συνδέονται οι παράμετροι μέσα στον κώδικα της μνήμης μεταβλητής αντίστασης (.param $x_0=\{steve\}$), η LTSpice εντολή (.param $W_U1=0.001$) καθώς και το SpiceLine χαρακτηριστικό (attribute) μέσα στο στοιχείο της μνήμης ($steve=W_U1$). Η τιμή 0.001 είναι η τιμή που δόθηκε για τη μεταβλητή κατάσταση της μνήμης μεταβλητής αντίστασης U1. Οι επιτρεπτές τιμές είναι από 0 μέχρι 1.

Στο σχήμα 18 φαίνεται ο τριγωνικός παλμός εισόδου μέσω LTSpice Piecewise Linear Function (PWL). Ο παλμός έχει περίοδο ένα δευτερόλεπτο και μέγιστο πλάτος $\pm 1.25V$.

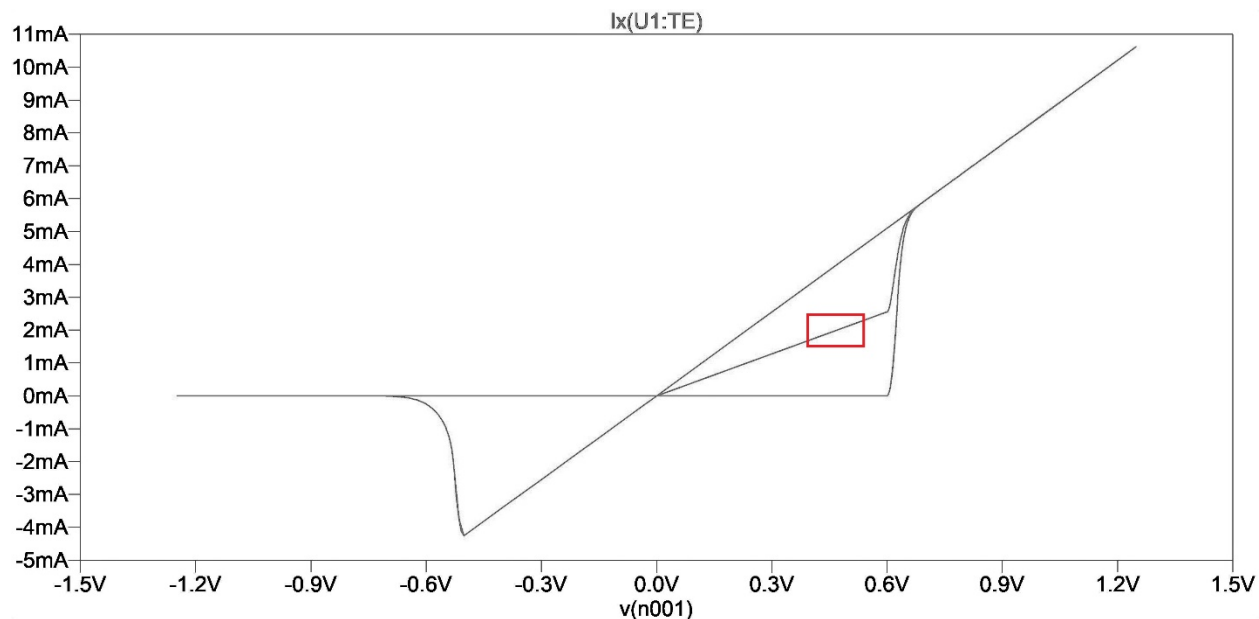
Στο σχήμα 19 φαίνεται η I-V χαρακτηριστική της μνήμης μεταβλητής αντίστασης. Η χρήση της μνήμης μεταβλητής αντίστασης ως σύναψη σε TND γίνεται στη γραμμική περιοχή του, στη κατάσταση υψηλής αντίστασης, για τάσεις εισόδου κάτω από την τάση κατωφλίου.

Στην εκπαίδευση του TND για την εφαρμογή της κατηγοριοποίησης που θα παρουσιαστεί, σε κάθε κύκλο εκπαίδευσης υπολογίζεται η απαιτούμενη μεταβολή της αγωγιμότητας (ΔW) της κάθε μνήμης μεταβλητής αντίστασης για τον επόμενο κύκλο ώστε να επιτευχθεί η βέλτιστη έξοδος (ρεύματα). Επειδή στην εφαρμογή LTSpice ο έλεγχος της αγωγιμότητας της κάθε μνήμης μεταβλητής αντίστασης γίνεται μέσω της μεταβλητής κατάστασης x_0 , πρέπει να κατασκευαστεί ένας πίνακας μετατροπής αγωγιμότητας – μεταβλητής κατάστασης για τη χρησιμοποιούμενη μνήμη μεταβλητής αντίστασης στην κατάσταση αποκοπής. Για το προσδιορισμό της αγωγιμότητας επιλέχθηκε ένα σημείο στο μέσον της γραμμικής περιοχής αγωγιμότητας (HRS), στη χρονική στιγμή $\sim 0.06s$. Με τη παραπάνω τάση εισόδου, η μνήμη μεταβλητής αντίστασης λειτουργεί στην επιθυμητή περιοχή στα χρονικά διαστήματα 0-0.12s, 1-1.12s και 2-2.12 s. Μεταβάλλοντας την μεταβλητή κατάσταση από 0 – 1, έγινε καταγραφή της αγωγιμότητας και δημιουργήθηκε ο παρακάτω πίνακας:

x_0	αγωγιμότητα (S)
0.01	0.0000850085
0.05	0.0004250092
0.1	0.0008500258
0.15	0.00127504
0.2	0.001700055
0.25	0.00212507
0.3	0.002550084
0.35	0.002975098
0.4	0.003400112
0.45	0.003825126
0.5	0.004250139
0.55	0.004675153
0.6	0.00510026
0.65	0.005525184
0.7	0.005950198
0.75	0.006375212
0.8	0.006800227
0.85	0.00722523
0.9	0.007650255
0.95	0.008075269
1	0.008500283

Πίνακας 7: Αγωγιμότητα - μεταβλητή κατάσταση μνήμης μεταβλητής αντίστασης

Ενδεικτικά για $x_0=0.5$, η περιοχή που μετρείται η αγωγιμότητα φαίνεται στο παρακάτω σχήμα:



Σχήμα 20: Μέτρηση αγωγιμότητας για $x_0=0.5$ (περιοχή στο κόκκινο πλαίσιο)

Από τον πίνακα 7, υπολογίστηκε η γραμμή παλινδρόμησης (regression line) η οποία θα χρησιμοποιηθεί αργότερα στην εκπαίδευση του νευρωνικού δικτύου.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-8.26707E-08	1.02607E-06	-0.080570477	0.936626179	-2.23025E-06	2.06491E-06	-2.23025E-06	2.06491E-06
conductance	117.6430818	0.000206508	569676.8125	1.11704E-98	117.6426496	117.643514	117.6426496	117.643514

Πίνακας 8: Γραμμή παλινδρόμησης για μετατροπή αγωγιμότητας σε μεταβλητή κατάστασης

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	7.02974E-10	8.72187E-09	0.080599	0.936603761	-1.75521E-08	1.8958E-08	-1.75521E-08	1.8958E-08
x0	0.008500287	1.49212E-08	569676.8	1.11704E-98	0.008500256	0.008500319	0.008500256	0.008500319

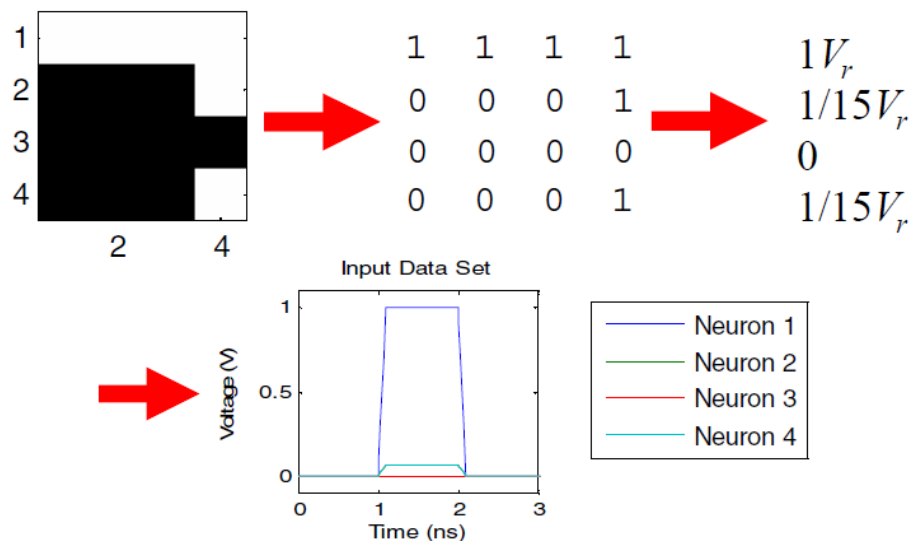
Πίνακας 9: Γραμμή παλινδρόμησης για μετατροπή της μεταβλητής κατάστασης σε αγωγιμότητα

3.4 Ανάλυση του συνόλου δεδομένων (dataset)

Το σύνολο δεδομένων που χρησιμοποιήθηκε για να γίνει κατηγοριοποίηση εικόνων ακολουθεί τη λογική της δημοσίευσης [3].

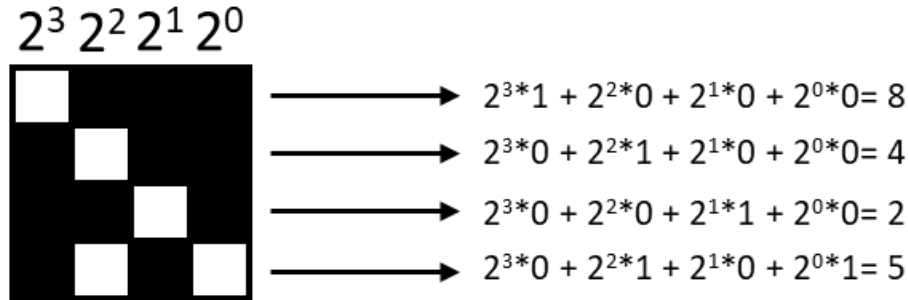
Το σύνολο δεδομένων αποτελείται από τέσσερις κατηγορίες, όπου η κάθε κατηγορία αντιπροσωπεύεται με δυο εικόνες των 16 pixels (4x4). Η κάθε κατηγορία έχει μια γραμμή πολύ πιο λευκή από τις άλλες τρεις. Κάθε pixel μπορεί να είναι είτε μαύρο ή λευκό.

Για να εισαχθούν στο δικτύωμα μνημών μεταβλητής αντίστασης οι εικόνες, κάθε γραμμή της κάθε εικόνας μετατράπηκε σε δεκαέξι διαφορετικές τάσεις, χωρισμένες γραμμικά από τα μηδέν volt μέχρι μια επιλεγμένη τάση (V_r), η οποία είναι κάτω από τη θετική τάση κατωφλίου της χρησιμοποιούμενης μνήμης μεταβλητής αντίστασης.



Σχήμα 21: Δυαδική κωδικοποίηση των εικόνων εισόδου [3]

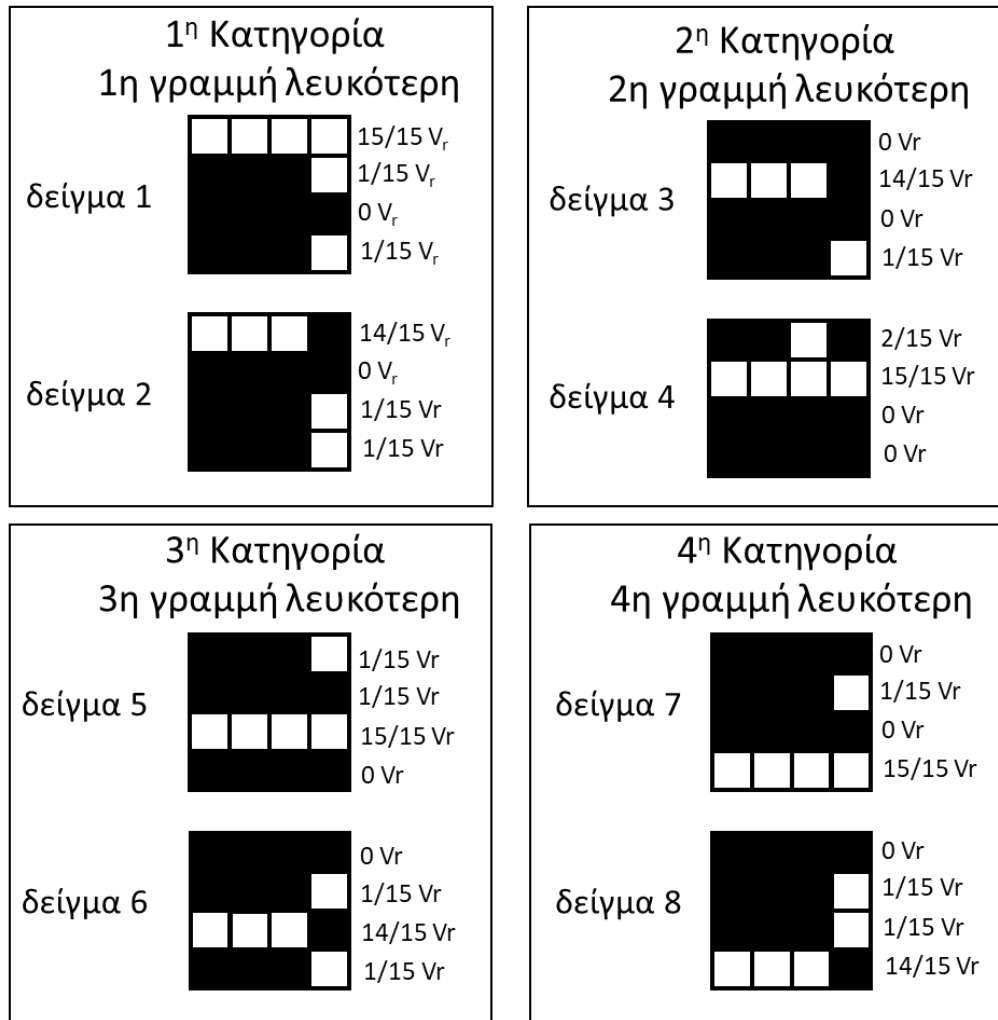
Στο παραπάνω σχήμα φαίνεται μια εικόνα της πρώτης κατηγορίας, αφού η πρώτη γραμμή είναι αρκετά πιο λευκή από τις άλλες τρεις. Το αποτέλεσμα της δυαδικής κωδικοποίησης φαίνεται δεξιά από την 4x4 εικόνα. Η πρώτη γραμμή είναι $15/15 \cdot V_r$, δηλαδή όλα λευκά pixels, $(1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0) \cdot V_r$. Η δεύτερη και η τέταρτη γραμμή κωδικοποιούνται ως $1/15 \cdot V_r$, δηλαδή $(0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0) \cdot V_r$. Η τρίτη γραμμή κωδικοποιείται ως $(0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0) \cdot V_r = 0V$, δηλαδή όλα μαύρα εικονοστοιχεία. Επίσης στο σχήμα 22 υπάρχει γράφημα με τις τάσεις εισόδου σε ένα δικτύωμα με τέσσερις νευρώνες εισόδου. Σε αυτήν την αναπαράσταση, η τάση V_r είναι 1 volt. Η παρακάτω εικόνα έχει άλλη μια, αντίστοιχης λογικής, αναπαράσταση της δυαδικής κωδικοποίησης.



Σχήμα 22: Παράδειγμα δυαδικής κωδικοποίησης τάσεων εισόδου

Για την παρούσα προσομοίωση επιλέχθηκε $V_r = 0.5V$, μια τάση κάτω από την θετική τάση κατωφλίου της μνήμης μεταβλητής αντίστασης που θα χρησιμοποιηθεί στο δικτύωμα.

Στο σχήμα που ακολουθεί, φαίνονται όλες οι εικόνες εισόδου, δυο δείγματα για κάθε κλάση καθώς και οι κωδικοποιημένες τάσεις της κάθε γραμμής.



Σχήμα 23: Σύνολο δεδομένων εκπαίδευσης νευρωνικού δικτύου

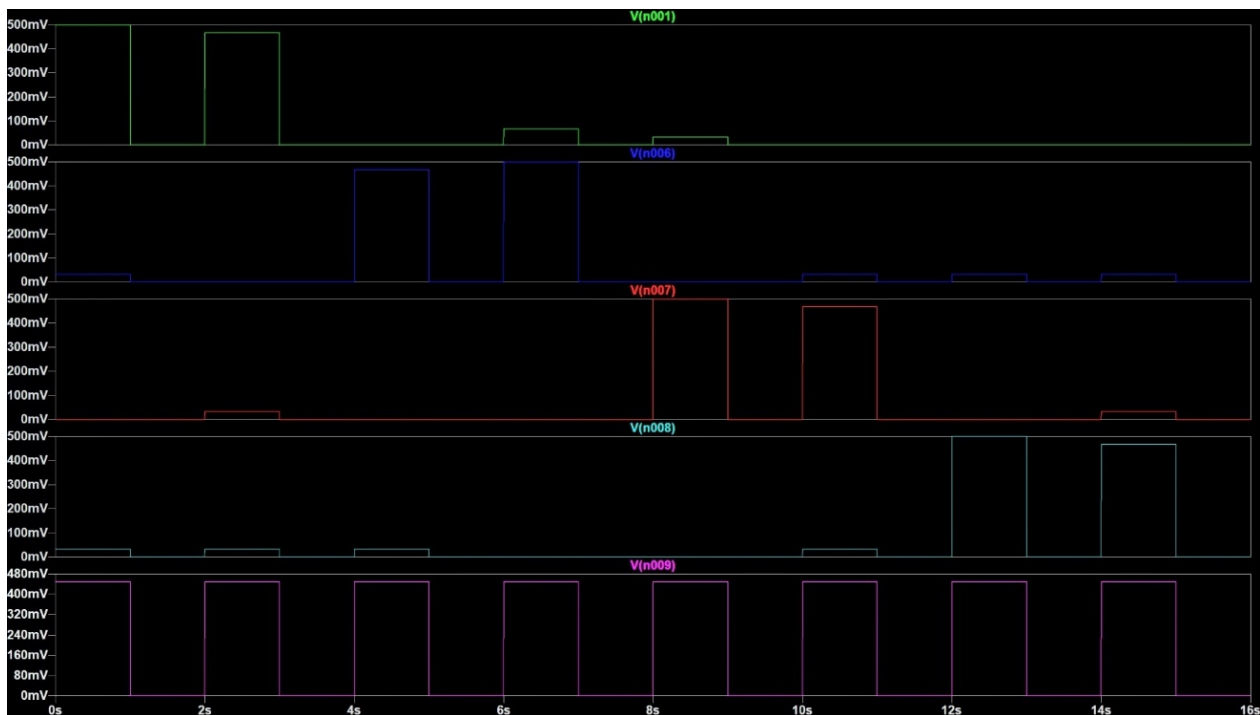
Με $V_r=0.5V$, οι παραπάνω εικόνες, σε μορφή πίνακα τάσεων είναι:

	χαρακτηριστικό-1	χαρακτηριστικό-2	χαρακτηριστικό-3	χαρακτηριστικό-4	bias	κατηγορία
	είσοδος-1 (V)	είσοδος-2 (V)	είσοδος-3 (V)	είσοδος-4 (V)	είσοδος-5 (V)	
δείγμα1	0.500	0.033	0.000	0.033	0.450	1
δείγμα2	0.467	0.000	0.033	0.033	0.450	1
δείγμα3	0.000	0.467	0.000	0.033	0.450	2
δείγμα4	0.067	0.500	0.000	0.000	0.450	2
δείγμα5	0.033	0.000	0.500	0.000	0.450	3
δείγμα6	0.000	0.033	0.467	0.033	0.450	3
δείγμα7	0.000	0.033	0.000	0.500	0.450	4
δείγμα8	0.000	0.033	0.033	0.467	0.450	4

Πίνακας 10: Πίνακας δεδομένων εκπαίδευσης σε volt

Στον πίνακα 10, φαίνεται και η πέμπτη είσοδος του δικτύωματος μνημών μεταβλητής αντίστασης, το bias, το οποίο είναι σταθερό στα 0.45V. Το bias χρησιμοποιείται ώστε να είναι πιο ευέλικτη η εκπαίδευση του νευρωνικού δικτύου και να συγκλίνει πιο γρήγορα, ενδεχομένως με πιο βέλτιστη λύση.

Όπως αναφέρθηκε στην παράγραφο που περιέγραψε το LTSpice, οι εικόνες (δείγματα) θα εισαχθούν στο δικτύωμα αντιστάσεων μεταβλητής αντίστασης μέσω γεννητριών τάσεων. Κάθε μια από τις τέσσερις τάσεις της κάθε εικόνας, μαζί με την τάση του bias, θα εισαχθούν στο δικτύωμα ταυτόχρονα και παράλληλα από διαφορετική γεννήτρια τάσης. Τα σήματα θα είναι τετραγωνικοί παλμοί. Ο χρονισμός στη παρούσα προσομοίωση είναι 1 δευτερόλεπτο τα σήματα εισόδου της πρώτης εικόνας, ένα δευτερόλεπτο μηδέν τάση σε όλες τις εισόδους, μετά ένα δευτερόλεπτο τα επόμενα σήματα εισόδου (δεύτερης εικόνας) και ούτω καθεξής. Ο χρονισμός επιλέχθηκε για να είναι ευδιάκριτα τα γραφήματα και όλη η λογική γενικότερα. Οι μνήμες μεταβλητής αντίστασης έχουν τη δυνατότητα να έχουν χρονισμό μέχρι μερικών ns. Παρακάτω είναι τα σήματα εισόδου μέσα από το LTSpice.

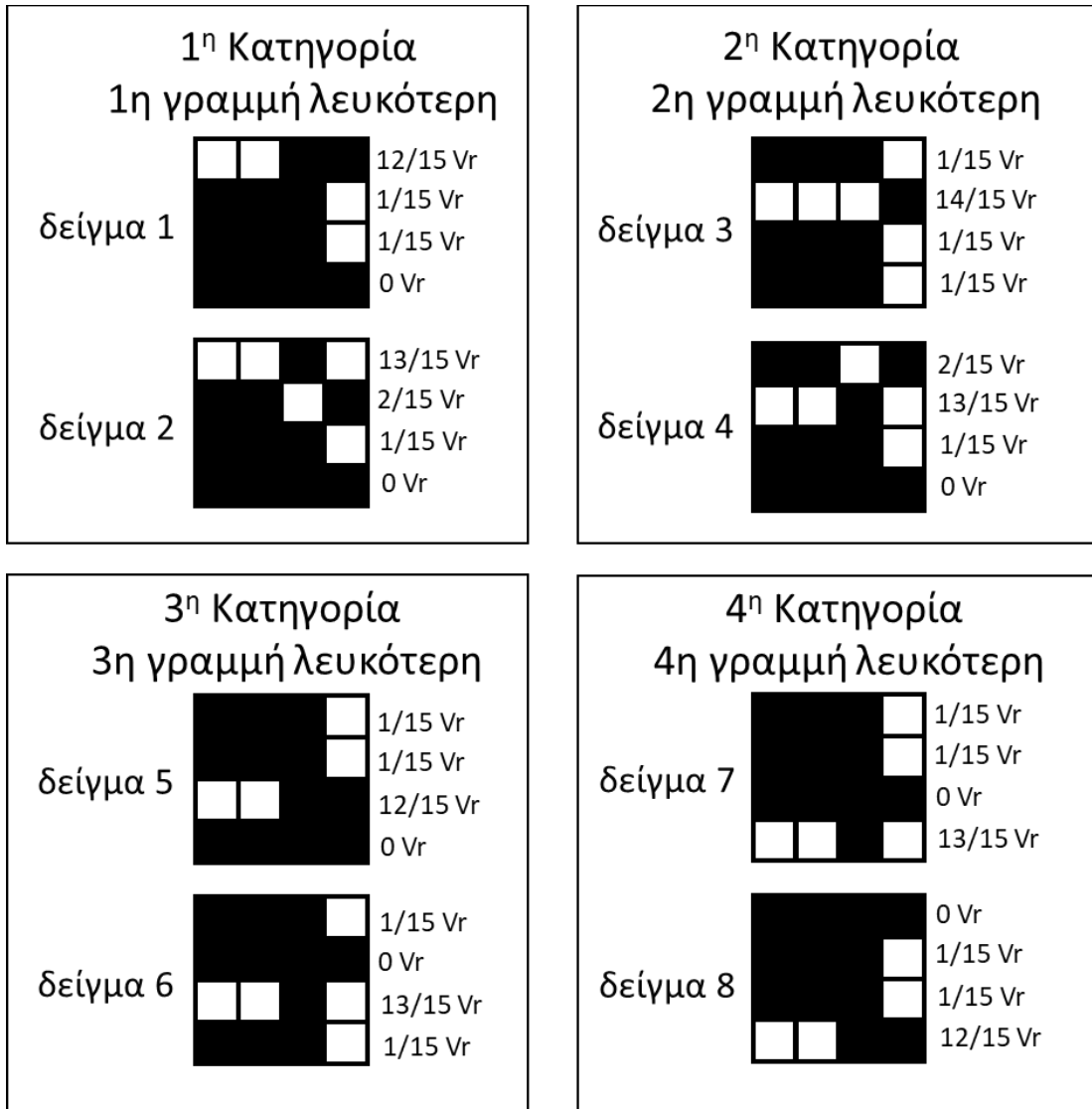


Σχήμα 24: Σήματα εισόδου μέσα στο LTSpice

Το σήμα V(n001) είναι το σήμα της πρώτης γραμμής των εικόνων. Αντίστοιχα, τα σήματα V(n006), V(n007), V(n008) και V(n009) είναι τα σήματα της δεύτερης, τρίτης και τέταρτης γραμμής των εικόνων καθώς και του bias.

Για να γίνει έλεγχος του εκπαιδευμένου νευρωνικού δικτύου, δηλαδή με τις βέλτιστες αγωγιμότητες που θα προκύψουν από τους υπολογισμούς της οπισθοδιάδοσης στους κύκλους εκπαίδευσης, θα χρησιμοποιηθεί ένα νέο σύνολο δεδομένων, τα δεδομένα ελέγχου (testing dataset), στα οποία το νευρωνικό δίκτυο δεν έχει προηγούμενη έκθεση και κατ' επέκταση "γνώση". Τα δεδομένα ελέγχου αποδεικνύουν ότι το εκπαιδευμένο νευρωνικό δίκτυο μπορεί να γενικεύει και να κάνει κατηγοριοποιήσεις σε άγνωστα δεδομένα.

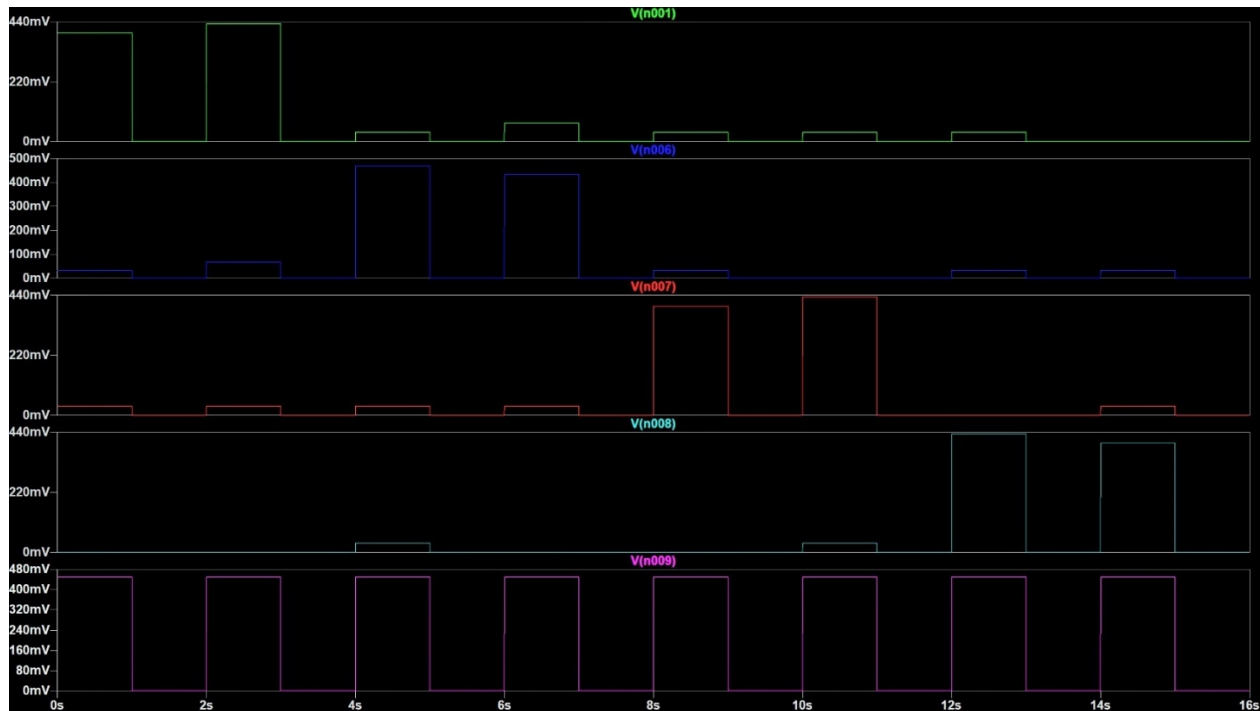
Τα δεδομένα ελέγχου φαίνονται στο παρακάτω σχήμα και πίνακα:



Σχήμα 25: Δεδομένα ελέγχου του νευρωνικού δικτύου

	χαρακτηριστικό-1	χαρακτηριστικό-2	χαρακτηριστικό-3	χαρακτηριστικό-4	bias	κατηγορία
	είσοδος-1 (V)	είσοδος-2 (V)	είσοδος-3 (V)	είσοδος-4 (V)	είσοδος-5 (V)	
δείγμα1	0.400	0.033	0.033	0.000	0.450	1
δείγμα2	0.433	0.067	0.033	0.000	0.450	1
δείγμα3	0.033	0.467	0.033	0.033	0.450	2
δείγμα4	0.067	0.433	0.033	0.000	0.450	2
δείγμα5	0.033	0.033	0.400	0.000	0.450	3
δείγμα6	0.033	0.000	0.433	0.033	0.450	3
δείγμα7	0.033	0.033	0.000	0.433	0.450	4
δείγμα8	0.000	0.033	0.033	0.400	0.450	4

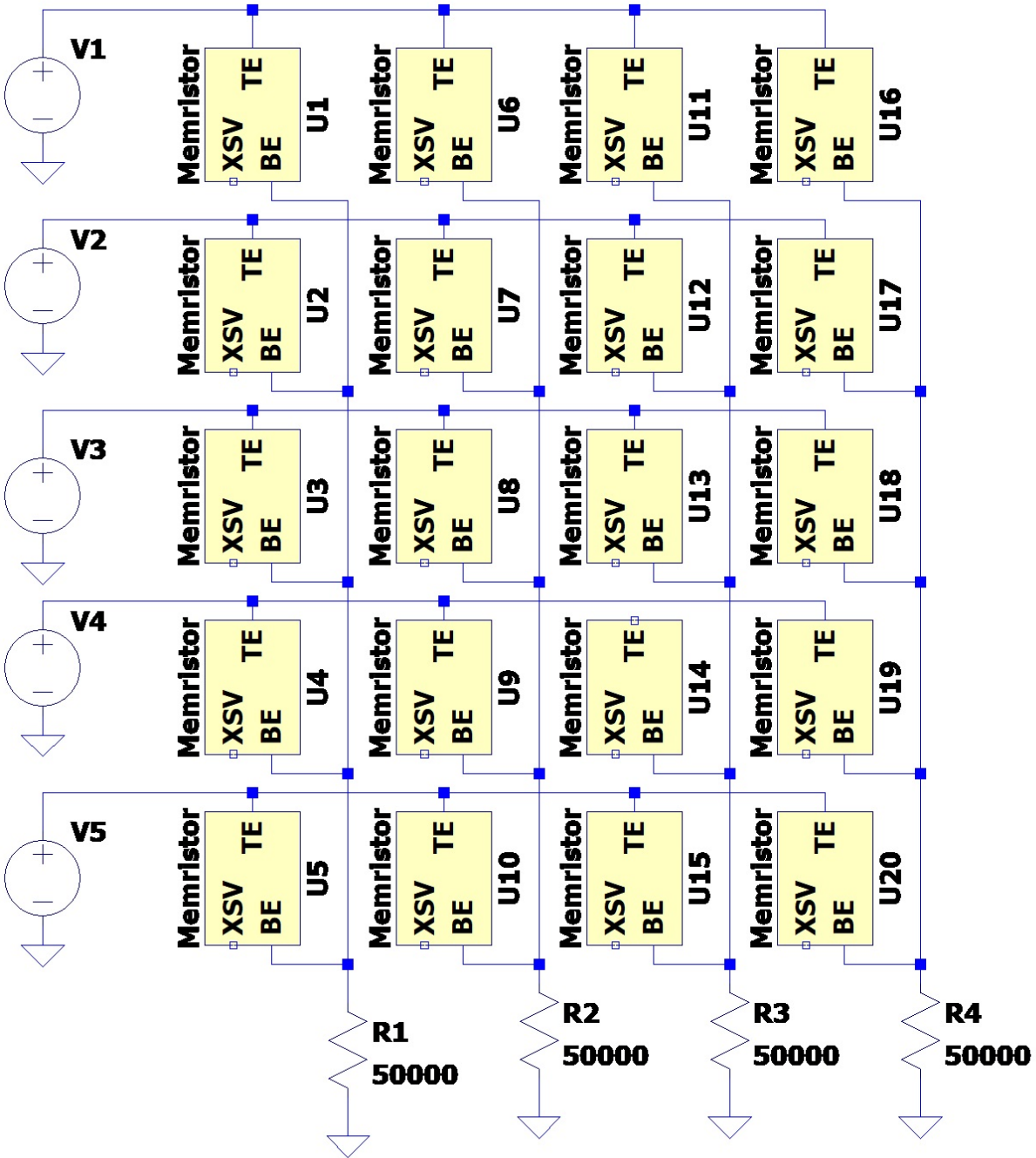
Πίνακας 11: Πίνακας δεδομένων ελέγχου



Σχήμα 26: Δεδομένα ελέγχου μέσα από το LTSpice

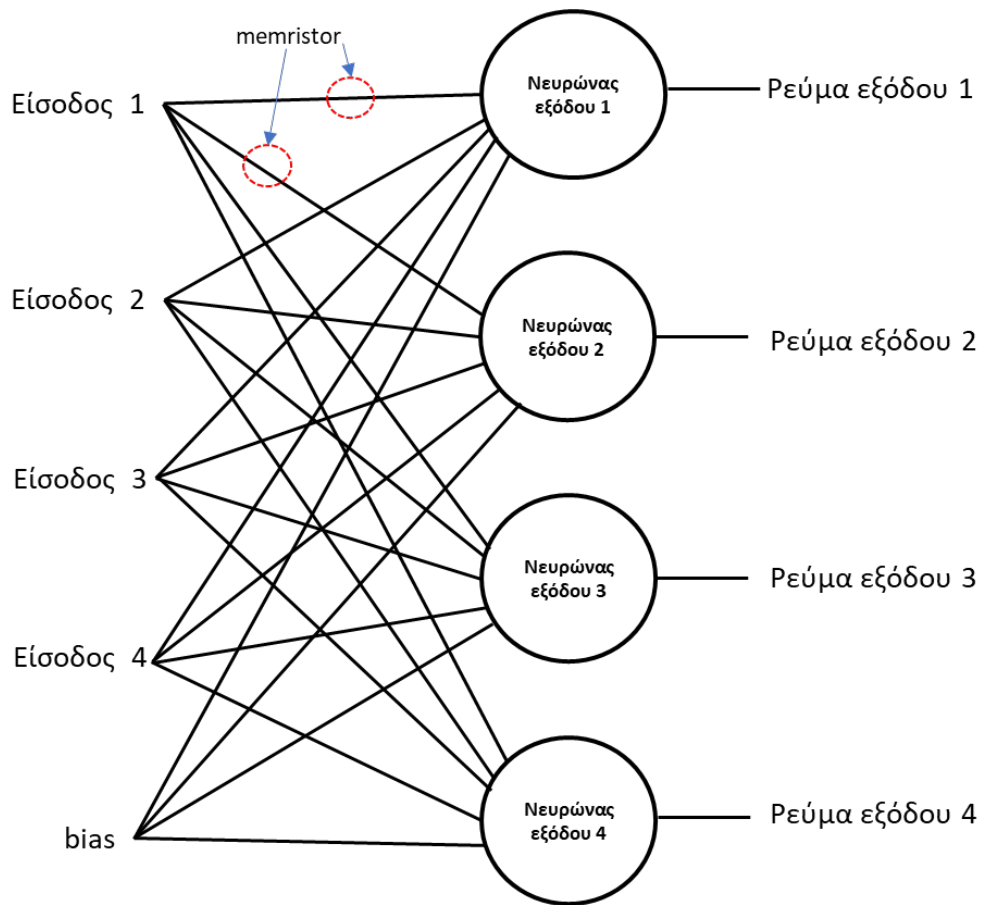
3.5 Αρχιτεκτονική του δικτύωματος μνημών μεταβλητής αντίστασης

Το δίκτυωμα μνημών μεταβλητής αντίστασης προσομοιώνει την εκπαίδευση ενός νευρωνικού δικτύου σε κύκλωμα, το οποίο θα εκτελέσει τη λειτουργία της κατηγοριοποίησης των εικόνων που αναλύθηκαν στη προηγούμενη παράγραφο. Το δίκτυωμα θα αποτελείται από είκοσι μνήμες μεταβλητής αντίστασης, οι οποίες θα λειτουργούν ως συνάψεις συνδέοντας τους πέντε νευρώνες εισόδου στους τέσσερις νευρώνες εξόδου. Τα βάρη των συνάψεων είναι οι αγωγιμότητες των μνημών μεταβλητής αντίστασης στην κατάσταση αποκοπής. Η λογική της κάθε συνάψης είναι 0T1M, δηλαδή δεν χρησιμοποιείται τρανζίστορ με τις μνήμες μεταβλητής αντίστασης ή κάποια άλλη διάταξη για τον έλεγχο και την αποφυγή των ρευμάτων ερπυσμού.



Σχήμα 27: : Το LTSpice δικτύωμα μνημών μεταβλητής αντίστασης της προσομοίωσης

Το παραπάνω LTSpice κύκλωμα αναπαρίσταται, σε κλασική απεικόνιση ΤΝΔ, ως εξής:



Σχήμα 28: Τυπική αναπαράσταση του ΤΝΔ της προσομοίωσης

3.6 Επαναληπτική μέθοδος εκπαίδευσης του δικτύωματος μνημών μεταβλητής αντίστασης

Το δίκτυωμα μνημών μεταβλητής αντίστασης εκπαιδεύτηκε με μια επαναληπτική διαδικασία οπισθοδιάδοσης, η οποία χρησιμοποιεί το LTSpice και ένα υπολογιστικό φύλλο. Οι εξισώσεις που εκτελούνται στο υπολογιστικό φύλλο μπορεί να γίνουν σε οποιαδήποτε εφαρμογή όπως Apache OpenOffice Calc, LibreOffice Calc ή Microsoft Excel.

Οι αγωγιμότητες των μνημών μεταβλητής αντίστασης στο δίκτυωμα αρχικοποιήθηκαν τυχαία γύρω από τη μέση τιμή της μεταβλητής κατάστασης, δηλαδή του $x_0 = 0.5$. Οι αρχικές τιμές των μεταβλητών καταστάσεων της κάθε μνήμης μεταβλητής αντίστασης φαίνονται στον παρακάτω πίνακα:

	χο (αρχικό)
U1	0.5
U2	0.42
U3	0.52
U4	0.58
U5	0.41
U6	0.4
U7	0.47
U8	0.51
U9	0.57
U10	0.46
U11	0.53
U12	0.51
U13	0.55
U14	0.57
U15	0.41
U16	0.49
U17	0.53
U18	0.5
U19	0.42
U20	0.5

Πίνακας 12: Αρχικές τιμές μεταβλητής κατάστασης των μνημών μεταβλητής αντίστασης

Η εκπαίδευση έγινε ανά μια εικόνα, χωρίς στοχαστικότητα στην επιλογή των εικόνων. Πρώτα εισήχθησαν οι εικόνες της πρώτης κατηγορίας, μετά της δεύτερης, της τρίτης και στο τέλος της τέταρτης κατηγορίας. Όπως θα δειχθεί στα αποτελέσματα, ο τρόπος επιλογής εισόδου των δειγμάτων επηρεάζει τον αριθμό των κύκλων εκπαίδευσης που απαιτούνται για 100% ακρίβεια στη τελική κατηγοριοποίηση αλλά χρησιμοποιείται ώστε να παρουσιαστεί πιο αποδοτικά η μεθοδολογία της εκπαίδευσης και η συμπεριφορά των μνημών μεταβλητής κατάστασης.

Στη διαδικασία της εκπαίδευσης, τροφοδοτείται το δικτύωμα με τις τάσεις εισόδου που αντιπροσωπεύουν τη πρώτη εικόνα. Το δικτύωμα εκτελεί το εσωτερικό γινόμενο ανά στήλη του κάθε νευρώνα εξόδου (μια από τα βασικές λειτουργίες του δικτυώματος) και παράγει τις εξόδους [4]. Οι εξόδοι ακολουθούν τον νόμο του Kirchhoff, δηλαδή:

$$\begin{aligned}
 I_{R1} &= V_1 * G_{U1} + V_2 * G_{U2} + V_3 * G_{U3} + V_4 * G_{U4} + V_5 * G_{U5} \\
 I_{R2} &= V_1 * G_{U6} + V_2 * G_{U7} + V_3 * G_{U8} + V_4 * G_{U9} + V_5 * G_{U10} \\
 I_{R1} &= V_1 * G_{U11} + V_2 * G_{U12} + V_3 * G_{U13} + V_4 * G_{U14} + V_5 * G_{U15} \\
 I_{R1} &= V_1 * G_{U16} + V_2 * G_{U17} + V_3 * G_{U18} + V_4 * G_{U19} + V_5 * G_{U20}
 \end{aligned}
 \quad (21)$$

όπου τα $G_{U_{xx}}$ είναι οι αγωγιμότητες των μνημών μεταβλητής αντίστασης στο δικτύωμα.

Τα ρεύματα εξόδου στους νευρώνες εξόδου χρησιμοποιήθηκαν ώστε να καθοριστεί ποια διόρθωση χρειάζονται τα βάρη των συνάψεων, δηλαδή οι αγωγιμότητες των μνημών μεταβλητής αντίστασης για

να επιτευχθεί σύγκλιση και σωστή κατηγοριοποίηση των εικόνων. Η εκπαίδευση ακολούθησε την λογική οπισθοδιάδοσης και επειδή δεν υπάρχει κρυμμένο επίπεδο στην εφαρμογή της παρούσας εργασίας δεν χρησιμοποιήθηκε κλίση καθόδου.

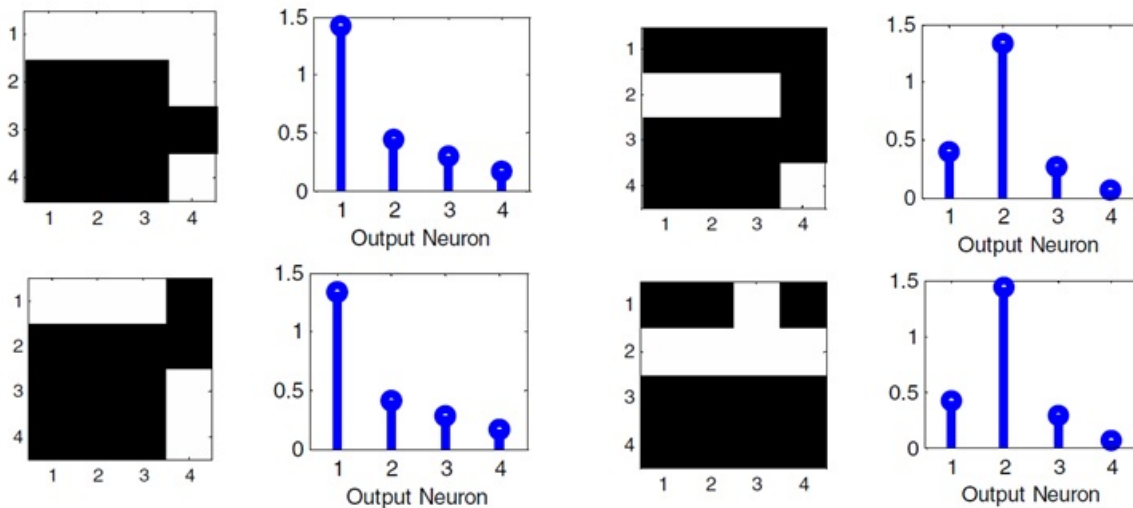
Η συνάρτηση ενεργοποίησης που χρησιμοποιήθηκε στα ρεύματα εξόδου ήταν η SoftMax, η οποία χρησιμοποιεί τον κανόνα του Baye's και δίνει μια φυσική αντιπροσώπευση της υστέρων κατανομής σε εφαρμογές κατηγοριοποίησης πολλών κατηγοριών. Η πιθανότητα αυτή, δηλαδή το ρεύμα ενός νευρώνα εξόδου να ανήκει σε μια συγκεκριμένη κατηγορία δίνεται ως εξής [5]:

$$y_c(n) = \frac{e^{k \cdot I_c(n)}}{\sum_{m=1}^C e^{k \cdot I_m(n)}} \quad (22)$$

Στη παραπάνω σχέση (22) έχουμε τα εξής:

- C ο αριθμός των κατηγοριών των δεδομένων. Στη παρούσα εφαρμογή, C=4
- $y_c(n)$ είναι η πιθανότητα η έξοδος του νευρώνα για το δείγμα n να ανήκει στη κατηγορία c
- $I_c(n)$ είναι το ρεύμα εξόδου του νευρώνα
- k είναι ένας συντελεστής ώστε να δοθεί μεγαλύτερη βαρύτητα στην έξοδο με το μεγαλύτερο ρεύμα. Στους υπολογισμούς της παρούσας προσομοίωσης χρησιμοποιήθηκε k=1000
- το άθροισμα του παρονομαστή είναι όλα τα ρεύματα εξόδου

Στην παρακάτω εικόνα αποτυπώνεται η λογική του εκπαιδευμένου νευρωνικού δικτύου, το οποίο δίνει μεγαλύτερη έξοδο (ρεύμα στη περίπτωση μας) με τις αντίστοιχες εικόνες εισόδου.



Σχήμα 29: Εικόνες εισόδου και έξοδοι νευρώνων [3]

Τα ρεύματα εξόδου του κάθε νευρώνα, μετά την τροφοδοσία του δικτυώματος με τις τάσεις αντιπροσώπευσης μιας εικόνας αποθηκευτήκαν σε log αρχείο με τις LTSprice εντολές .means tran result που αναλύθηκαν στην παράγραφο του LTSprice. Τα ρεύματα αυτά περάστηκαν στο υπολογιστικό φύλλο και υπολογίστηκαν τα τέσσερα y_c από τη σχέση (22).

Το βάρη των συνάψεων μετά από κάθε κύκλο εκπαίδευσης (εισαγωγή μιας εικόνας) ανανεώνονταν σύμφωνα με τις παρακάτω σχέσεις [5], [6]:

$$\delta_j = t_j - y_j \quad (23)$$

$$\Delta W_{ij} = a * \delta_j * V_{in-i} \quad (24)$$

$$W_{next-epoch} = W_{current} + \Delta W_{ij} \quad (25)$$

Για τις παραπάνω τρεις σχέσεις έχουμε τα εξής:

- Το δ_j είναι το σφάλμα του j νευρώνα εξόδου
- Το y_j είναι το y_c της σχέσης (22)
- Το t_j είναι 1 αν η έξοδος είναι η σωστή κατηγορία και 0 για την άλλη περίπτωση
- ΔW_{ij} είναι η τροποποίηση της αγωγιμότητας που πρέπει να γίνει στη μνήμη μεταβλητής αντίστασης της i γραμμής και j στήλης του δικτύματος για τον επόμενο κύκλο εκπαίδευσης
- V_{in-i} είναι η τάση εισόδου της γραμμής i του δικτύματος
- a είναι ο ρυθμός μάθησης (learning rate). Ο ρυθμός μάθησης είναι υπέρ-παράμετρος στα νευρωνικά δίκτυα και χρειάζεται να βρεθεί η βέλτιστη τιμή του. Με πειραματισμούς επιλέχθηκε $a=0.001$ για τη προσομοίωση

Ειδικά για το $W_{next-epoch}$, πρέπει να σημειωθεί ότι είναι η αγωγιμότητα που θα χρησιμοποιηθεί στον επόμενο κύκλο εκπαίδευσης. Η τιμή αυτή προκύπτει από την υπάρχουσα αγωγιμότητα της κάθε μνήμης μεταβλητής αντίστασης προσθέτοντας το ΔW από την παραπάνω σχέση. Το ΔW παίρνει θετικές καθώς και αρνητικές τιμές. Για να παραμετροποιηθούν οι μνήμες μεταβλητής αντίστασης στις νέες τιμές αγωγιμότητας, πρέπει να μετατραπούν σε μεταβλητές κατάστασης x_0 (αυτή είναι η μεταβλητή που παραμετροποιούμε) κάτι το οποίο επιτυγχάνεται με την εξίσωση (γραμμή) παλινδρόμησης που περιγράφηκε στην παράγραφο του μοντέλου της μνήμης μεταβλητής αντίστασης.

$$x_0 = 117.635 * W - 8.2671E-08 \quad (26)$$

Οι παραπάνω σχέσεις χρησιμοποιούνται αν η εκπαίδευση του νευρωνικού δικτύου γίνει με βήμα μιας εικόνας ανά κύκλο εκπαίδευσης, κάτι το οποίο εφαρμόζεται στη παρούσα εργασία. Αν χρησιμοποιηθούν όλες οι εικόνες (batch) για τον υπολογισμό το ΔW , τότε οι παραπάνω σχέσεις γίνονται:

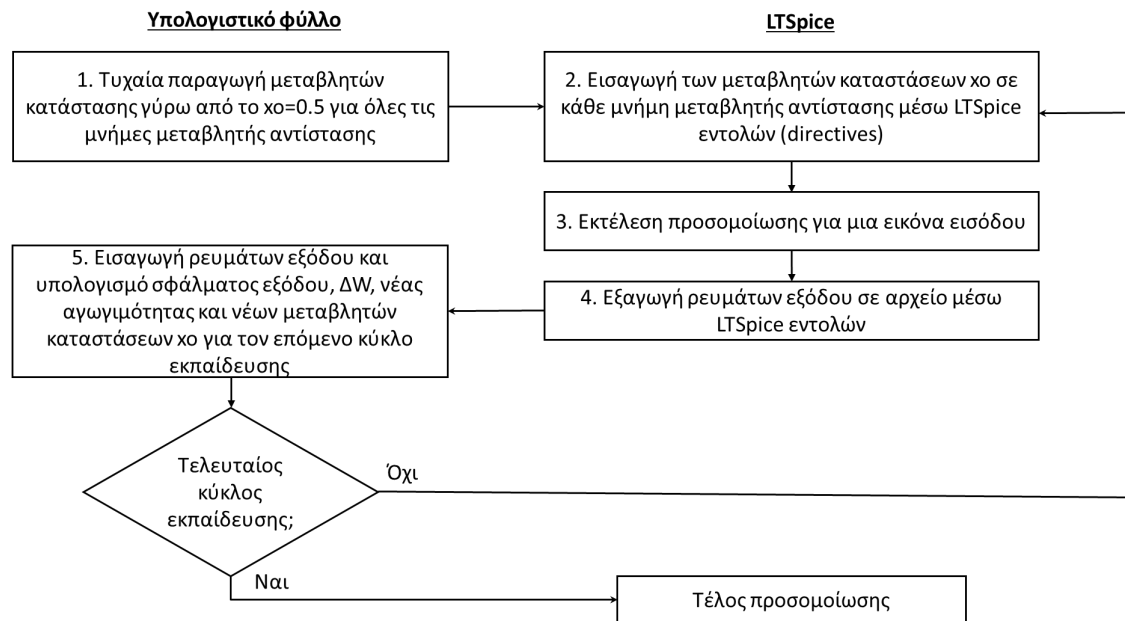
$$\delta_j = y_j - t_j \quad (27)$$

$$\Delta W_{ij} = a * \sum_{n=1}^S \delta_j * V_{in-i} \quad (28)$$

όπου S ο συνολικός αριθμός των δειγμάτων εκπαίδευσης.

Τα μόνα που αλλάζουν είναι η σειρά των μελών στην αφαίρεση που υπολογίζει το σφάλμα καθώς και στο ΔW , το οποίο υπολογίζεται όπως πριν αλλά σαν άθροισμά σε όλα τα δείγματα. Οι ίδιες σχέσεις ισχύουν και στη περίπτωση μίνι-παρτίδων (mini-batch), όπου το σύνολο των δειγμάτων χωρίζεται σε έναν αριθμό N υποσυνόλων και το ΔW υπολογίζεται με το άθροισμα στο N .

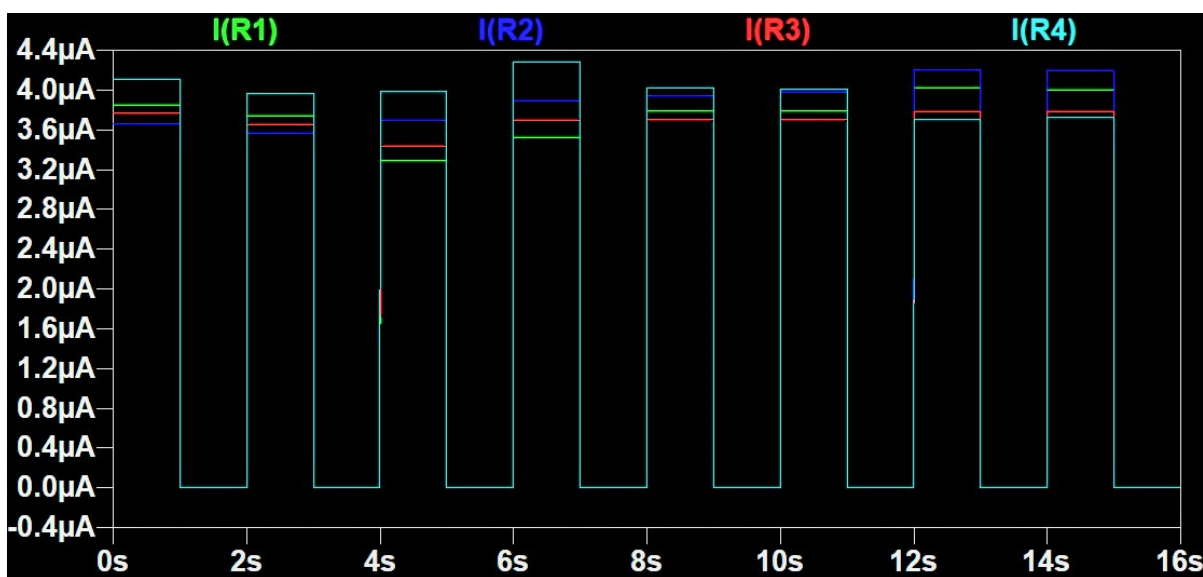
Στο παρακάτω διάγραμμα παρουσιάζεται όλη η παραπάνω διαδικασία εκπαίδευσης.



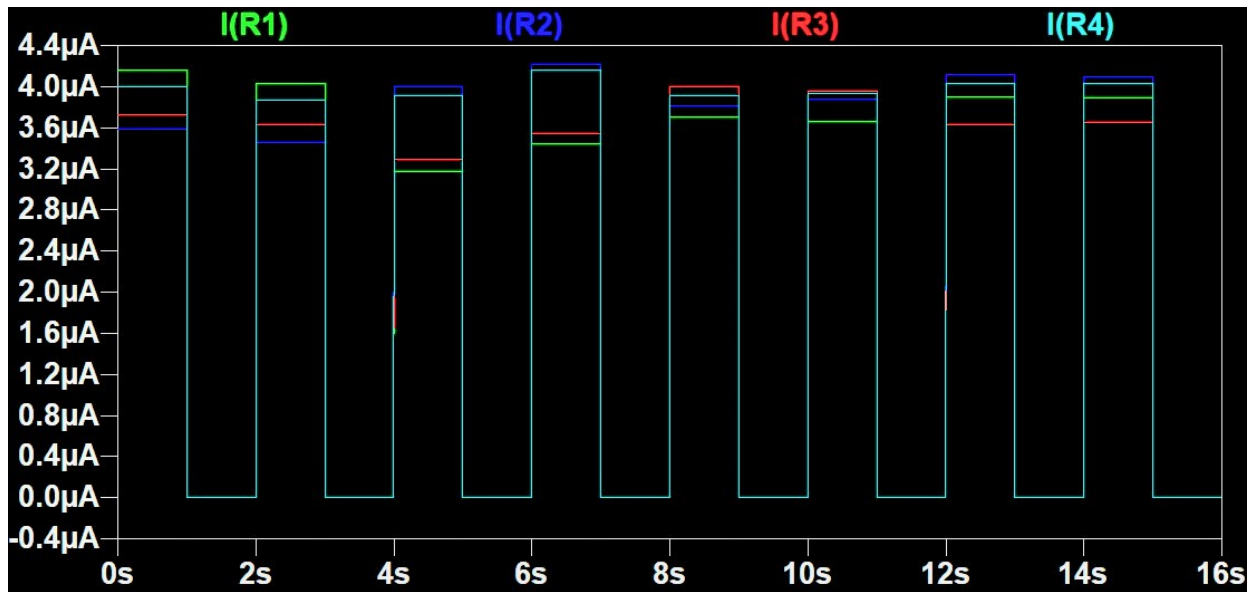
Σχήμα 30: Διαδικασία προσομοίωσης και εκπαίδευσης του δικτυώματος μνημών μεταβλητής αντίστασης

3.7 Αποτελέσματα της προσομοίωσης

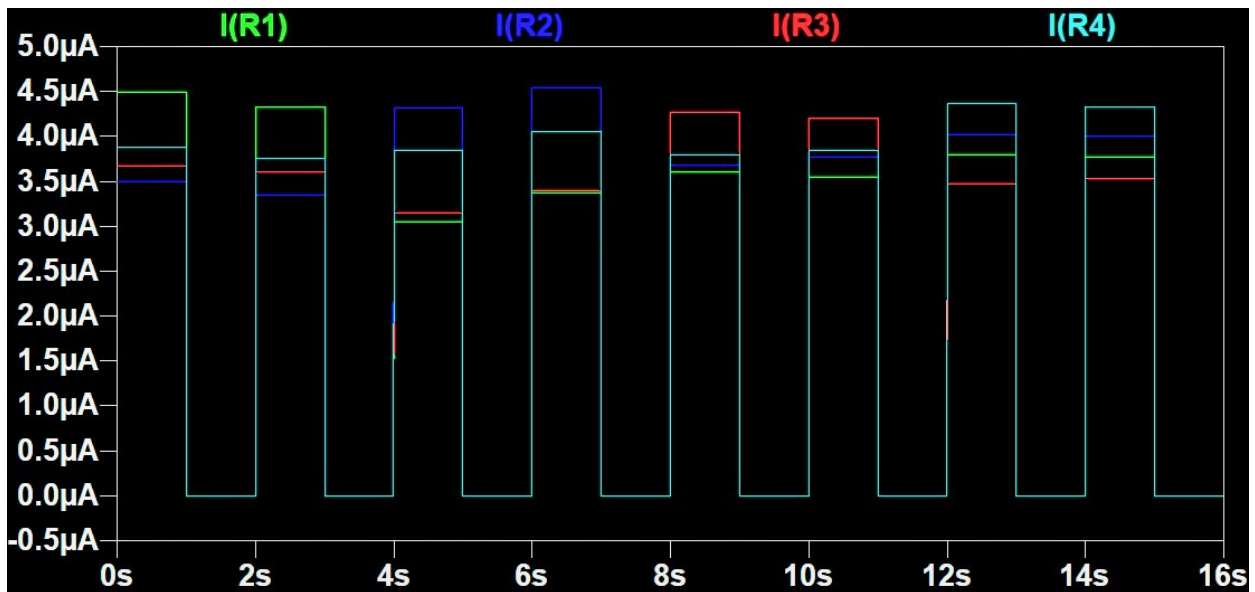
Η εκπαίδευση του νευρωνικού δικτύου έγινε με μια εικόνα ανά κύκλο εκπαίδευσης. Στα σχήματα που ακολουθούν φαίνονται όλα τα ρεύματα εξόδου αλλά μόνο μια εικόνα χρησιμοποιήθηκε για τον υπολογισμό το ΔW σε κάθε κύκλο εκπαίδευσης. Ακολουθούν σχήματα για την αρχική έξοδο του δικτυώματος και τις εξόδους μετά το όγδοο, το δέκατο έκτο και εικοστό τέταρτο κύκλο εκπαίδευσης, ώστε να φανεί το αποτέλεσμα ενός πλήρους κύκλου εκπαίδευσης από όλα τα δείγματα εισόδου. Όλα τα αποτελέσματα παρουσιάζονται στο δεύτερο παράρτημα.



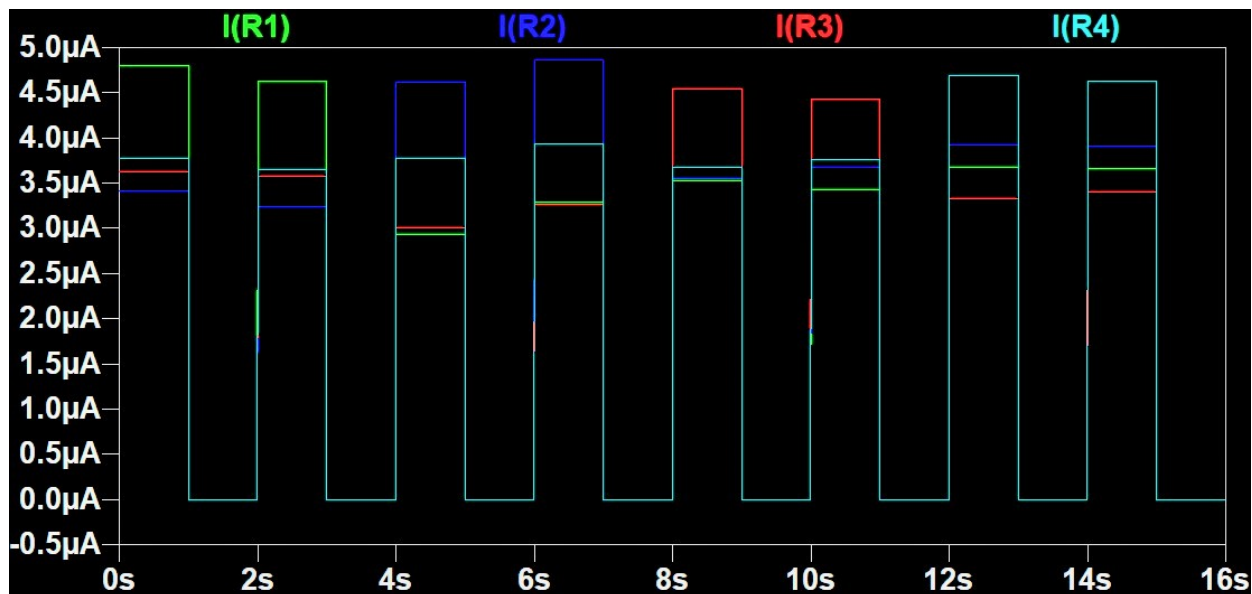
Σχήμα 31: Αρχική έξοδος του δικτυώματος με αρχικές μεταβλητές καταστάσεων



Σχήμα 32: Έξοδος του δικτύματος μετά από 8 κύκλους εκπαίδευσης



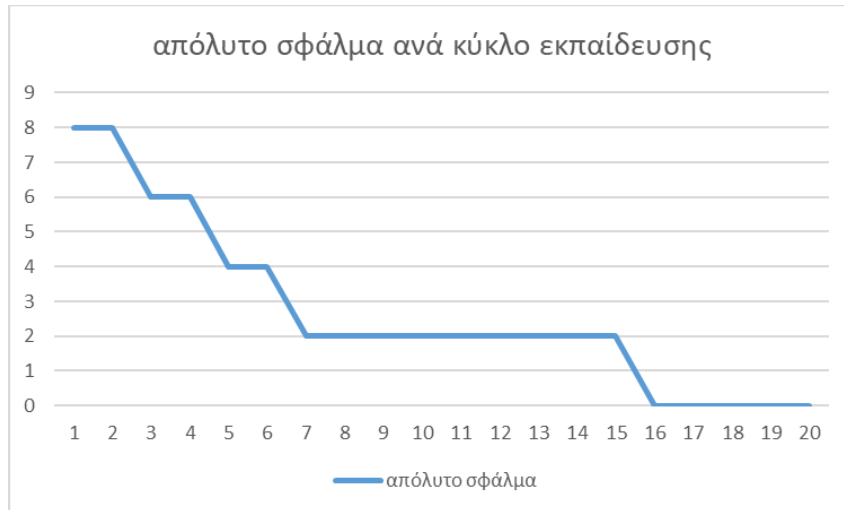
Σχήμα 33: Έξοδος του δικτύματος μετά από 16 κύκλους εκπαίδευσης



Σχήμα 34: Έξοδος του δικτύωματος μετά από 24 κύκλους εκπαίδευσης

Στα παραπάνω σχήματα με τα ρεύματα εξόδου φαίνεται η σταδιακή αύξηση του ρεύματος εξόδου στις σωστές κατηγορίες. Για παράδειγμα το ρεύμα στην αντίσταση R1 (έξοδος των πρώτου νευρώνα εξόδου) για τα σήματα εισόδου στις χρονικές στιγμές 0-1s και 2-3s, που είναι της πρώτης κατηγορίας, είναι το μεγαλύτερο στο 24^ο κύκλο εκπαίδευσης (ήδη γίνεται σωστή κατηγοριοποίηση από τον όγδοο κύκλο).

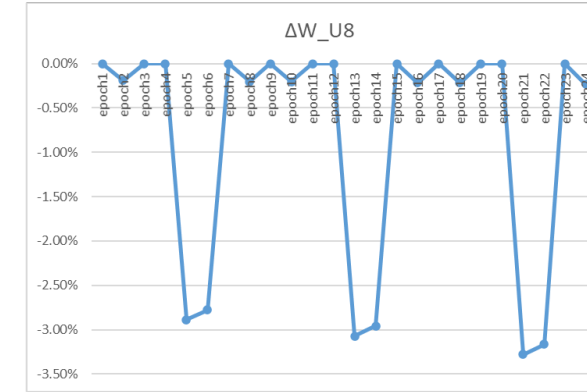
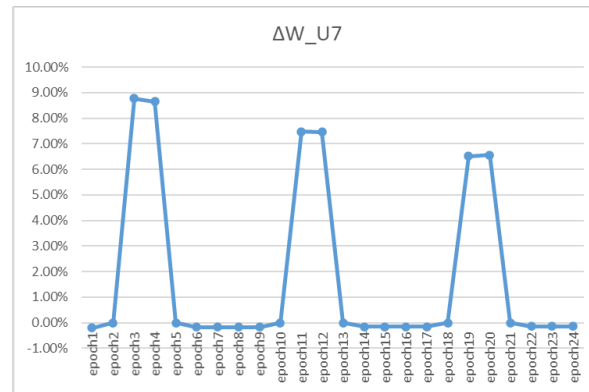
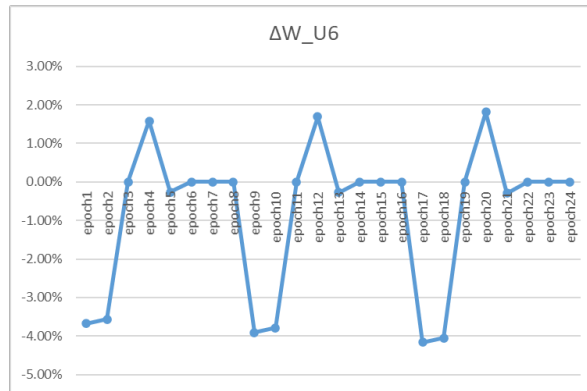
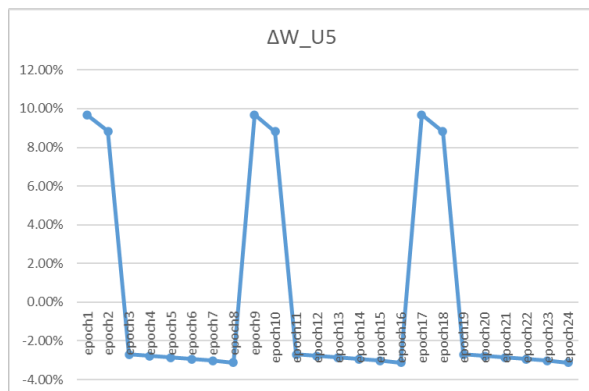
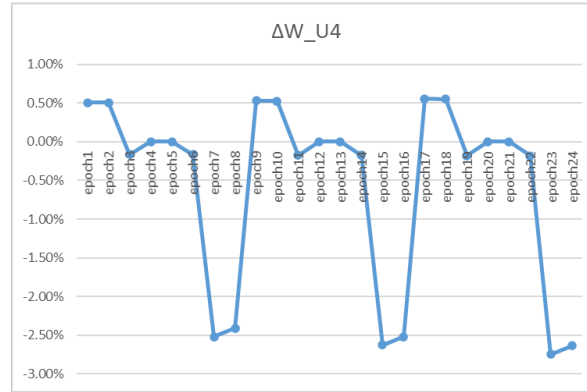
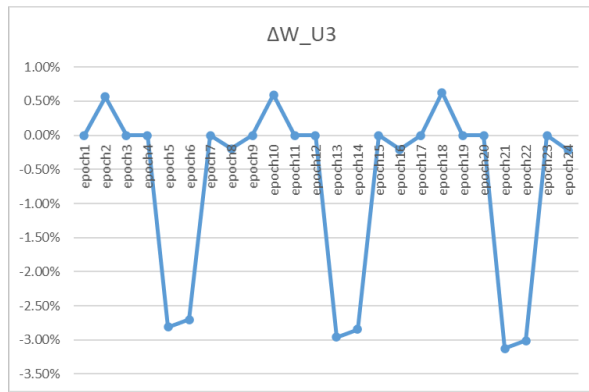
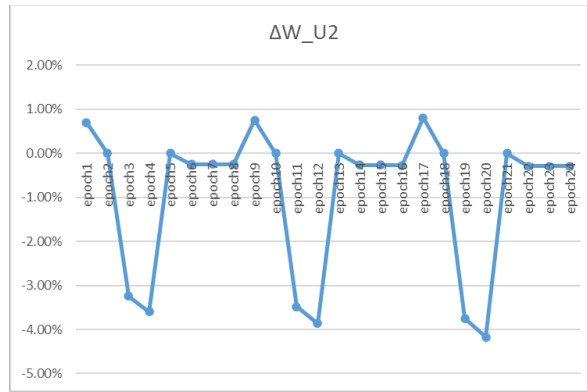
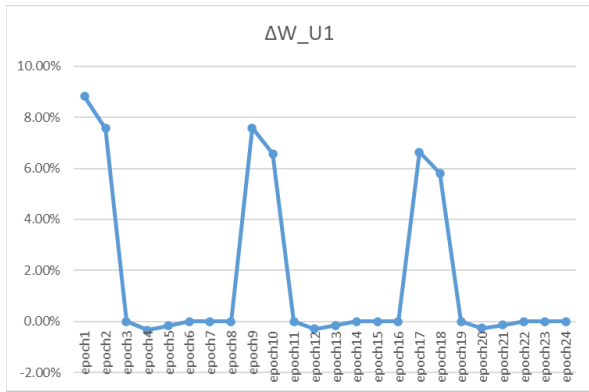
Από τα παραπάνω γραφήματα φαίνεται ότι στην αρχική κατάσταση το δίκτυωμα μνημών μεταβλητής αντίστασης δεν κάνει καμία σωστή κατηγοριοποίηση. Επίσης μετά το πέρας του 8^{ου} κύκλου εκπαίδευσης, το δίκτυωμα κατηγοριοποιεί οριακά σωστά τις τρεις πρώτες κατηγορίες. Μετά το πέρας 16 κύκλων εκπαίδευσης, το δίκτυωμα επιτυγχάνει 100% ακρίβεια, αφού χρησιμοποιηθούν δυο φορές οι εικόνες της τέταρτης κατηγορίας (εκεί που ήταν τα σφάλματα) για την εκπαίδευση των μνημών μεταβλητής αντίστασης. Η στοχαστικότητα στην σειρά επιλογής των δειγμάτων εισόδου επηρεάζει τον τελικό αριθμό κύκλων εκπαίδευσης που χρειάζονται ώστε να φτάσει το δίκτυωμα σε 100% ακρίβεια στην κατηγοριοποίηση.



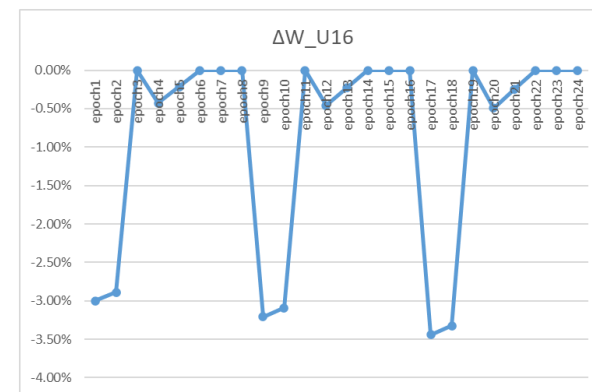
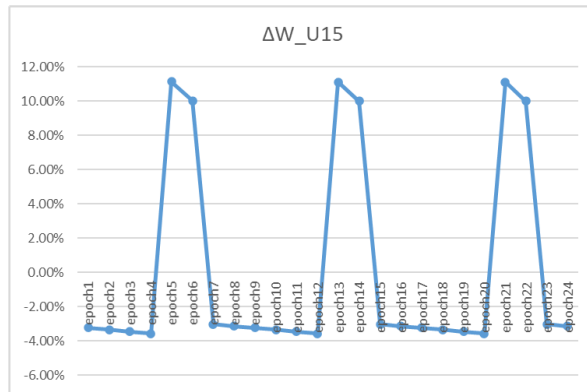
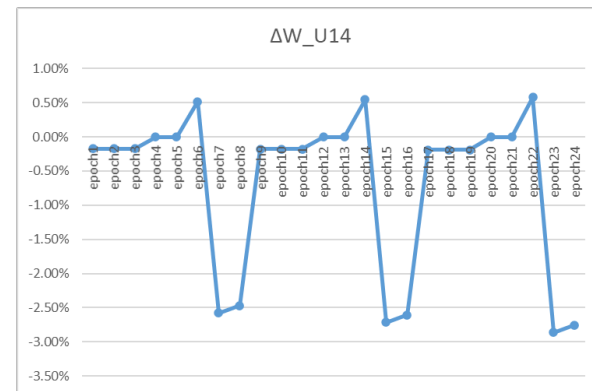
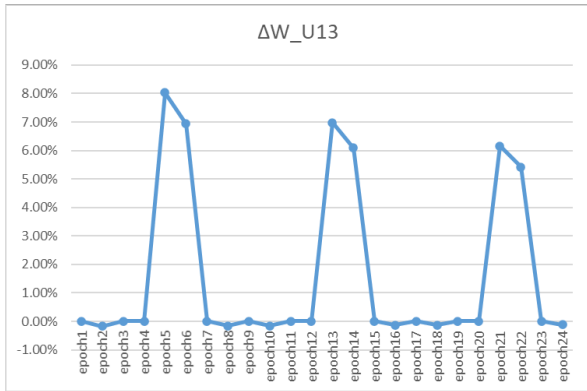
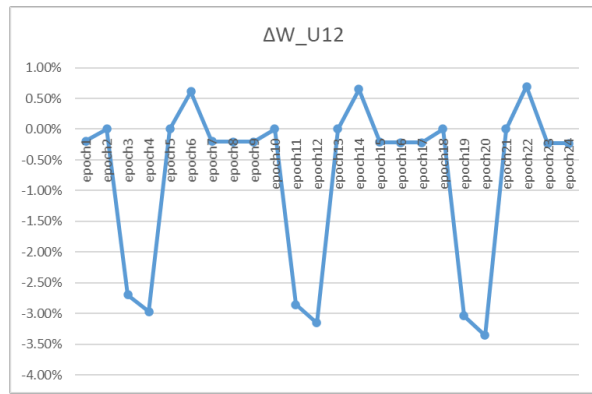
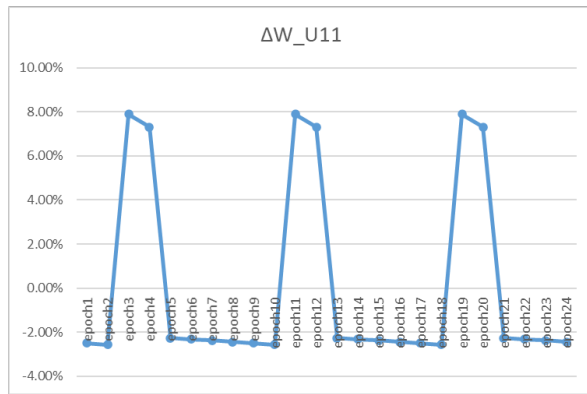
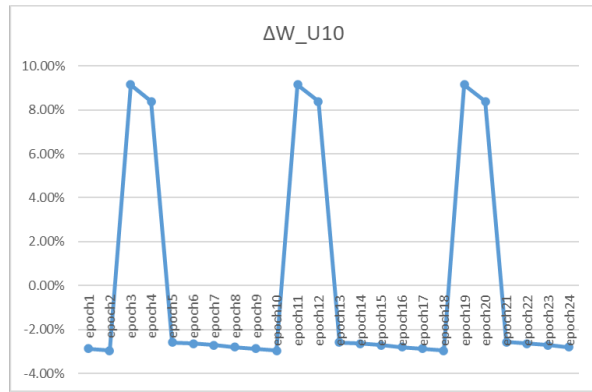
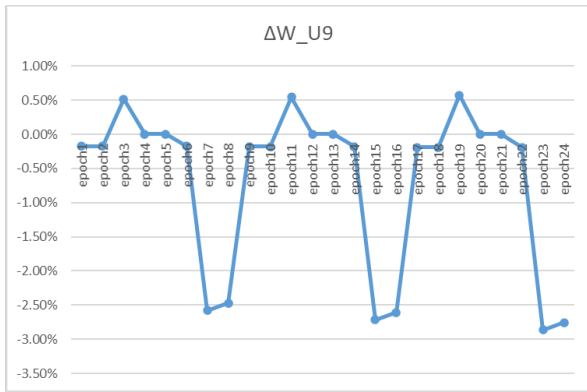
Σχήμα 35: Απόλυτη ακρίβεια κατηγοριοποίησης ανά κύκλο εκπαίδευσης

Η εκπαίδευση του δικτυώματος μνημών μεταβλητής αντίστασης θα μπορούσε να συνεχιστεί πέρα των 24 κύκλων εκπαίδευσης, κάτι το οποίο θα είχε σαν αποτέλεσμα να μεγαλώσει η διαφορά του ρεύματος εξόδου μεταξύ της σωστής κατηγορίας και μη. Το όριο στην εκπαίδευση θα το καθορίσουν το υλικό της μνήμης μεταβλητής αντίστασης, η συμπεριφορά των μνημών στον χρόνο και την χρήση (conductance drift) [7] καθώς και η κρισιμότητα της εφαρμογής.

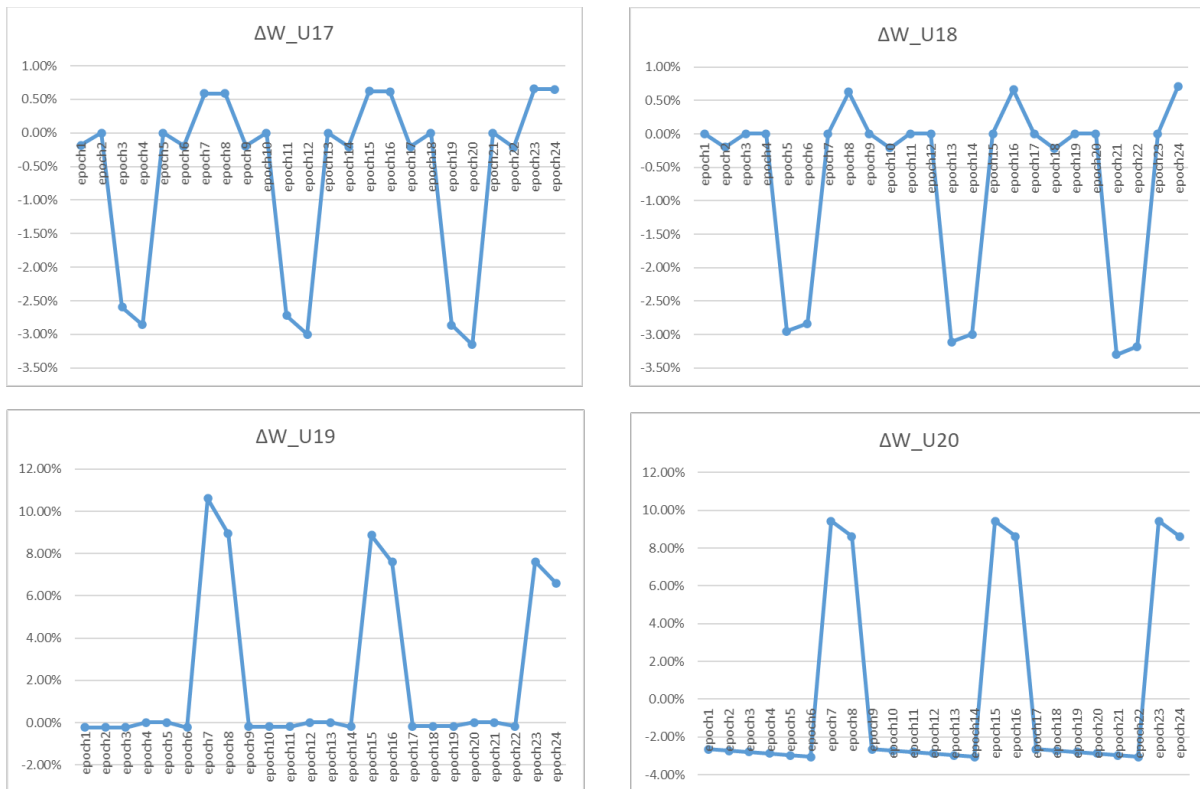
Λόγω του μικρού μεγέθους του δικτυώματος, την προκαθορισμένη σειρά των δειγμάτων εισόδου και του τρόπου εκπαίδευσης (μια εικόνα εισόδου ανά κύκλο εκπαίδευσης) μπορούν να παρατηρηθούν οι συμπεριφορές και οι αλλαγές της αγωγιμότητας της κάθε μνήμης μεταβλητής αντίστασης ανά κύκλο εκπαίδευσης.



Σχήμα 36: Μεταβολές αγωγιμότητας ανά μνήμη μεταβλητής αντίστασης (U1 - U8) ανά κύκλο εκπαίδευσης



Σχήμα 37: Μεταβολές αγωγιμότητας ανά μνήμη μεταβλητής αντίστασης (U9 - U16) ανά κύκλο εκπαίδευσης



Σχήμα 38: Μεταβολές αγωγιμότητας ανά μνήμη μεταβλητής αντίστασης (U17 - U20) ανά κύκλο εκπαίδευσης

Από τα παραπάνω γραφήματα παρατηρούμε ότι σε κάθε κύκλο εκπαίδευσης η πλειοψηφία των μνημών μεταβλητής αντίστασης έχουν μεταβολή στις αγωγιμότητές τους. Για την πρώτη κατηγορία εικόνων, φαίνεται από τα ΔW_{U1} και ΔW_{U5} στους κύκλους εκπαίδευσης (1, 2, 9, 10, 17 και 18) που χρησιμοποιούνται δεδομένα της πρώτης κλάσης ότι οι μνήμες μεταβλητής αντίστασης U1 και U5 με αυξανόμενη αγωγιμότητα και οι U6 και U16 με μειούμενη αγωγιμότητα επηρεάζουν περισσότερο τη σωστή αναγνώριση της κατηγορίας αυτής. Αντίστοιχα, από τα παραπάνω γραφήματα προκύπτει ότι για η σωστή κατηγοριοποίηση της δεύτερης κατηγορίας επηρεάζεται θετικά από τις U7, U10 και U11 και αρνητικά από τις U2, U12 και U17, της τρίτης θετικά από τις U13 και U15 και αρνητικά από τις U3, U8 και U18 και της τέταρτης θετικά από τις U19 και U20 και αρνητικά από τις U4, U9 και U14.

Η μέγιστη μεταβολή που παρατηρήθηκε με ρυθμό μάθησης 0.001 ήταν στο +11%. Η σύγκλιση του νευρωνικού δικτύου θα μπορούσε να επιτευχθεί πιο γρήγορα με αύξηση του ρυθμού μάθησης καθώς και με χρήση όλων των εικόνων εισόδου σε κάθε κύκλο εκπαίδευσης (batch).

3.7 Δεύτερο μοντέλο μνήμης μεταβλητής αντίστασης για προσομοίωση δικτύματος

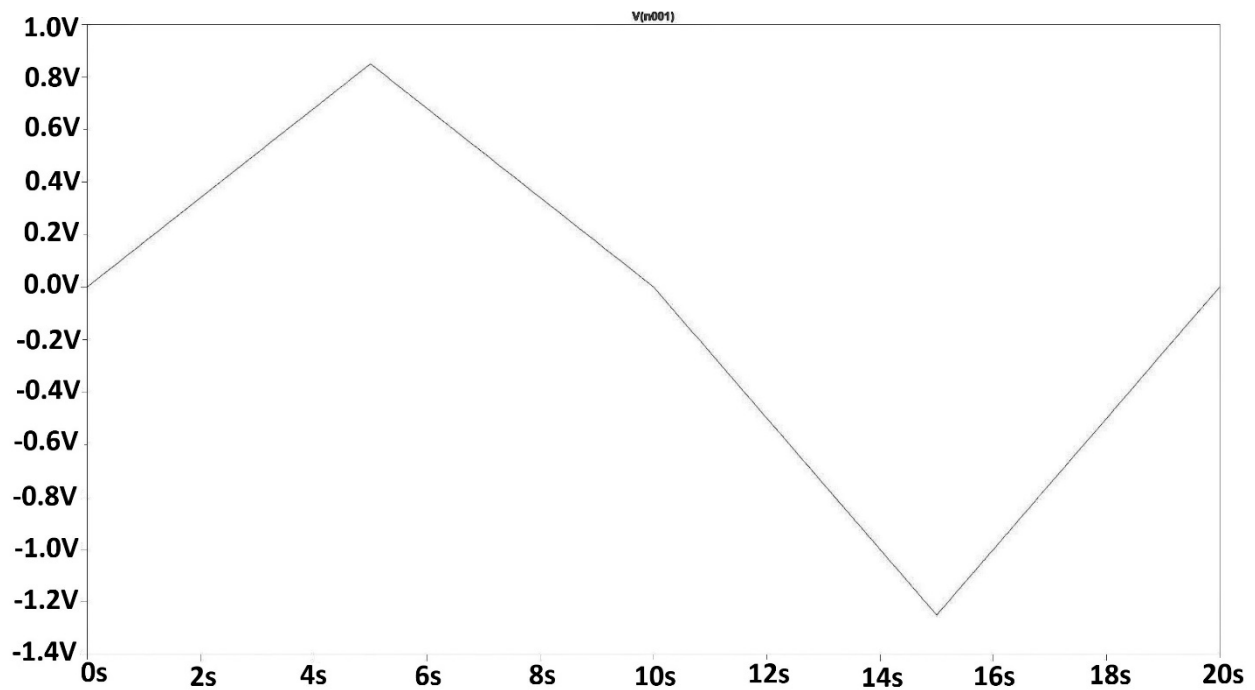
Για να εξεταστεί πόσο και πως αλλάζουν τα αποτελέσματα της προσομοίωσης η κλίση της χαρακτηριστικής I-V της μνήμης μεταβλητής αντίστασης στην κατάσταση υψηλής αντίστασης, όλη η παραπάνω διαδικασία προσομοίωσης έγινε πάλι αλλά για διαφορετική μνήμη μεταβλητής αντίστασης. Η μνήμη μεταβλητής αντίστασης που χρησιμοποιήθηκε είναι από τη δημοσίευση [8] που αναλύει μνήμη

μεταβλητής αντίστασης από οξειδίο του τανταλίου, με τις παρακάτω παραμέτρους για το μοντέλο του Yakorcić [1]:

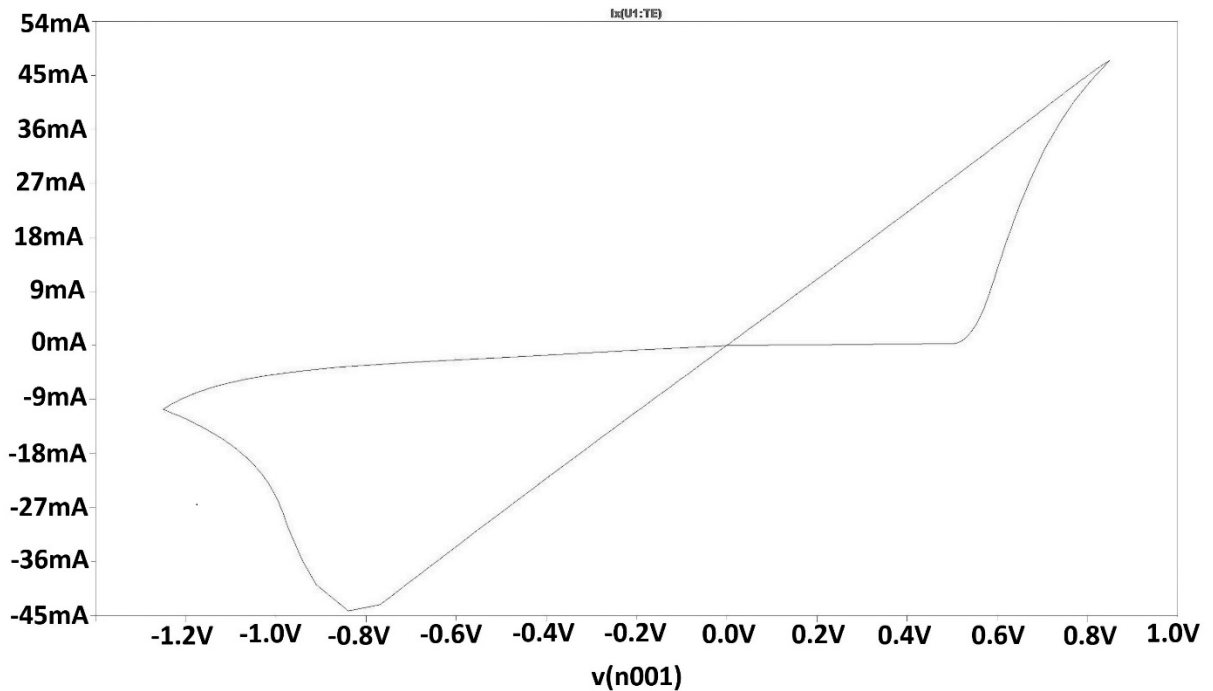
Παράμετρος	Τιμή 2 ^{ης} μνήμης μεταβλητής αντίστασης	Τιμή 1 ^{ης} μνήμης μεταβλητής αντίστασης
α_1	0.11	0.17
α_2	0.11	0.17
b	0.5	0.05
V_p	0.5V	0.6
V_n	0.75V	0.5
A_p	7.5	4000
A_n	2	4000
x_p	0.3	0.5
x_n	0.5	0.5
α_p	1	1
α_n	5	5
η	1	1

Πίνακας 13: Παράμετροι μοντέλου Yakorcić για τη 1η και 2η μνήμη μεταβλητής αντίστασης

Παρακάτω είναι τα γραφήματα τάσης εισόδου και I-V για τη 2^η μνήμη μεταβλητής αντίστασης.



Σχήμα 39: Τάση εισόδου 2^{ης} μνήμης μεταβλητής αντίστασης



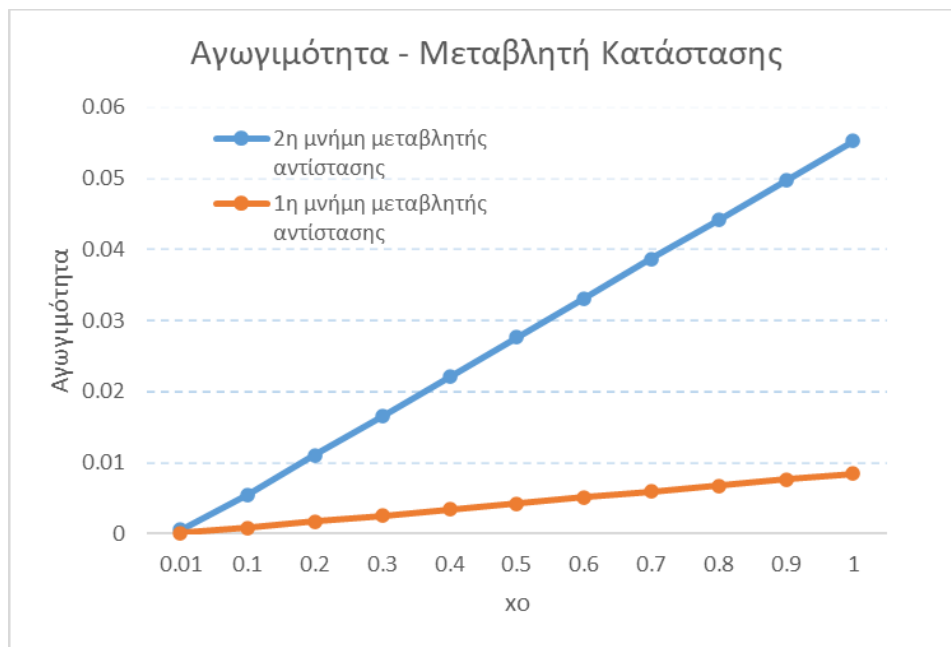
Σχήμα 40: I-V γράφημα 2^{ης} μνήμης μεταβλητής αντίστασης ($\chi_0=0.01$)

Όπως για την πρώτη μνήμη μεταβλητής αντίστασης, βρέθηκε η εξίσωση γραμμής (παλινδρόμηση) ώστε να μετατρέπεται η αγωγιμότητα στη μεταβλητή κατάσταση στη φάση της εκπαίδευσης του νευρωνικού δικτύου.

	Coefficients	standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	-5.6E-06	1.49E-06	-3.74864	0.004564	-9E-06	-2.2E-06	-9E-06	-2.2E-06
cond-300mV	18.10152	4.57E-05	395759.6	2.14E-47	18.10142	18.10162	18.10142	18.10162

Πίνακας 14: Συντελεστές μετατροπής αγωγιμότητας σε μεταβλητή κατάσταση

Συγκρίνοντας τις κλίσεις των δυο γραμμών αγωγιμότητας/μεταβλητή κατάσταση διακρίνουμε ότι η δεύτερη μνήμη μεταβλητής αντίστασης είναι κατά 6.5 φορές πιο αγωγή σε σχέση με την πρώτη.



Σχήμα 41: Σύγκριση αγωγιμότητας στην κατάσταση υψηλής αντίστασης των δυο μηνμών μεταβλητής αντίστασης

x0	1η μνήμη μεταβλητής αντίστασης αγωγιμότητα (S1)	2η μνήμη μεταβλητής αντίστασης αγωγιμότητα (S2)	S2 / S1
0.01	8.50085E-05	0.000552667	6.501312795
0.1	0.000850026	0.005524907	6.499693131
0.2	0.001700055	0.011049353	6.499409692
0.3	0.002550084	0.01657332	6.499128239
0.4	0.003400112	0.02209776	6.499127646
0.5	0.004250139	0.0276222	6.499128164
0.6	0.00510026	0.033146631	6.499008436
0.7	0.005950198	0.038671067	6.499123153
0.8	0.006800227	0.044195517	6.499124075
0.9	0.007650255	0.049719933	6.499121059
1	0.008500283	0.0552444	6.499124817

Πίνακας 15: Οι αγωγιμότητες των δυο μηνμών μεταβλητής αντίστασης και η σχέση τους

Η εκπαίδευση του δικτύματος με τη χρήση του μοντέλου της 2^{ης} μνήμης μεταβλητής αντίστασης έγινε **ακριβώς** όπως περιεγράφηκε στις προηγούμενες υπό-παραγράφους του κεφαλαίου ώστε να είναι συγκρίσιμα τα αποτελέσματα και η επίπτωση της διαφορετικής συμπεριφοράς της μνήμης μεταβλητής αντίστασης στη κατάσταση υψηλής αντίστασης. Λόγω όμως της απόλυτης γραμμικής συμπεριφοράς στην κατάσταση υψηλής αντίστασης των δυο μηνμών, παρατηρήθηκε η ίδια συμπεριφορά στην εκπαίδευση του δικτύματος με τη δεύτερη μνήμη μεταβλητής αντίστασης αλλά με αλλαγή του ρυθμού μάθησης κατά ένα συντελεστή 6.5, ο οποίος είναι και η διαφορά της αγωγιμότητας των δυο μηνμών μεταβλητής αντίστασης (η δεύτερη κατά x6.5 πιο αγωγή).

Από τα παραπάνω, η μη γραμμική συμπεριφορά της αγωγιμότητας στην κατάσταση υψηλής αντίστασης θα επιφέρει διαφορά στην εκπαίδευση του δικτύματος.

3.8 Αναφορές 3^{ου} κεφαλαίου

- [1] Chris Yakopcic, Tarek M. Taha, Robinson E. Pino, “Generalized Memristive Device SPICE Model and its Application in Circuit Design”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 32, No. 8, August 2013.
- [2] LTSpice XVII, Analog Devices [Online], <https://www.analog.com/en/design-center/design-tools-and-calculators/ltspice-simulator.html>
- [3] Chris Yakopcic, Tarek M. Taha, “Energy Efficient Perceptron Pattern Recognition Using Segmented Memristor Crossbar Arrays”, Proceedings of International Joint Conference on Neural Networks, Dallas, Texas, USA, August 4-9, 2013.
- [4] Raqibul Hasan, Chris Yakopcic, Tarek M. Taha, “Ex-situ training of large memristor crossbar for neural network applications”, Springer Nature, Published online, 29 August 2018.
- [5] Can Li, Daniel Belkin, Yunling Li, Peng Yan, Miao Hu, Ning Ge, Hao Jiang, Eric Montgomery, Peng Lin, Zhongrui Wang, Wenhao Song, John Paul Strachan, Mark Barnell, Qing Wu, R. Stanley Williams, J. Joshua Yang, and Qiangfei Xia, “Efficient and self-adaptive in-situ learning in multilayer memristor neural networks”, Nature Communications, 9:2385, 2018
- [6] Cong Wang, Zaizheng Yang, Shuang Wang, Pengfei Wang, Chen-Yu Wang, Chen Pan, Bin Cheng, Shi-Jun Liang, and Feng Miao, “A Braitenberg Vehicle Based on Memristive Neuromorphic Circuits”, Advanced Intelligent Systems, 1900103, 2019
- [7] Xiaochen Peng, Shanshi Huang, Yandong Luo, Xiaoyu Sun and Shimeng Yu, “DNN+NeuroSim: An End-to-End Benchmarking Framework for Compute-in-Memory Accelerators with Versatile Device Technologies”, IEEE Xplore, 2019
- [8] Feng Miao, John Paul Strachan, J. Joshua Yang, Min-Xian Zhang, Ilan Goldfarb, Antonio C. Torrezan, Peter Eschbach, Ronald D. Kelley, Gilberto Medeiros-Ribeiro, and R. Stanley Williams, “Anatomy of a Nanoscale Conduction Channel Reveals the Mechanism of a High-Performance Memristor”, Advance Materials, 2011.

Κεφάλαιο 4: Συμπεράσματα και μελλοντική έρευνα

Η παρούσα διπλωματική εργασία κατέδειξε τη δυνατότητα χρήσης δικτυωμάτων μνημών μεταβλητής αντίστασης σαν ΤΝΔ με ένα παράδειγμα κατηγοριοποίησης εικόνων. Η εφαρμογή της κατηγοριοποίησης υλοποιήθηκε με ένα δικτύωμα μνημών μεταβλητής αντίστασης χωρίς κρυφά επίπεδα. Το δικτύωμα αυτό είχε διαστάσεις 5x4, με τέσσερα σήματα εισόδου, τα κωδικοποιημένα σήματα των εικόνων προς κατηγοριοποίηση, ένα σήμα εισόδου για το σταθερό bias και τέσσερις εξόδους, που αντιπροσώπευαν τις τέσσερις κατηγορίες εικόνων. Η κατηγοριοποίηση αφορούσε ασπρόμαυρες εικόνες διαστάσεων 4x4 εικονοστοιχείων, όπου η κάθε κατηγορία εικόνων είχε τη μια γραμμή αρκετά πιο λευκή από τις υπόλοιπες τρεις. Η εκπαίδευση του δικτύου έγινε μέσω της οπισθοδιάδοσης, όπου κάθε ανανέωση της αγωγιμότητας της κάθε μνήμης μεταβλητής αντίστασης έγινε ανά μια εικόνα από το σύνολο δεδομένων εκπαίδευσης. Οι εικόνες εισήχθησαν στο δικτύωμα για εκπαίδευση διαδοχικά για κάθε κατηγορία χωρίς στοχαστικότητα στην επιλογή. Ο αλγόριθμος βελτιστοποίησης που χρησιμοποιήθηκε για την εφαρμογή κατηγοριοποίησης στο δικτύωμα μνημών μεταβλητής αντίστασης, χωρίς κρυφά επίπεδα, ήταν να μεγιστοποιηθεί το ρεύμα εξόδου στη σωστή κατηγορία, δηλαδή να μεγιστοποιηθεί η SoftMax έξοδος για κάθε σωστή κατηγορία. Η SoftMax ήταν η συνάρτηση ενεργοποίησης της εξόδου του δικτυώματος μνημών μεταβλητής αντίστασης.

Από τα αποτελέσματα της εκπαίδευσης ανά κύκλο εκπαίδευσης φάνηκε ο τρόπος που λειτουργεί ένα ΤΝΔ. Ανάλογα με τα στατιστικά χαρακτηριστικά των δεδομένων εισόδου, τα χαρακτηριστικά του ΤΝΔ (αριθμό επιπέδων, συναρτήσεις ενεργοποίησης, ρυθμός μάθησης και συνάρτηση σφάλματος) και της μεθόδου εκπαίδευσης, τα βάρη των συνάψεων ανανεώνονται αναλόγως. Στην εφαρμογή της παρούσας πτυχακής φάνηκε από την προσομοίωση ότι το δικτύωμα χρειάστηκε 16 κύκλους εκπαίδευσης ώστε να εκπαιδευτεί και να μπορεί να κατηγοριοποιεί με 100% ακρίβεια τις εικόνες του συνόλου δεδομένων της εκπαίδευσης. Η ίδια ακρίβεια θα μπορούσε να πραγματοποιηθεί πιο γρήγορα αν είχε επιλεγεί στοχαστικότητα στην επιλογή των εικόνων εισόδου για την εκπαίδευση καθώς και αν είχε επιλεγεί να χρησιμοποιηθούν όλες οι εικόνες εισόδου (batch) για τον προσδιορισμό μιας ανανέωσης των αγωγιμοτήτων των μνημών μεταβλητής αντίστασης. Από τα γραφήματα των μεταβολών των αγωγιμοτήτων της κάθε μνήμης μεταβλητής αντίστασης ανά κύκλο εκπαίδευσης φαίνεται ποια μνήμη μεταβλητής αντίστασης συνεισφέρει περισσότερο αλλά και με ποιόν τον τρόπο (αυξανόμενη ή μειούμενη αγωγιμότητα) στην κατηγοριοποίηση της κάθε εικόνας. Για την σωστή κατηγοριοποίηση της πρώτης κατηγορίας εικόνων, οι μνήμες μεταβλητής αντίστασης U1 και U5 (σχήμα 28 η αναπαράσταση/ονοματολογία των μνημών) αυξάνουν την αγωγιμότητά τους ενώ οι U6 και U16 την μειώνουν. Αντίστοιχα για την δεύτερη κατηγορία, οι U7, U10 και U11 αυξάνουν την αγωγιμότητά τους ενώ οι U2, U12 και U17 την μειώνουν, για την τρίτη κατηγορία οι U13 και U15 αυξάνουν την αγωγιμότητά τους ενώ οι U3, U8 και U18 την μειώνουν και για την τέταρτη κατηγορία οι U19 και U20 αυξάνουν την αγωγιμότητά τους ενώ οι U4, U9 και U14 την μειώνουν.

Για να γίνει επαλήθευση της γενίκευσης της σωστής κατηγοριοποίησης του εκπαιδευμένου δικτυώματος μνημών μεταβλητής αντίστασης χρησιμοποιήθηκε ένα σύνολο δεδομένων ελέγχου, που ήταν εικόνες που δεν είχε προηγουμένως εκτεθεί το δικτύωμα. Το αποτέλεσμα κατηγοριοποίησης του συνόλου δεδομένων ελέγχου ήταν 100% ακρίβεια κατηγοριοποίησης κάτι που απέδειξε ότι το εκπαιδευμένο δικτύωμα μπορεί να γενικεύσει σε άγνωστα δεδομένα.

Η χρήση του ελεύθερου λογισμικού LTSpice με κώδικα μοντέλων μνημών μεταβλητής αντίστασης που αποτυπώνουν σωστά τη συμπεριφορά των φυσικών διατάξεων δίνουν τη δυνατότητα έρευνας αρχιτεκτονικών και υλικών ως προπομπό για τη φυσική υλοποίηση συστημάτων. Από τη χρήση του λογισμικού LTSpice φάνηκε η ευκολία χρήσης του καθώς και η ευκολία εισαγωγής διάφορων μοντέλων διατάξεων για δικτυώματα μικρών διαστάσεων. Όσο όμως αυξάνονται οι διαστάσεις και τα επίπεδα των δικτυωμάτων προς προσομοίωση, θα πρέπει να χρησιμοποιηθεί το λογισμικό MATLAB σε συνδυασμό με το LTSpice για αυτοματισμό. Ενδεικτικά, ένα δικτύωμα διαστάσεων 100x100 για εκπαίδευση με αντίστοιχων διαστάσεων δεδομένα, χρειάζεται σχεδόν μια ολόκληρη μέρα. Σε αυτές τις περιπτώσεις προκρίνεται η χρήση μόνο του λογισμικού Matlab, που προσεγγίζει την ακρίβεια των αποτελεσμάτων του LTSpice και δίνει μια πολύ καλή ένδειξη της συμπεριφοράς του δικτυώματος σε πολύ λίγο χρόνο (λεπτά). Για μεγαλύτερη ακρίβεια στα ηλεκτρικά αποτελέσματα την προσομοίωσης από το MATLAB καθώς και μεγαλύτερη ταχύτητα εκτέλεσης της προσομοίωσης σε σχέση με το LTSpice, επιβάλλεται η χρήση του λογισμικού προσομοίωσης ανοικτού κώδικα Xyce της Sandia National Laboratories (<https://xyce.sandia.gov/>), που εκμεταλλεύεται την παράλληλη επεξεργασία δεδομένων.

Μελλοντικά πεδία έρευνας με βάση τη παρούσα πτυχιακή θα μπορούσαν να είναι η δημιουργία και προσομοίωση πιο σύνθετων δικτυωμάτων σε συνδυασμό με τη χρήση μοντέλων μνημών μεταβλητής αντίστασης διατάξεων του Εργαστηρίου Νανοϋλικών και Διατάξεων του Εθνικού Μετσόβιου Πολυτεχνείου. Τα πιο σύνθετα δικτυώματα θα μπορούσαν να είναι με συνάψεις υλοποιημένες με διπλή μνήμη μεταβλητής αντίστασης, με χρήση επιλογέν (τρανζίστορ ή διόδων) σε σειρά με τις μνήμες μεταβλητής αντίστασης για την αποφυγή ρευμάτων ερπυσμού καθώς και με περισσότερα του ενός επίπεδα για εφαρμογές με πιο σύνθετα δεδομένα όπως οι εικόνες MNIST. Ειδικά προσομοιώσεις με εφαρμογές για το σύνολο δεδομένων MNIST έχει το πλεονέκτημα ότι όποια αποτελέσματα προκύψουν από τις προσομοιώσεις να είναι άμεσα συγκρίσιμα με άλλες έρευνες. Παράλληλα με τις παραλλαγές στην αρχιτεκτονική των δικτυωμάτων και τα είδη των συνάψεων, έρευνα στην δημιουργία μοντέλων LTSpice ή/και MATLAB για τις μνήμες μεταβλητής αντίστασης του Εργαστηρίου Νανοϋλικών και Διατάξεων μπορεί να γίνει και τα μοντέλα αυτά να χρησιμοποιηθούν στις διάφορες, προς έρευνα, εφαρμογές. Οι εφαρμογές αυτές μπορεί να είναι κλασικά ΤΝΔ όπως αναφέρονται παραπάνω και τα αποτελέσματα των προσομοιώσεων να συγκριθούν με άλλες διατάξεις σε επίπεδα ταχύτητας εκπαίδευσης, ακρίβειας του στόχου λειτουργίας του δικτυώματος καθώς και συμπεριφοράς στα ρεύματα ερπυσμού με εύρεση του μέγιστου δυνατού λειτουργικού δικτυώματος χωρίς τη χρήση επιλογέν. Επιπρόσθετα, με δεδομένα από διάφορους αισθητήρες, θα μπορεί να γίνει έρευνα στον συνδυασμό αισθητήρων με δικτυώματα μνημών μεταβλητής αντίστασης για την προσομοίωση και υλοποίηση έξυπνων αυτόνομων συστημάτων.

Τα μοντέλα των διατάξεων του Εργαστηρίου Νανοϋλικών και Διατάξεων μπορούν επίσης να χρησιμοποιηθούν και σε αρχιτεκτονικές ΤΝΔ αιχμών και να γίνει έρευνα στις εφαρμογές αυτού του χώρου.

Παράρτημα 1^ο : Spice κώδικας της μνήμης μεταβλητής αντίστασης

```
* SPICE model for memristive devices
* Created by Chris Yakopcic
* Last Update: 8/9/2011
*
* Connections:
* TE - top electrode
* BE - bottom electrode
* XSV - External connection to plot state variable
* that is not used otherwise

.subckt Memristor TE BE XSV

* Fitting parameters to model different devices
* a1, a2, b:      Parameters for IV relationship
* Vp, Vn:        Pos. and neg. voltage thresholds
* Ap, An:        Multiplier for SV motion intensity
* xp, xn:        Points where SV motion is reduced
* alphap, alphan: Rate at which SV motion decays
* xo:            Initial value of SV
* eta:           SV direction relative to voltage

*.params steve=0.5
.params xo={steve}

.params a1=0.17 a2=0.17 b=0.05 Vp=0.6 Vn=0.5 Ap=4000
+An=4000 xp=0.5 xn=0.5 alphap=1 alphan=5 eta=1

* Multiplicative functions to ensure zero state
* variable motion at memristor boundaries
.func wp(V) = xp/(1-xp) - V/(1-xp) + 1
.func wn(V) = V/(1-xn)

* Function G(V(t)) - Describes the device threshold
.func G(V) = IF(V <= Vp, IF(V >= -Vn, 0, -An*(exp(-
+V)-exp(Vn))), Ap*(exp(V)-exp(Vp)))

* Function F(V(t),x(t)) - Describes the SV motion
.func F(V1,V2) = IF(eta*V1 >= 0, IF(V2 >= xp, exp(-
+alphap*(V2-xp))*wp(V2) ,1), IF(V2 <= (1-xn),
+exp(alphan*(V2+xn-1))*wn(V2) ,1))

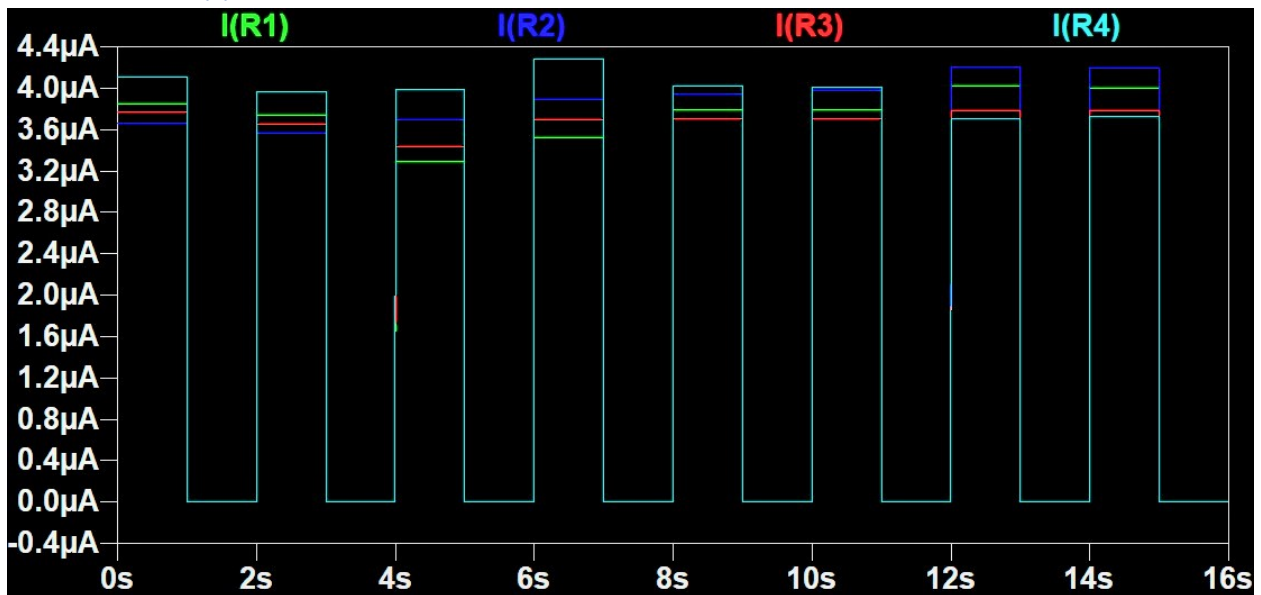
* IV Response - Hyperbolic sine due to MIM structure
.func IVRel(V1,V2) = IF(V1 >= 0, a1*V2*sinh(b*V1),
+a2*V2*sinh(b*V1) )

* Circuit to determine state variable
* dx/dt = F(V(t),x(t))*G(V(t))
Cx XSV 0 {1}
.ic V(XSV) = xo
Gx 0 XSV value={eta*F(V(TE,BE),V(XSV,0))*G(V(TE,BE))}
* Current source for memristor IV response
Gm TE BE value = {IVRel(V(TE,BE),V(XSV,0))}

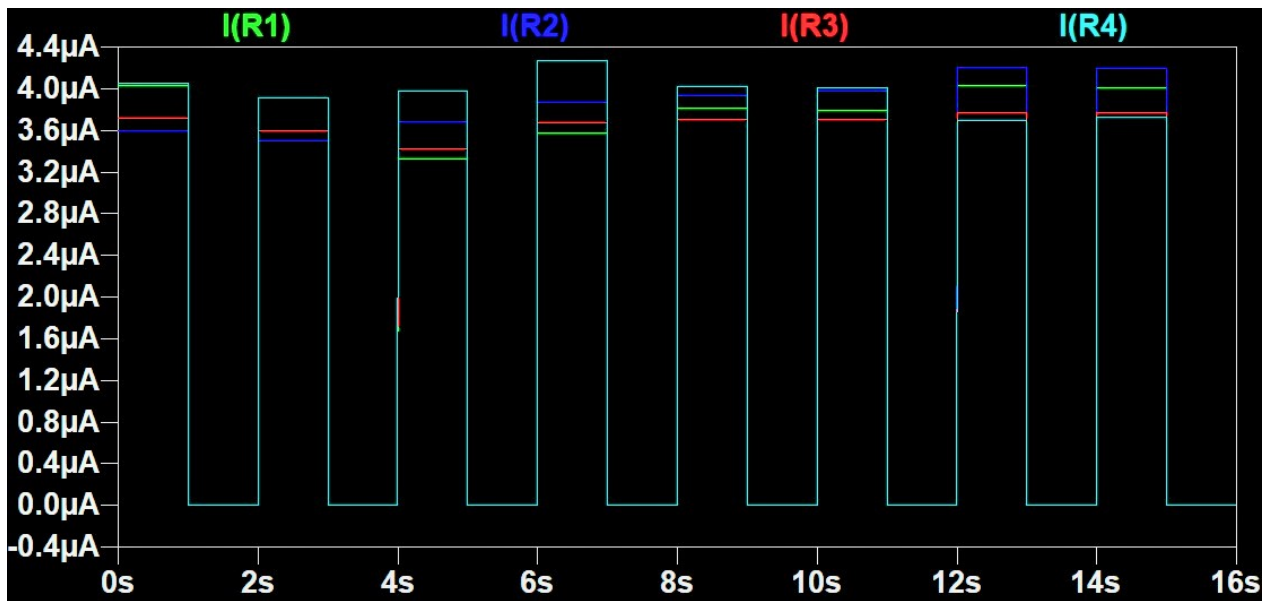
.ends Memristor
```

Σχήμα 42: Κώδικας Spice για το μοντέλο Yakopcic

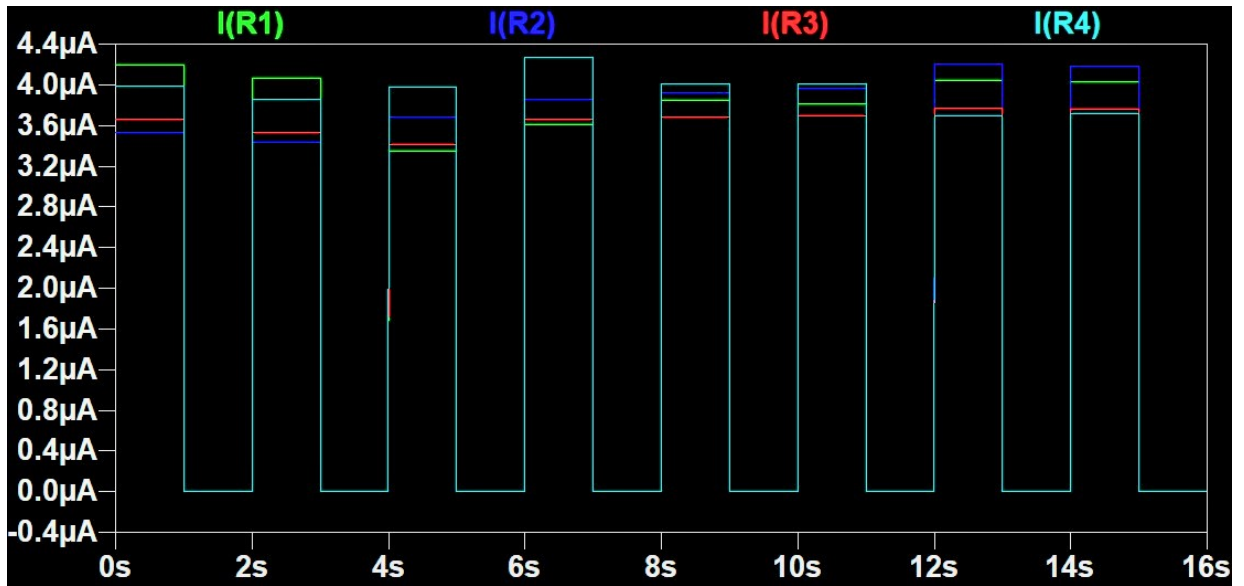
Παράρτημα 2^ο : Ρεύματα εξόδου νευρωνικού δικτυώματος ανά κύκλο εκπαίδευσης



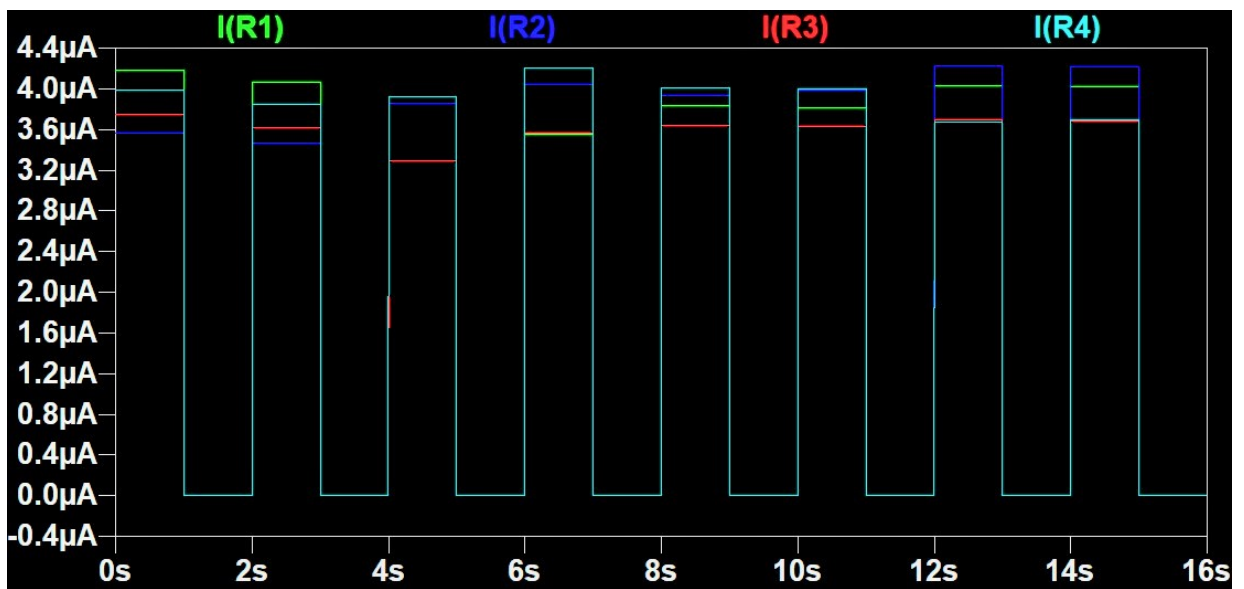
Σχήμα 43: Αρχική έξοδος του δικτυώματος με αρχικές μεταβλητές καταστάσεων



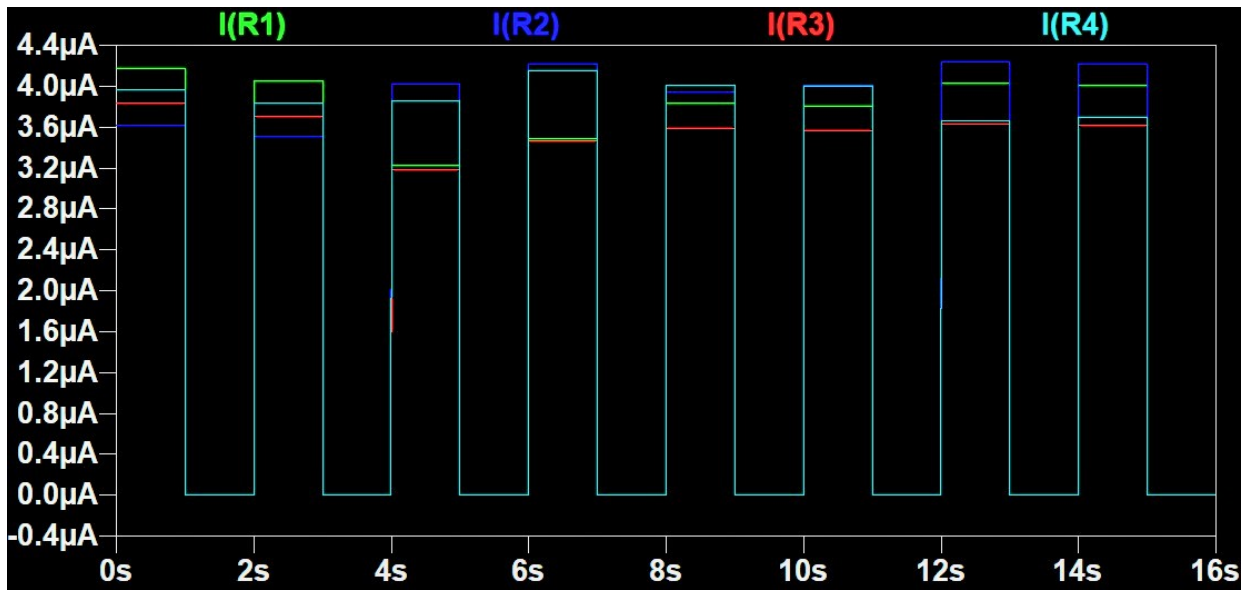
Σχήμα 44: Έξοδος του δικτυώματος μετά από 1 κύκλο εκπαίδευσης



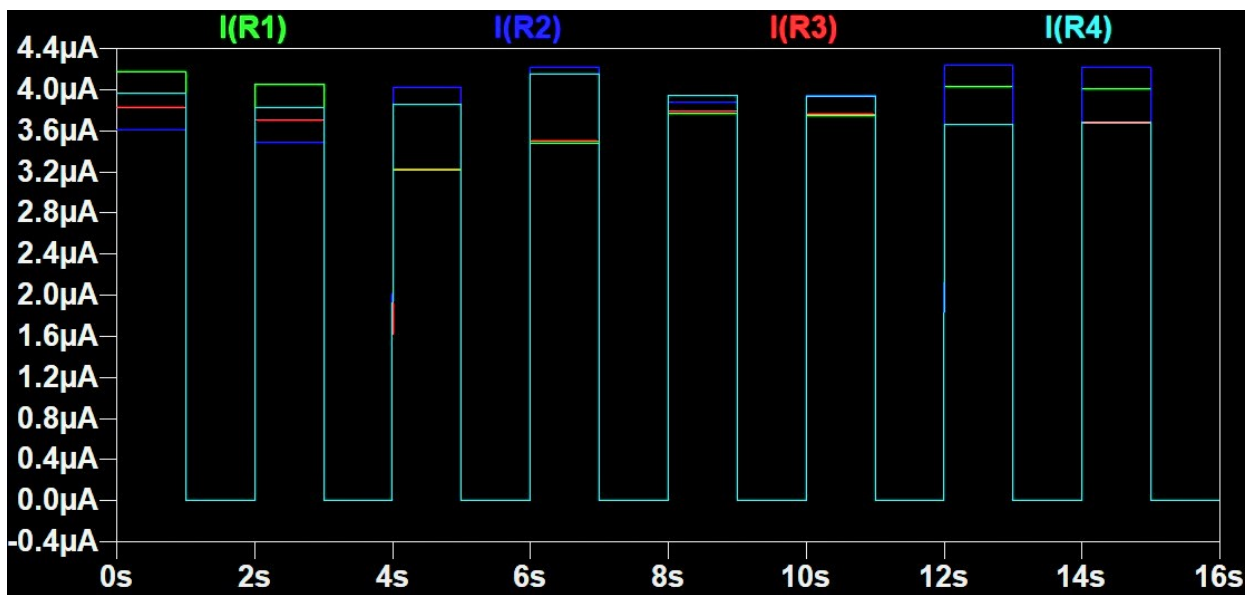
Σχήμα 45: Έξοδος του δικτύωματος μετά από 2 κύκλους εκπαίδευσης



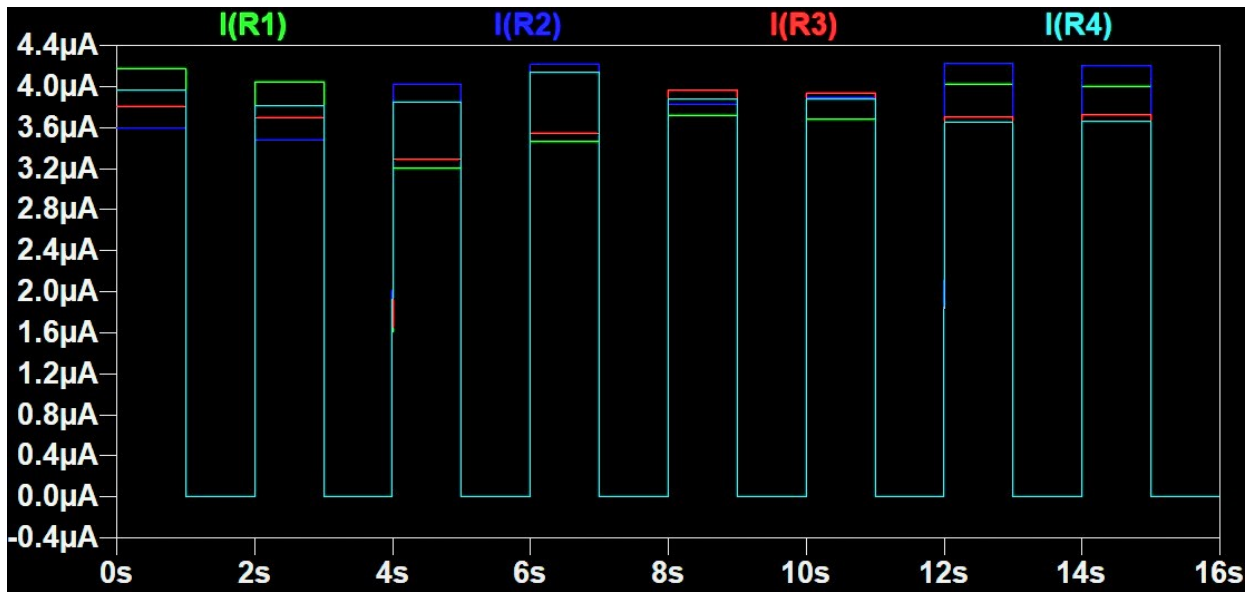
Σχήμα 46: Έξοδος του δικτύωματος μετά από 3 κύκλους εκπαίδευσης



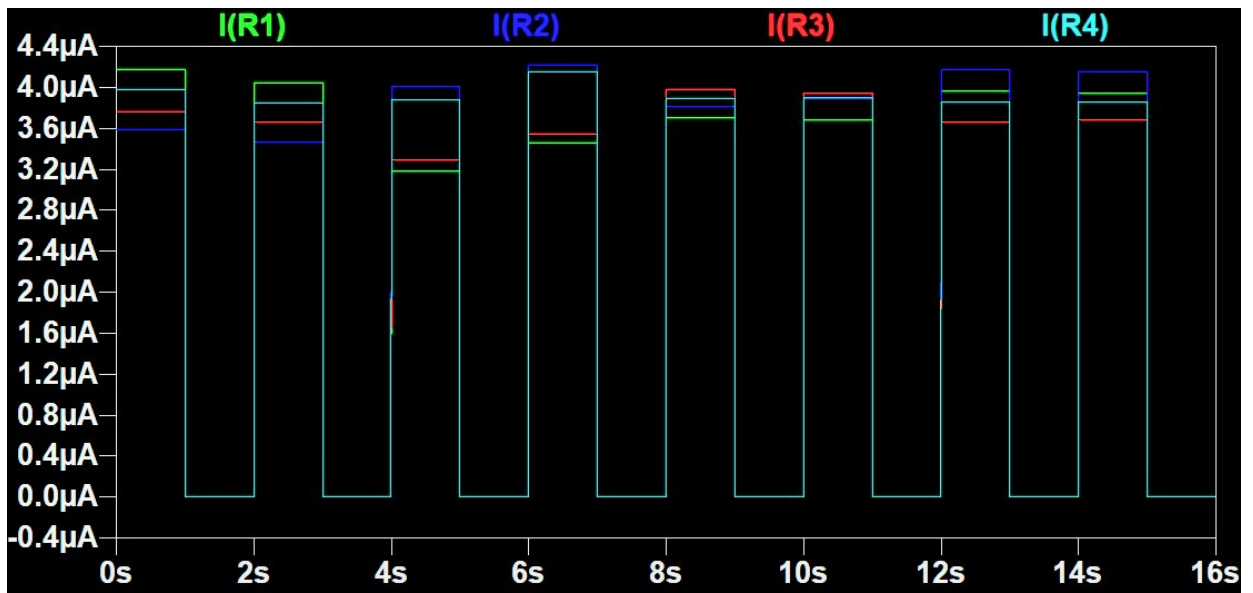
Σχήμα 47: Έξοδος του δικτύματος μετά από 4 κύκλους εκπαίδευσης



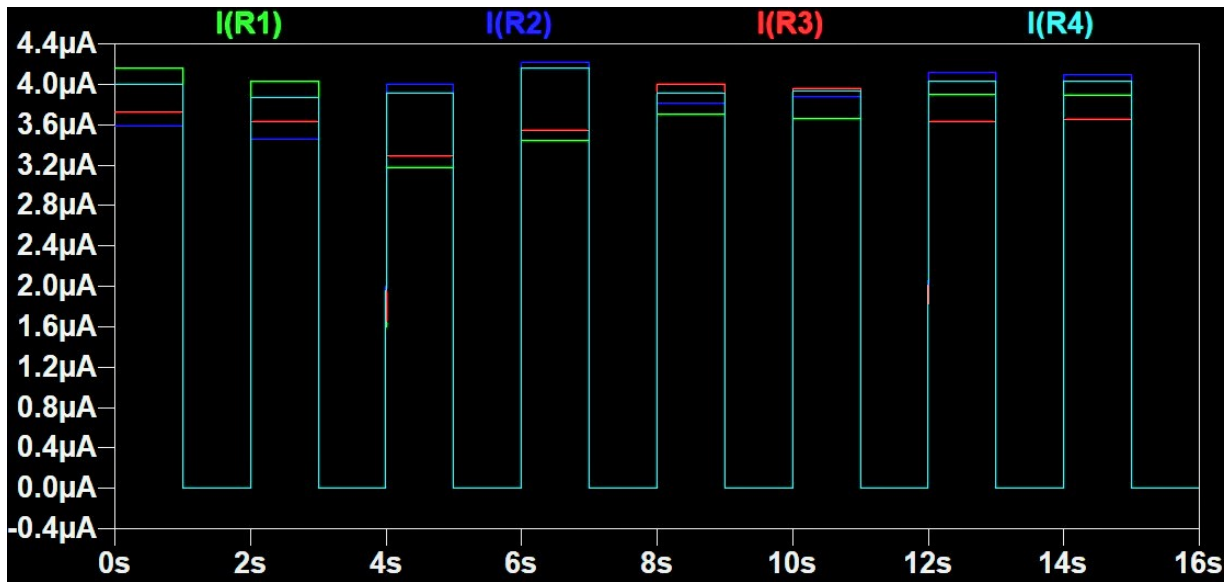
Σχήμα 48: Έξοδος του δικτύματος μετά από 5 κύκλους εκπαίδευσης



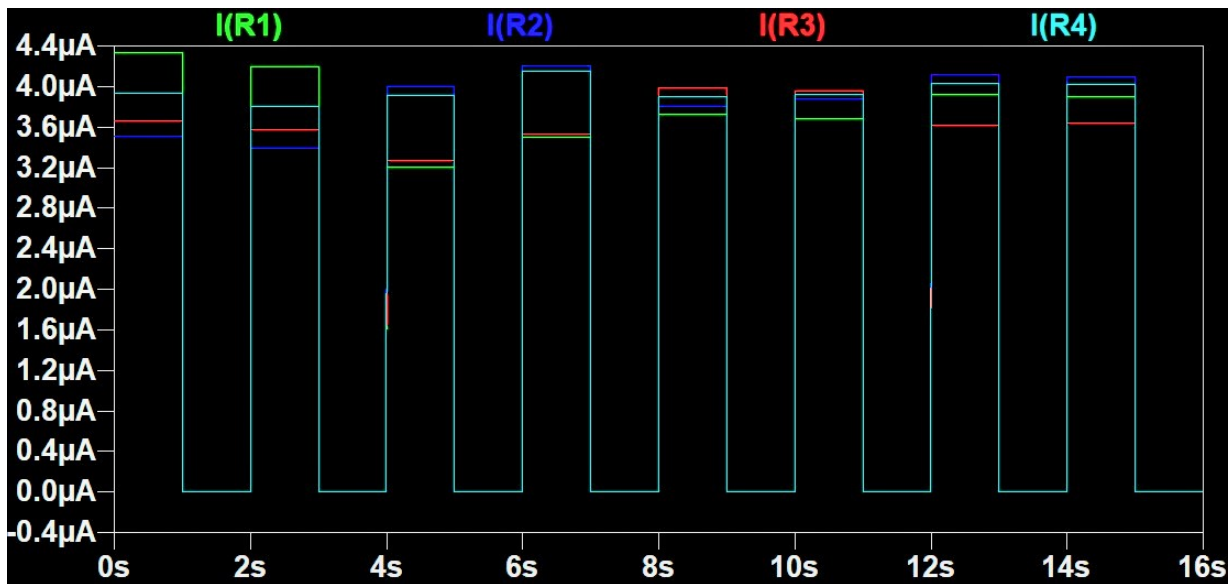
Σχήμα 49: Έξοδος του δικτυώματος μετά από 6 κύκλους εκπαίδευσης



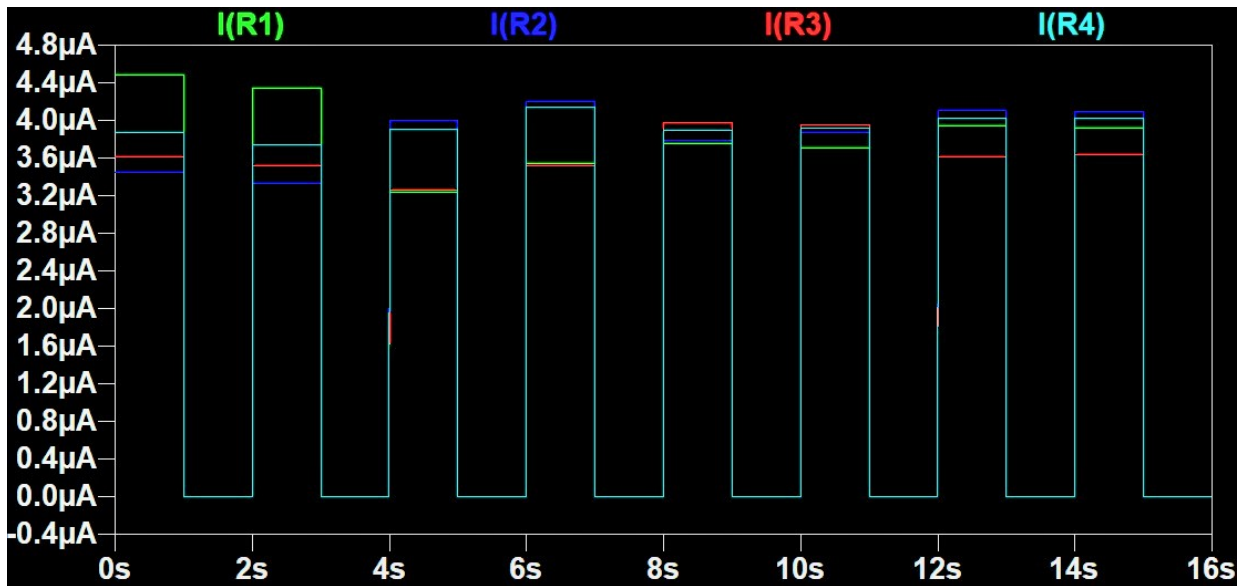
Σχήμα 50: Έξοδος του δικτυώματος μετά από 7 κύκλους εκπαίδευσης



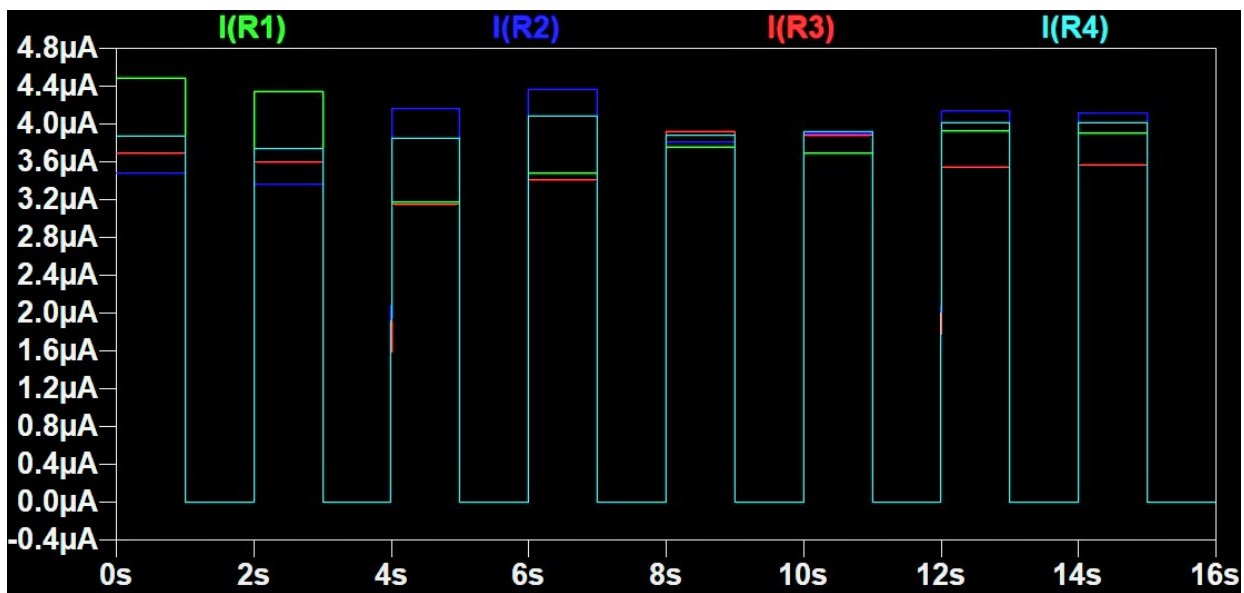
Σχήμα 51: Έξοδος του δικτύματος μετά από 8 κύκλους εκπαίδευσης



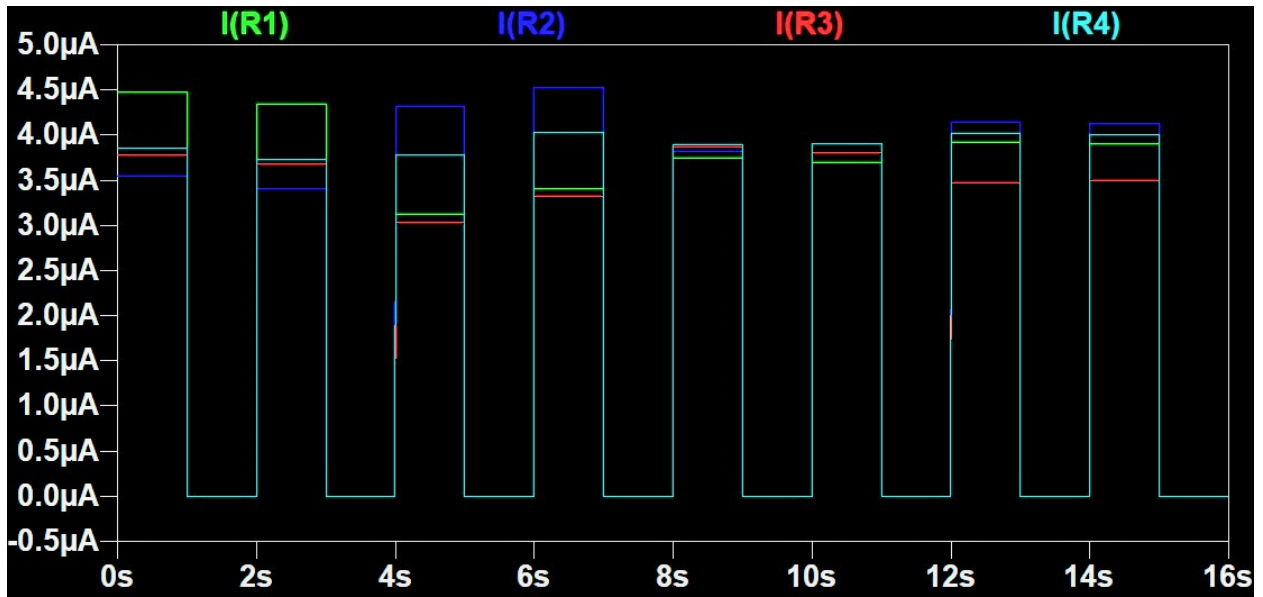
Σχήμα 52: Έξοδος του δικτύματος μετά από 9 κύκλους εκπαίδευσης



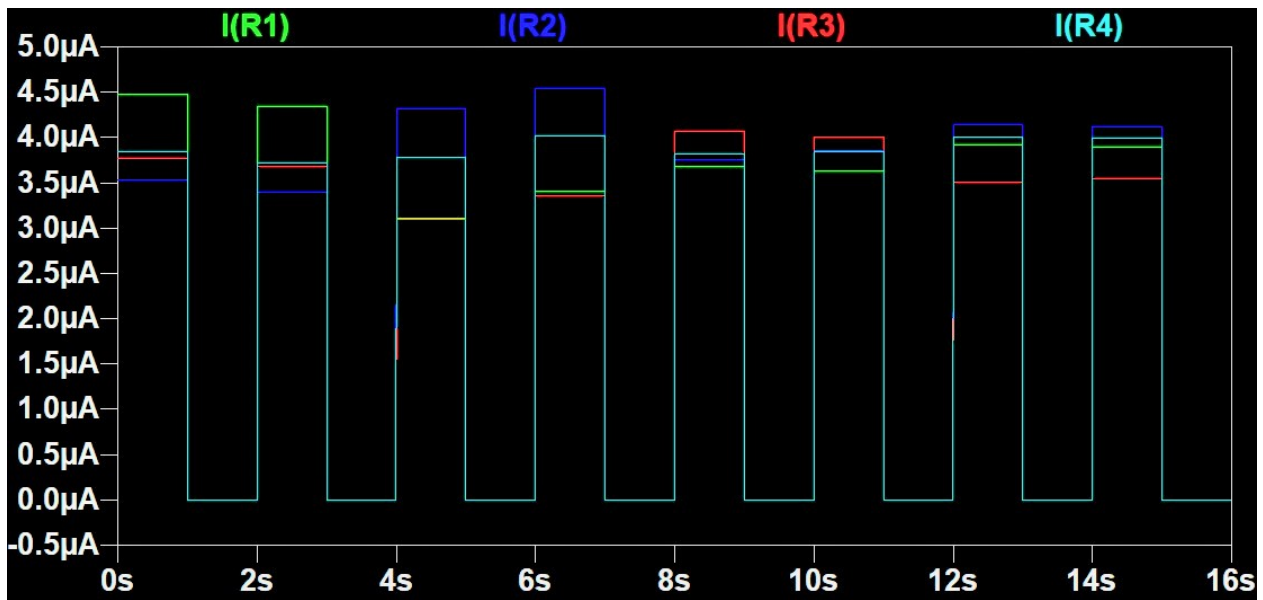
Σχήμα 53: Έξοδος του δικτύωματος μετά από 10 κύκλους εκπαίδευσης



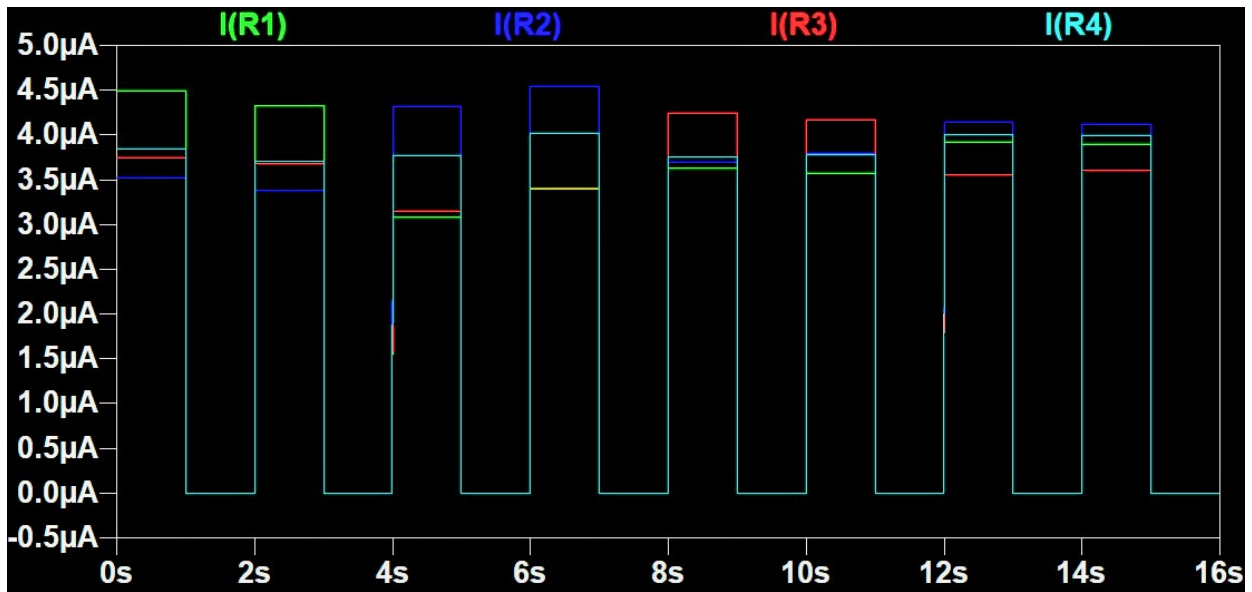
Σχήμα 54: Έξοδος του δικτύωματος μετά από 11 κύκλους εκπαίδευσης



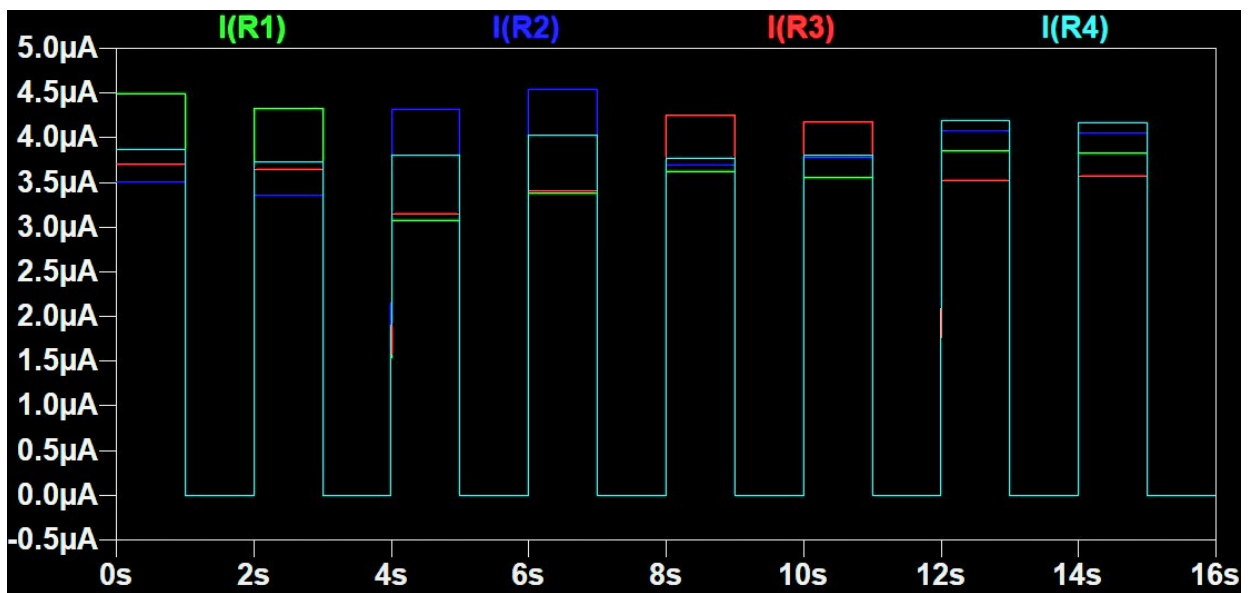
Σχήμα 55: Έξοδος του δικτυώματος μετά από 12 κύκλους εκπαίδευσης



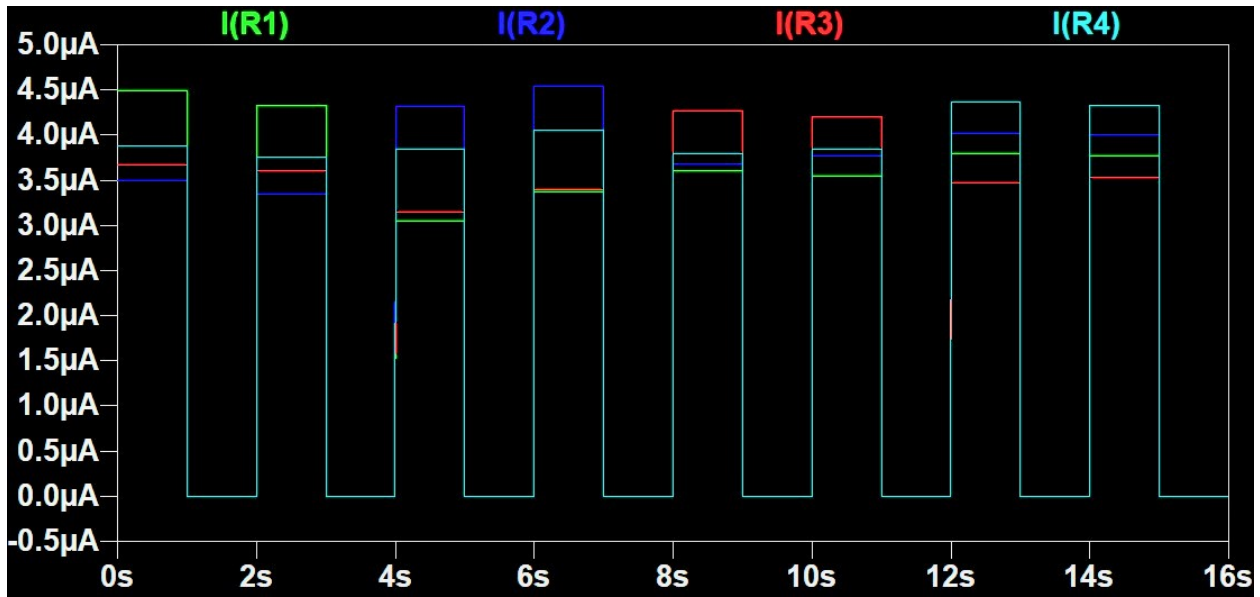
Σχήμα 56: Έξοδος του δικτυώματος μετά από 13 κύκλους εκπαίδευσης



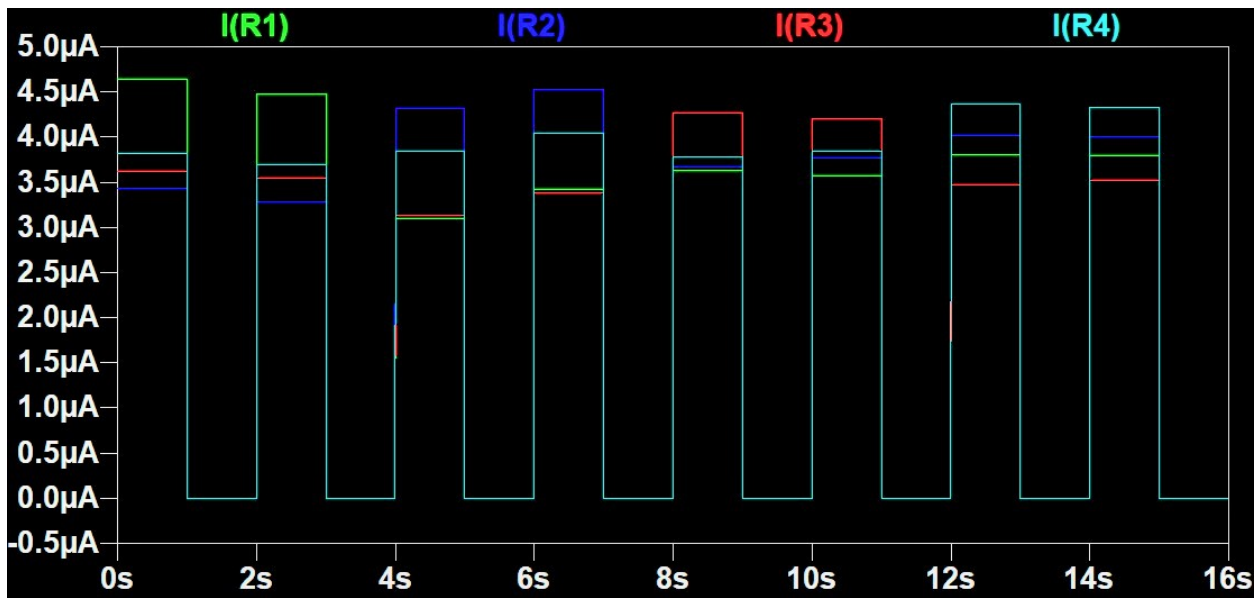
Σχήμα 57: Έξοδος του δικτύματος μετά από 14 κύκλους εκπαίδευσης



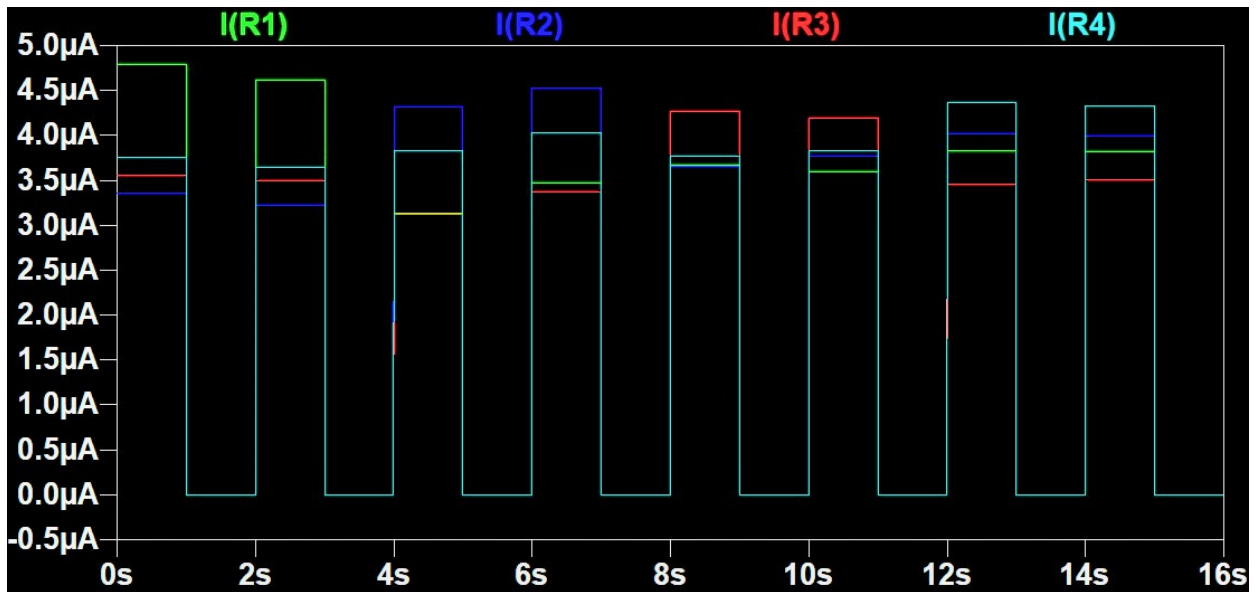
Σχήμα 58: Έξοδος του δικτύματος μετά από 15 κύκλους εκπαίδευσης



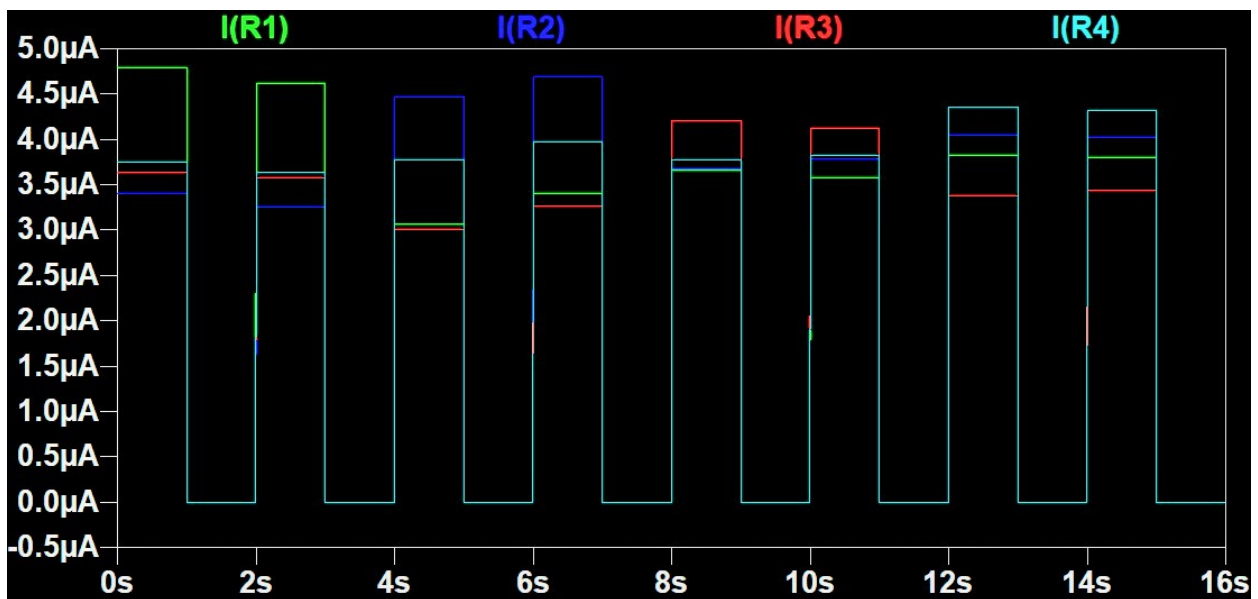
Σχήμα 59: Έξοδος του δικτύωματος μετά από 16 κύκλους εκπαίδευσης



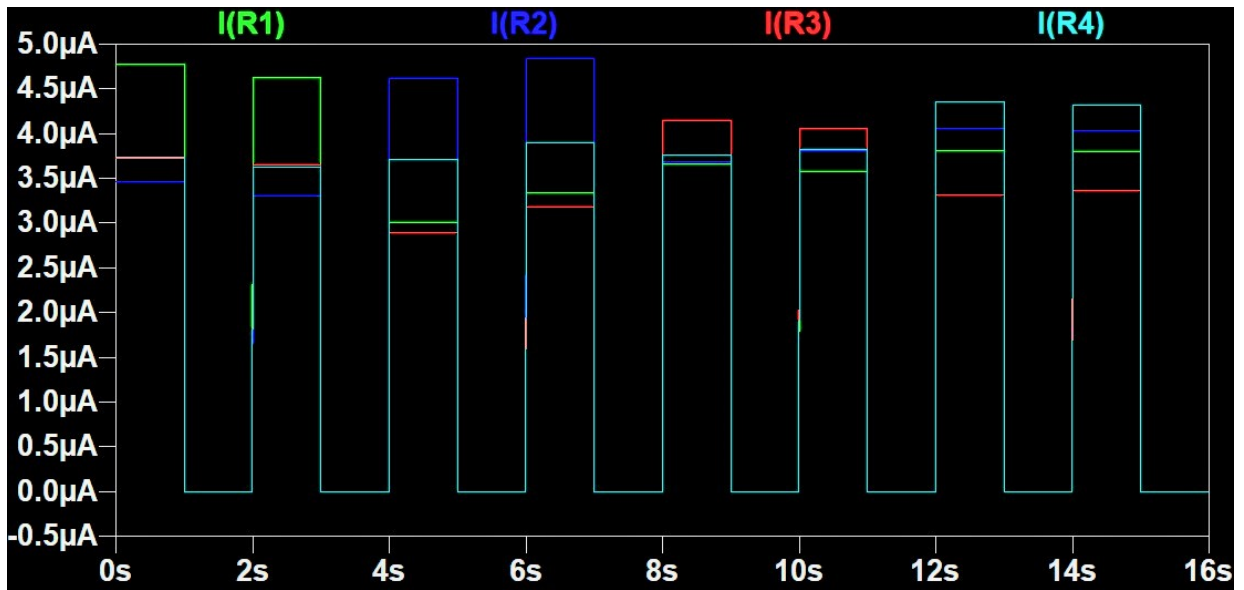
Σχήμα 60: Έξοδος του δικτύωματος μετά από 17 κύκλους εκπαίδευσης



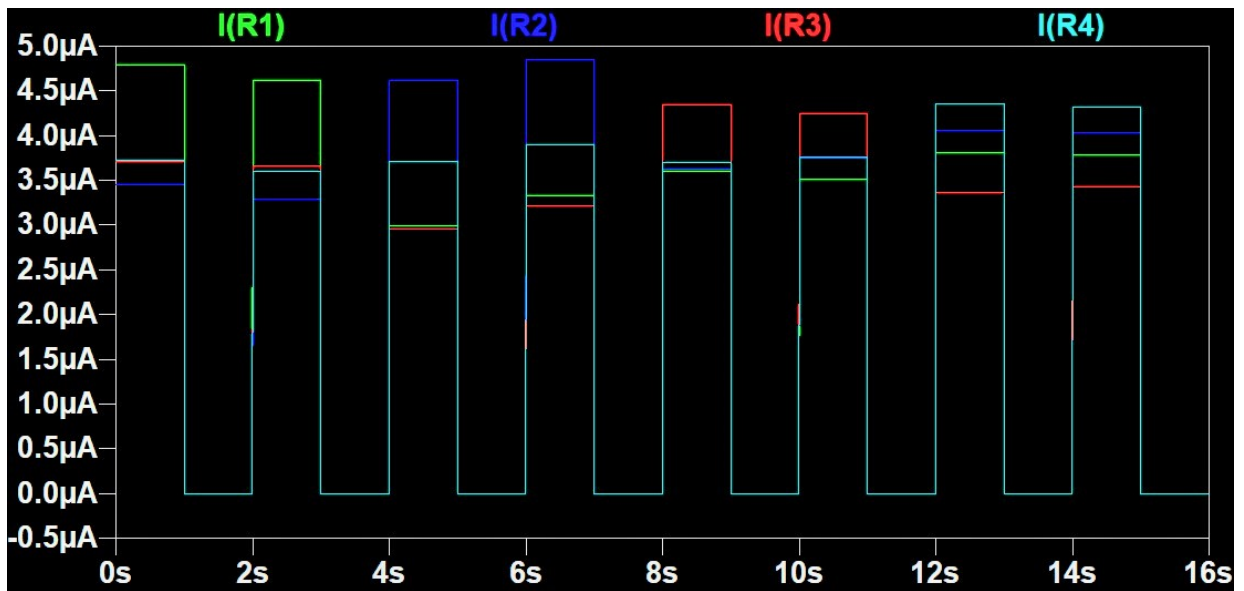
Σχήμα 61: Έξοδος του δικτύματος μετά από 18 κύκλους εκπαίδευσης



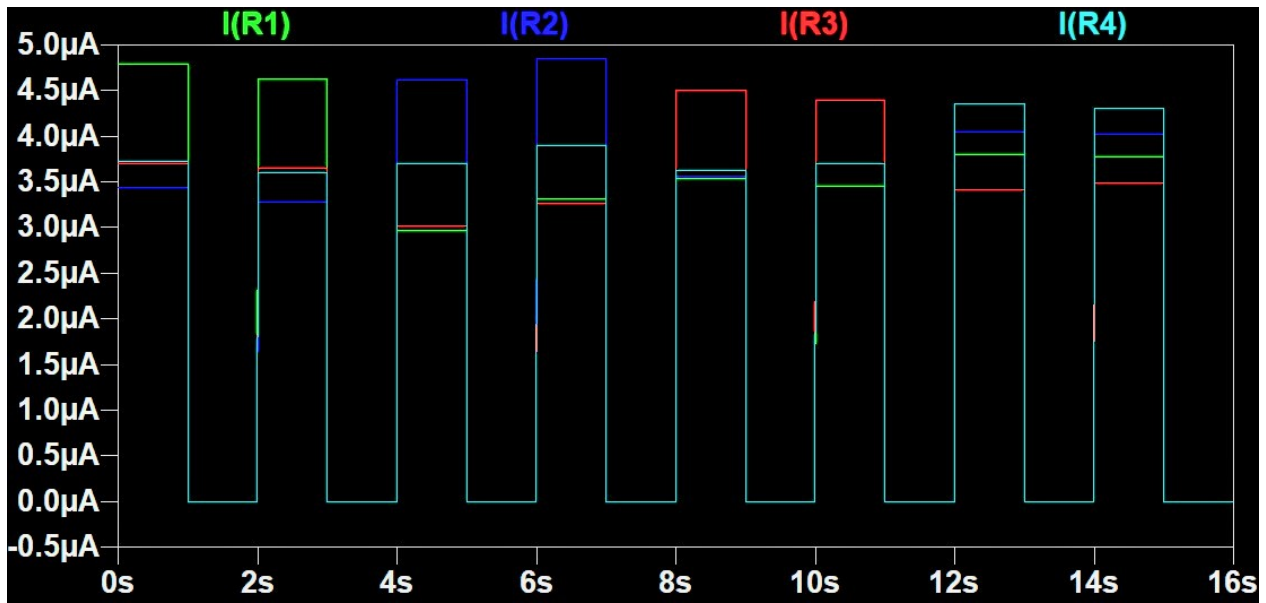
Σχήμα 62: Έξοδος του δικτύματος μετά από 19 κύκλους εκπαίδευσης



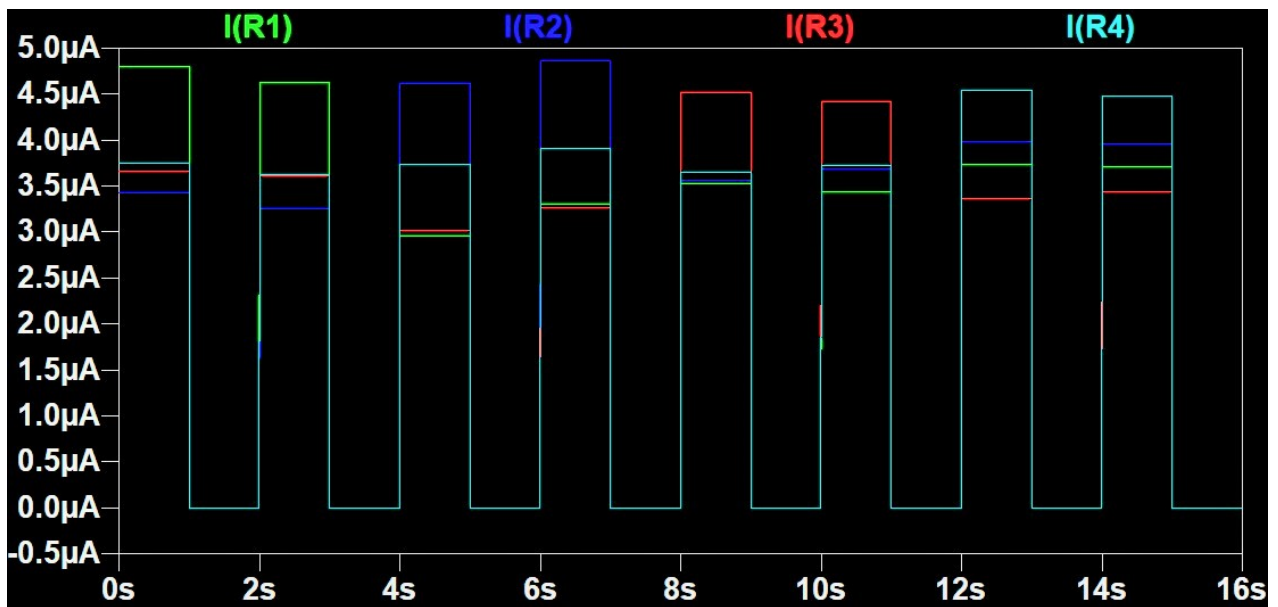
Σχήμα 63: Έξοδος του δικτύματος μετά από 20 κύκλους εκπαίδευσης



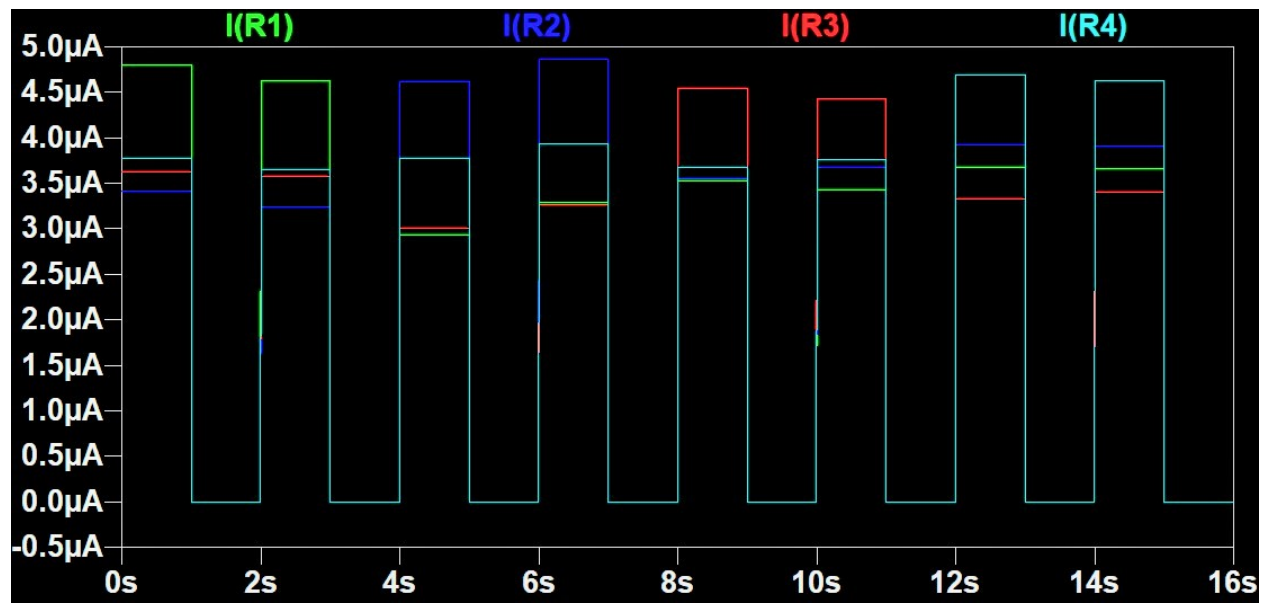
Σχήμα 64: Έξοδος του δικτύματος μετά από 21 κύκλους εκπαίδευσης



Σχήμα 65: Έξοδος του δικτυώματος μετά από 22 κύκλους εκπαίδευσης



Σχήμα 66: Έξοδος του δικτυώματος μετά από 23 κύκλους εκπαίδευσης



Σχήμα 67: Έξοδος του δικτύωματος μετά από 24 κύκλους εκπαίδευσης