



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**ΜΝΗΜΕΣ ΤΥΧΑΙΑΣ ΠΡΟΣΠΕΛΑΣΗΣ ΜΕ ΒΑΣΗ ΤΗΝ ΑΝΤΙΣΤΑΣΗ
ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΕ ΕΝΤΟΣ ΜΝΗΜΗ ΥΠΟΛΟΓΙΣΤΙΚΗ**

Διπλωματική Εργασία

Χαλυβδόπουλος Σπυρίδων

Επιβλέπων: Ευμορφόπουλος Νέστωρας

Ιούλιος 2024



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

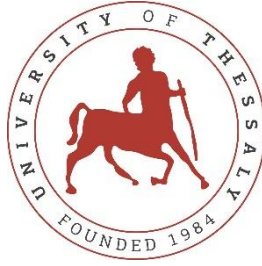
**ΜΝΗΜΕΣ ΤΥΧΑΙΑΣ ΠΡΟΣΠΕΛΑΣΗΣ ΜΕ ΒΑΣΗ ΤΗΝ ΑΝΤΙΣΤΑΣΗ
ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΕ ΕΝΤΟΣ ΜΝΗΜΗ ΥΠΟΛΟΓΙΣΤΙΚΗ**

Διπλωματική Εργασία

Χαλυβδόπουλος Σπυρίδων

Επιβλέπων: : Ευμορφόπουλος Νέστωρας

Ιούλιος 2024



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

**RESISTIVE RANDOM ACCESS MEMORY
AND IN MEMORY COMPUTING APPLICATIONS**

Diploma Thesis

Chalyvdopoylos Spyridon

Supervisor: Evmorfopoulos Nestoras

July 2024

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων

Ευμορφόπουλος Νέστωρας

Αναπληρωτής Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και
Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

Σταμούλης Γεώργιος

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

Πλέσσας Φώτιος

Αναπληρωτής Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και
Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελούν αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλουν οποιασδήποτε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχουν έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο Δηλών

Χαλυβδόπουλος Σπυρίδων

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism.

The Declarant

Chalyvdopoylos Spyridon

ΜΝΗΜΕΣ ΤΥΧΑΙΑΣ ΠΡΟΣΠΕΛΑΣΗΣ ΜΕ ΒΑΣΗ ΤΗΝ ΑΝΤΙΣΤΑΣΗ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΕ ΕΝΤΟΣ ΜΝΗΜΗ ΥΠΟΛΟΓΙΣΤΙΚΗ

Χαλυβδόπουλος Σπυρίδων

ΠΕΡΙΛΗΞΗ

Στην παρούσα εργασία μελετώνται οι μνήμες RE-RAM και όλες οι εφαρμογές αυτών. Αρχικά αφού κατανοηθεί η έννοια της μνήμης RAM και η αρχή λειτουργίας διάφορων ειδών RAM, επικεντρώνονται το ενδιαφέρον σε ένα συγκεκριμένο τύπο RAM την RE-RAM. Για την RE-RAM αναφέρεται μια ιστορική αναδρομή, ο εξειδικευμένος μηχανισμός λειτουργίας της και οι χρήσεις της. Σε επόμενα κεφάλαια κατηγοριοποιούνται οι τύποι των RE-RAM και η συμβολή τους στην πράξη πινάκων. Σκοπός της συγκεκριμένης βιβλιογραφικής εργασίας είναι η ανάδειξη της χρήσης του συγκεκριμένου τύπου μνήμης σε εφαρμογές γραμμικής άλγεβρας, όπως ο πολλαπλασιασμός πινάκων. Τα νευρωνικά δίκτυα μια ταχέως αναπτυσσόμενη τεχνολογία χρησιμοποιούν κατά κύριο λόγο RE-RAM για την μεταφορά δεδομένων κάτι που καθιστά τις RE-RAM ως την καλύτερη επιλογή παρά τα υπάρχοντα προβλήματα στην λειτουργία τους τα οποία θα πρέπει να αντιμετωπιστούν. Τέλος, αναφέρονται τα πλεονεκτήματα και μειονεκτήματά τους και γίνεται μια σφαιρική συζήτηση των αποτελεσμάτων. Αφήνεται ως πεδίο περαιτέρω μελλοντικής μελέτης η βελτίωση των μειονεκτημάτων τους και η των τεχνικών παραγωγής τους.

Λέξεις-κλειδιά: RAM, RE-RAM, Switching mechanism, DNN Accelerators, In Memory Computing

RESISTIVE RANDOM ACCESS MEMORY AND IN MEMORY COMPUTING APPLICATION

Chalyvdopoylos Spyridon

Abstract

This paper deals with RERAM memories and their applications. Initially, after understanding the concept of RAM and the principle of operation of various types of RAMS, attention is focused on a specific type of RAM, RE-RAM. A historical review is given for RE-RAM, its specialized operating mechanism, and its uses. In subsequent chapters, the types of RE-RAMS and their contribution to array operation are categorized.

The purpose of this bibliographic work is to highlight the use of this memory in linear algebra applications, such as matrix multiplication. After all, neural networks, a rapidly developing technology, primarily use RERAM for data transfer, which makes RERAM the best choice despite existing performance issues that should be addressed.

Finally, their advantages and disadvantages are mentioned, and a global discussion of the results is made. The improvement of their disadvantages and of their production techniques is left as a field of further future study.

Keywords: RAM, RE-RAM, Switching mechanism, DNN Accelerators, In Memory Computing

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη	xv
Abstract	xvii
Κεφάλαιο 1 Εισαγωγή.....	22
1.1 Η μνήμη RAM	22
1.2 Η σημασία δημιουργίας νέων RAM	22
1.3 RE-RAM και γιατί την επιλέξαμε	23
1.4 Οργάνωση του τόμου	24
Κεφάλαιο 2 Τεχνολογική βάση των RE-RAM	25
2.1 Τι είναι η RE-RAM	25
2.2 Ιστορική αναδρομή	25
2.3 Εσωτερική κατασκευή και λειτουργία μιας RE-RAM	26
2.4 Switching mechanism-Μηχανισμός αλλαγής	27
2.4.1 CMOS BEOL	28
2.5 Ιδιότητες και Χρήσεις των RE-RAMS	29
Κεφάλαιο 3 Κατηγορίες και Τύποι RE-RAM	31
3.1 Είδη των RE-RAMS	31
3.2 Anion-type RE-RAM	31
3.3 Cation-type RE-RAM	33
3.4 Carbon-based RE-RAM	34
3.5 Oxide-based electrode RE-RAM	35
Κεφάλαιο 4 Εφαρμογές για In-Memory Computing	37
4.1 Τι είναι το in memory computing	37
4.2 Πράξεις πινάκων με in memory computing	39
4.2.1 Επαναχρησιμοποίηση μοτίβου δεδομένων στις στρώσεις συνέλιξης (convolution layers)	40
4.2.2 RE-RAM based DNN Accelerators	41

Κεφάλαιο 5 Αξιολόγηση RE-RAM	42
3.1 Πλεονεκτήματα των RE-RAM	42
3.2 Μειονεκτήματα των RE-RAM	43
Συμπεράσματα	45
Βιβλιογραφία	46

Πίνακας εικόνων

Εικόνα 1 : Χαρακτηριστική δομή RE-RAM,(Amit Prakash,Hyunsang hwang)	
Εικόνα 2 : Σχηματική αναπαράσταση τομής TaOx διάταξης που αποτελείται από μια λεπτή Ta ₂ O ₅ -x μονωτική στρώση και μια στρώση βάσης TaO ₂ -x	
Εικόνα 3 : Anion type RE-RAM,(Ji-Hyun Hur)	
Εικόνα 4 : Σχηματικό διάγραμμα του μηχανισμού resistive switching μιας CBRAM	
Εικόνα 5 : Μοντέλο Hydrogen redox μιας μνήμης pt/a-C:H/TiN στις καταστάσεις LRS και HRS	
Εικόνα 6 : Μοντέλο resistance switching για διάταξη τριών στρώσεων	
Εικόνα 7 : Ιεραρχία memory / storage	
Εικόνα 8 : Τα δυο παράθυρα κύλισης σε ένα 2D CNN	
Εικόνα 9 : Πολλαπλασιασμός διανύσματος πίνακα σε μια RE-RAM	

Κεφάλαιο 1 Εισαγωγή

1.1 Η μνήμη RAM

Η μνήμη τυχαίας προσπέλασης (random access memory), αναφερόμενη και ως RAM, είναι μια μνήμη προσωρινής αποθήκευσης δεδομένων (short term data storage). Είναι εκεί που τα δεδομένα αποθηκεύονται ώστε να χρησιμοποιηθούν από τον επεξεργαστή για να «τρέξουν» οι εφαρμογές και τα αρχεία που θέλουμε.

Η RAM ενός computer, είναι κατασκευασμένη από μικροσκοπικούς πυκνωτές και τρανζίστορ, και εμφανίζεται στην μορφή συνήθως τετράγωνου επίπεδου κυκλώματος. Οι υπολογιστές συνήθως έχουν τουλάχιστον δύο RAM modules, με χώρο ώστε να προστεθεί και άλλο αν χρειαστεί. Τα RAM modules, δουλεύουν σε συνεργασία με το CPU (central processing unit) [5].

Η RAM διακρίνεται σε δύο κατηγορίες

- volatile (πτητική)
- non-volatile (μη πτητική)

Μια volatile μνήμη χάνει τα δεδομένα που είχαν αποθηκευτεί σε αυτή όταν κόψουμε την τροφοδοσία. Παραδείγματα: DRAM (Dynamic Random Access Memory), SRAM (Static Random Access Memory).

Χαρακτηριστική ιδιότητα μιας μη πτητικής μνήμης είναι ότι με διακοπή της τροφοδοσίας, τα δεδομένα δεν «σβήνουν» από αυτή. Παράδειγμα είναι η flash memory και η RE-RAM (Resistive Random Access Memory).

1.2 Η σημασία δημιουργίας νέων RAM

Η σημασία δημιουργίας νέων RAM έγκειται στο γεγονός ότι θα πρέπει να πληρούνται ορισμένα κριτήρια και όσο το δυνατόν περισσότερα σε συνδυασμό σε μια RAM όπως αυτά που αναφέρονται παρακάτω δηλαδή :

- χαμηλή τάση λειτουργίας (<1),
- αντοχή σε πολλούς κύκλους (>10),
- αυξημένη κράτηση δεδομένων (>10 χρόνια),

- χαμηλή κατανάλωση ενέργειας (fJ/bit),
- μεγάλη ικανότητα επεξεργασίας (<10nm).

Ως εκ τούτου έχουν αρχίσει να αναπτύσσονται καινούργιες τεχνολογίες οι οποίες βασίζονται στην αντίσταση πάρα από το φορτίο(πυκνωτή) για να αποθηκεύσουν την πληροφορία έτσι ώστε αυτές οι μνήμες να ανταποκριθούν στις απαιτήσεις μας.

Τέτοιες τεχνολογίες είναι:

- Phase change memory (PCM)
- Spin transfer torque magneto-resistive random-access memory (STT-MRAM)
- Resistive Random Access Memory (RE-RAM)

1.3 RE-RAM και γιατί την επιλέξαμε

Η RE-RAM αποτελείται από μια στιβάδα μονωτή (insulating layer) που γίνεται “σάντουιτς” μεταξύ δύο μεταλλικών ηλεκτροδίων (metal electrodes). Η λειτουργία της RE-RAM αποδίδεται στο γεγονός του σχηματισμού και της θραύσης των αγώγιμων οδών (filaments) των LRS (low resistance state) και HRS (high resistance state) αντιστάσεων και της μετάβασης από την μία κατάσταση στην άλλη.

Οι μνήμες RE-RAM είναι στο επίκεντρο ερευνών καθώς η αρχή λειτουργίας τους τους επιτρέπει τον συνδυασμό των παραπάνω ιδιοτήτων. Αναφορικά ιδιότητες που τις καθιστούν προνομιακές είναι οι εξής [8]:

1. Non volatile memory (NVM)
2. In memory computing
3. Neuromorphic Computing
4. Edge devices and IoT
5. Crossbar Arrays
6. Analog Computing

1.4 Οργάνωση του τόμου

Στο πρώτο κεφάλαιο εισάγεται η έννοια των RE-RAM. Αρχικά παρουσιάζεται συνοπτικά η έννοια των μνημών RAM και πιο συγκεκριμένα ένα είδος RAM η RE-RAM η οποία αποτελεί και το αντικείμενο της πτυχιακής. Ακόμα, αναφέρονται επιγραμματικά ορισμένα πλεονεκτήματα αυτού του τύπου μνήμης και ο λόγος που αποτελεί αντικείμενο έρευνας και οι εφαρμογές της.

Στο δεύτερο κεφάλαιο επεξηγείται το τι είναι μια resistance random access memory (RE-RAM), μια γρήγορη ιστορική αναδρομή, πως λειτουργεί το resistance switching, εσωτερική κατασκευή οι ιδιότητες και οι χρήσεις της.

Στο τρίτο κεφάλαιο θα δούμε διαφορά είδη RE-RAMS όπως η :

- 1) Anion-type RE-RAM
- 2) Cation-type RE-RAM
- 3) Carbon-based RE-RAM
- 4) Oxide-based electrode RE-RAM

Στο τέταρτο κεφάλαιο θα αναλύσουμε τι είναι το memory computing και τι μας προσφέρει ως μια ιδιότητα των RE-RAMS. Τέλος επεξηγείται πως υλοποιούν πράξεις πινάκων

Στο πέμπτο κεφάλαιο θα αναφερθούμε στα πλεονεκτήματα αλλά και τα μειονεκτήματα των RE-RAMS.

Κεφάλαιο 2 Τεχνολογική βάση των RE-RAM

2.1 Τι είναι η RE-RAM

Η (RE-RAM ή RRAM) είναι ένας τύπος μη πτητικής (non volatile) μνήμης υπολογιστή και η αρχή λειτουργίας είναι η μεταβολή της αντίστασης ενός διηλεκτρικού υλικού στερεάς κατάστασης, γνωστό και ως memristor. Ένα διηλεκτρικό υλικό δεν άγει το ηλεκτρικό ρεύμα. Όταν όμως εφαρμόσουμε πάνω από κάποια τάση τότε προκαλείται μόνιμη παραμόρφωση και αστοχία του υλικού. Η ιδιότητα που έχει όμως το memristor είναι ότι με την εφαρμογή της υψηλής αυτής τάσης η παραμόρφωση δεν είναι μόνιμη αλλά αντιστρεπτή διαδικασία. Με την εφαρμογή κατάλληλης τάσης σε ένα memristor, δημιουργείται ένα αγώγιμο μονοπάτι στο διηλεκτρικό. Η κατάσταση αυτή είναι γνωστή και ως κατάσταση χαμηλής αντίστασης (Low Resistance State LRS). Διακόπτοντας ή αναστρέφοντας την τάση το αγώγιμο μονοπάτι παύει να υπάρχει και επομένως η κατάσταση αυτή ονομάζεται κατάσταση υψηλής αντίστασης (High Resistance State HRS) [9]

Σε αντίθεση με την παραδοσιακή μνήμη RAM (Random Access Memory), η οποία αποθηκεύει δεδομένα προσωρινά, η RE-RAM διατηρεί πληροφορίες ακόμα και όταν η τροφοδοσία είναι απενεργοποιημένη.

2.2 Ιστορική αναδρομή

Η εφεύρεση της RE-RAM αποδίδεται στον συνεργάτη της Hewlett Packard R. Stanley Williams στα εργαστήρια της HP. Χρησιμοποίησε δύο στρώσεις φιλμ διοξειδίου του τιτανίου για να αποθηκεύσει την αντίσταση [7].

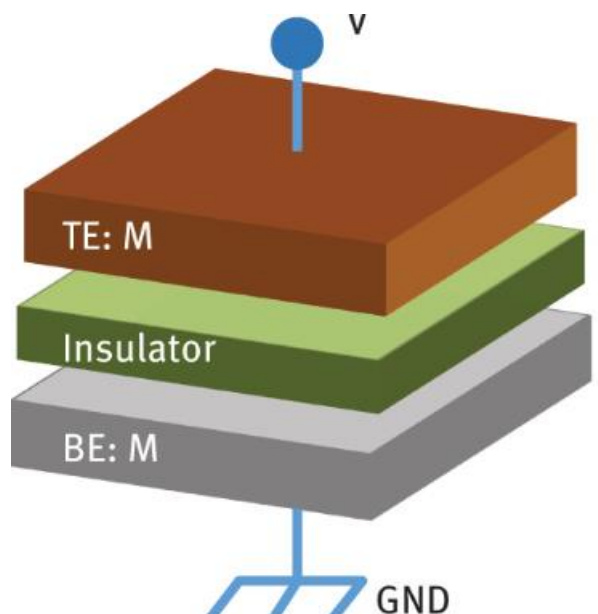
Ωστόσο, η έννοια του memristor προτάθηκε αρχικά σε μια εργασία από τον καθηγητή Leon Chua το 1971 στο Πανεπιστήμιο της Καλιφόρνια. Στην εργασία του το 1971 με τίτλο «Memristor—The Missing Circuit Element», ο Chua πρότεινε την ύπαρξη ενός τέταρτου θεμελιώδους στοιχείου κυκλώματος παράλληλα με την αντίσταση (resistor), τον πυκνωτή (capacitor) και τον επαγωγέα (inductor). Περιέγραψε το memristor ως μια συσκευή με μνήμη της προηγούμενης κατάστασης αντίστασης.

Στις αρχές της δεκαετίας του 2000, αρκετές εταιρείες ανέπτυξαν ενεργά την τεχνολογία RE-RAM. Ορισμένες από αυτές τις εταιρείες υπέβαλαν αιτήσεις για διπλώματα ευρεσιτεχνίας για διάφορες πατέντες.

Η RE-RAM έχει εισέλθει στην εμπορευματοποίηση, αν και αρχικά σε περιορισμένη κλίμακα με χωρητικότητα μόλις μερικά KB. Τα τελευταία χρόνια, οι τεχνολογίες RE-RAM έχουν ερευνηθεί εντατικά από μεγάλες εταιρείες όπως η Panasonic και η Samsung, καθώς και πολλές νέες νεοφυείς επιχειρήσεις. Το 2013, η Panasonic παρουσίασε τα MCU 8-bit MN-101L που χρησιμοποιούν μνήμη RERAM 64 Kb, σηματοδοτώντας τα πρώτα προϊόντα RERAM που διατίθενται στην αγορά.

2.3 Εεσωτερική κατασκευή και λειτουργία μια RE-RAM

Μια RE-RAM αποτελείται από μια κυψέλη resistive switching memory η οποία έχει δομή μέταλλο-μονωτής-μέταλλο γνωστό και ως MIM (Metal Insulator Metal). Η δομή μιας μνήμης RE-RAM αποτελείται από ένα μονωτικό στρώμα (διηλεκτρικό υλικό) (I) το οποίο γίνεται «σάντουιτς» μεταξύ των δύο μεταλλικών (M) ηλεκτροδίων. Το σχήμα και η διατομή μιας χαρακτηριστικής δομής RE-RAM φαίνεται στην παρακάτω εικόνα (Hwang et al 2016) [2]



Εικόνα 1: Χαρακτηριστική δομή RERAM, (Amit Prakash, Hyunsang Hwang)

Εφαρμόζοντας εξωτερική τάση με την μορφή παλμού σε μια μονάδα της RE-RAM επιτρέπεται η μετάβαση της συσκευής από μια υψηλή κατάσταση αντίστασης (HRS) ή κατάσταση OFF που αναφέρεται γενικά ως λογική τιμή «0» σε κατάσταση χαμηλής αντίστασης (LRS) ή ON κατάσταση που αναφέρεται γενικά ως λογική τιμή «1» και αντίστροφα.

Με την εφαρμογή παλμού υψηλής τάσης, δημιουργούνται μονοπάτια αγωγίμου στην στρώση του διηλεκτρικού υλικού. Η τάση αυτή πολλές φορές συμβολίζεται με V_f (forming voltage) και εξαρτάται από την επιφάνεια του κελιού και το πάχος του οξειδίου. Για την μετάβαση από την κατάσταση LRS στο HRS, εφαρμόζεται τάση παλμού που αναφέρεται ως τάση RESET (V_{reset}). Αντίθετα, η τάση στην οποία πραγματοποιείται η μετάβαση από το HRS στο LRS αναφέρεται ως τάση SET (V_{set}).

Για την αποτελεσματική ανάγνωση δεδομένων από την RE-RAM, χρειάζεται να εφαρμόσουμε μια μικρή τάση <<ανάγνωσης>> η οποία δεν διαταράσσει την τρέχουσα κατάσταση του κελιού η οποία είναι ή λογικό 0 (HRS) ή λογικό 1 (LRS). Αφού λοιπόν διατηρείται η κατάσταση της RE-RAM ακόμα και όταν αφαιρέσουμε την μικρή αυτή τάση μπορούμε να χαρακτηρίσουμε την RE-RAM ως μια περίπτωση μη πτητικής μνήμης.

2.4 Switching mechanism – Μηχανισμός εναλλαγής

Έχει παρατηρηθεί ότι σε ορισμένους μονωτές, η αντίσταση στο ηλεκτρικό ρεύμα μεταβάλλεται κατά την εφαρμογή εφαρμοζόμενου ηλεκτρικού πεδίου. Αυτή η ιδιότητα ονομάζεται resistance switching και πραγματοποιείται με μεταβολή της αντίστασης του διηλεκτρικού, έχει πρόσφατα διερευνηθεί για την ανάπτυξη μη πτητικών (non volatile) μνημών.

Ο μηχανισμός εναλλαγής (switching mechanism), βασίζεται στην δημιουργία αγωγίμου νήματος (conductive filament- CF) μέσα σε ένα διηλεκτρικό υλικό. Το conductive filament είναι ένα κανάλι με πολύ μικρότερη διάμετρο της τάξης των

νανομέτρων που συνδέει το επάνω και το κάτω ηλεκτρόδιο της κυψέλης μνήμης. Μια κατάσταση χαμηλής αντίστασης (LRS) με υψηλή αγωγιμότητα επιτυγχάνεται όταν το CF είναι συνδεδεμένο και η υψηλή αντίσταση (HRS) προκύπτει όταν το CF αποσυνδέεται με ύπαρξη ενός κενού μεταξύ των ηλεκτροδίων [5]

Με βάση τη σύνθεση του CF, μια RE-RAM μπορεί να ταξινομηθεί στους ακόλουθους δύο τύπους:

- RE-RAM με βάση τα μεταλλικά ιόντα που αναφέρεται επίσης και ως μνήμη τυχαίας πρόσβασης αγωγίμης γέφυρας (CBRAM) και
- RE-RAM με βάση το CF όταν αυτό δημιουργείται από «κενά οξυγόνου» που αναφέρεται ως «OxRRAM».

Πρέπει να σημειωθεί εδώ ότι το CBRAM μερικές φορές αναφέρεται ως μνήμη ηλεκτροχημικής επιμετάλλωσης (ECM), ενώ το «OxRRAM» μερικές φορές είναι επίσης γνωστό ως μνήμη αλλαγής σθένους (VCM).

Το φαινόμενο resistance switching ή αλλαγή αντίστασης βρέθηκε μέσω ερευνών ότι συμβαίνει σε διάφορα οξειδία. Τα δυαδικά οξείδια μετάλλων έχουν μελετηθεί εκτενώς ως προτιμώμενο υλικό μονωτή για μελλοντικές εφαρμογές non volatile μνήμης, κυρίως λόγω της συμβατότητάς τους με την τεχνολογία CMOS BEOL.

2.4.1 CMOS BEOL

Με τον όρο CMOS (complementary metal-oxide semiconductor) αναφερόμαστε στην τεχνολογία ημιαγωγών που χρησιμοποιείται σε πληθώρα από τα σημερινά ολοκληρωμένα κυκλώματα (IC- Integrated Circuits), γνωστά και ως τσιπ ή μικροτσιπ. Μια κατηγορία ηλεκτρονικών εξαρτημάτων, τα τρανζίστορ τεχνολογίας CMOS βασίζονται στην τεχνολογία MOSFET (metal-oxide semiconductor field-effect transistor). Τα MOSFET χρησιμεύουν ως διακόπτες ή ενισχυτές που ελέγχουν την ποσότητα ηλεκτρικής ενέργειας που ρέει μεταξύ των ακροδεκτών πηγής και τερματικών, με βάση το μέγεθος της εφαρμοζόμενης τάσης.

Με τον όρο BEOL (back end of line) εννοούμε την τελική κατάσταση επεξεργασίας των ημιαγωγών και αφορά τις διασυνδέσεις που βρίσκονται στο επάνω μέρος ενός τσιπ. Στα σημερινά προηγμένα IC, αυτό αποτελείται από έως και δεκαπέντε στρώματα σύνθετης καλωδίωσης.

2.5 Ιδιότητες και χρήσεις των RE-RAMS

Οι RE-RAMS ως μια νέα και συνεχώς εξελισσόμενη τεχνολογία έχει κάποιες ιδιότητες η χρήση των οποίων μας προσφέρουν διευκολύνσεις και αναβαθμίσεις στου υπολογιστές μας

1. Non volatile memory (NVM)

Προσφέρει μεγαλύτερη ταχύτητα έγγραφής δεδομένων, χαμηλότερη ισχύ και μεγάλη ικανότητα επεξεργασίας.

Μπορεί να χρησιμοποιηθεί στους SSD και σε άλλες συσκευές αποθήκευσης.

2. In memory computing

Οι συσκευές που έχουν κύκλωμα RE-RAM, μπορούν να ικανοποιήσουν ταυτόχρονα αποθήκευση δεδομένων, και υπολογισμούς (in memory computing). Αυτό επιταχύνει συγκεκριμένες λειτουργίες, όπως μηχανική μάθηση και αναγνώριση μοτίβων, καθώς μειώνονται οι κινήσεις δεδομένων μεταξύ CPU και μνήμης.

3. Neuromorphic Computing

Οι RE-RAM, μπορούν να αποθηκεύσουν τους συντελεστές βαρύτητας των νευρωνικών δικτύων και να επιταχύνουν τον πολλαπλασιασμό πινάκων. Αυτή η ιδιότητα χρησιμοποιείται για να χτίσει τεχνητές συνάψεις και νευρώνες, καθιστώντας εφικτές εφαρμογές των νευρωνικών δικτύων με χαμηλή ενέργεια.

4. Edge devices and IoT

Η χαμηλή κατανάλωση ενέργειας τις καθιστά επιλογή σε εφαρμογές IoT.

Μπορούν να ενσωματωθούν σε αισθητήρια (sensors) και σε χαμηλής ισχύος συσκευές.

5. Crossbar Arrays

Κάθε κυψέλη των RE-RAM, αναπαριστά ένα στοιχείο μνήμης. Έτσι, μπορούμε να αποθηκεύσουμε αρκετή πληροφορία σε μικρό χώρο και να διαβάσουμε και να γράψουμε αρκετά δεδομένα.

6. Analog Computing

Οι RE-RAM λόγω του φαινομένου resistive switching χρησιμοποιούνται για αναλογικούς υπολογισμούς όπως η επεξεργασία σήματος, η αναγνώριση εικόνας και η βελτιστοποίηση.

Κεφάλαιο 3 Κατηγορίες και τύποι RE-RAM

3.1 Είδη RE-RAMS

Η λειτουργία μιας RE-RAM βασίζεται στον αντιστρεπτό σχηματισμό και διάσπαση μιας αγώγιμης διαδρομής σε μια στοιβάδα με αντίσταση, και δημιουργεί μια κατάσταση χαμηλού ρεύματος και μια κατάσταση υψηλού ρεύματος (HIGH, LOW).

Υπάρχουν αρκετές ταξινομήσεις. Στην παρούσα διπλωματική θα διακρίνουμε τις RE-RAMS ως προς τον μηχανισμό του resistance switching. Έτσι έχουμε τις:

- Anion-Type RE-RAM: ο μηχανισμός αυτού του τύπου RE-RAM, γίνεται από ιόντα οξυγόνου (αρνητικό φορτίο)
- Cation-type RE-RAM: Ο μηχανισμός αυτών των RE-RAM, γίνεται από την οξείδωση και μετακίνηση μεταλλικών ιόντων (θετικό φορτίο). Αυτός ο τύπος RE-RAM είναι γνωστός και ως «αγώγιμη γέφυρα» (conductive bridge ram CBRAM)
- Carbon based RE-RAM: αυτά τα υλικά χρησιμοποιούν τον άνθρακα ως βάση στην στοιβάδα αλλαγής κατάστασης (switching layer)
- Oxide-based electrode RE-RAM: χρησιμοποιούνται υλικά ως ηλεκτρόδια σε μια RE-RAM χαρακτηριστικά των οποίων θα αναλυθούν και παρακάτω

Εν κατακλείδι, θα μελετηθεί πως τα υλικά όπως η στοιβάδα αλλαγής κατάστασης και τα ηλεκτρόδια (electrodes) αλλά και οι τεχνικές κατασκευής επηρεάζουν της ιδιότητες μιας RE-RAM. Ακόμη θα μελετηθεί το φαινόμενο resistance switching για κάθε μια από τις παραπάνω κατηγορίες

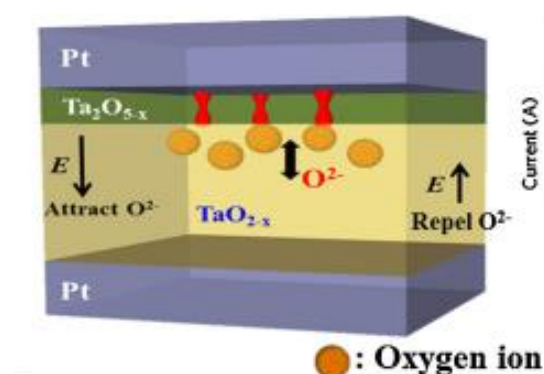
3.2 Anion-type RE-RAM

Σε αυτόν τον τύπο RE-RAM το resistance switching επιτυγχάνεται με το σχηματισμό κενών θέσεων οξυγόνου και τη μεταφορά ιόντων οξυγόνου. Οι αγώγιμες διαδρομές σχηματίζονται από συνεχείς κενές θέσεις οξυγόνου που οδηγούν την διάταξη στη μετάβαση σε LRS (set). Ωστόσο, στη διαδικασία επαναφοράς (reset), οι κενές θέσεις οξυγόνου ανασυνδυάζονται με ιόντα οξυγόνου, οδηγώντας τη διάταξη να επιστρέψει

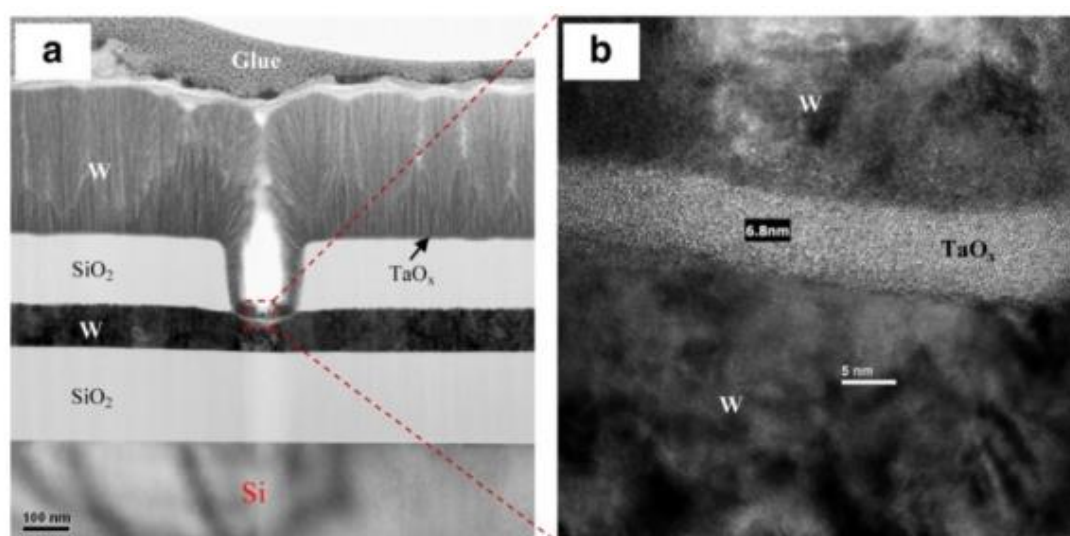
στο HRS . Επιπλέον, η anion type RE-RAM χρειάζεται ένα ενεργό ηλεκτρόδιο ιόντων οξυγόνου ή ένα φιλμ φτωχό σε οξυγόνο ως δεξαμενή ιόντων για το resistance switching.

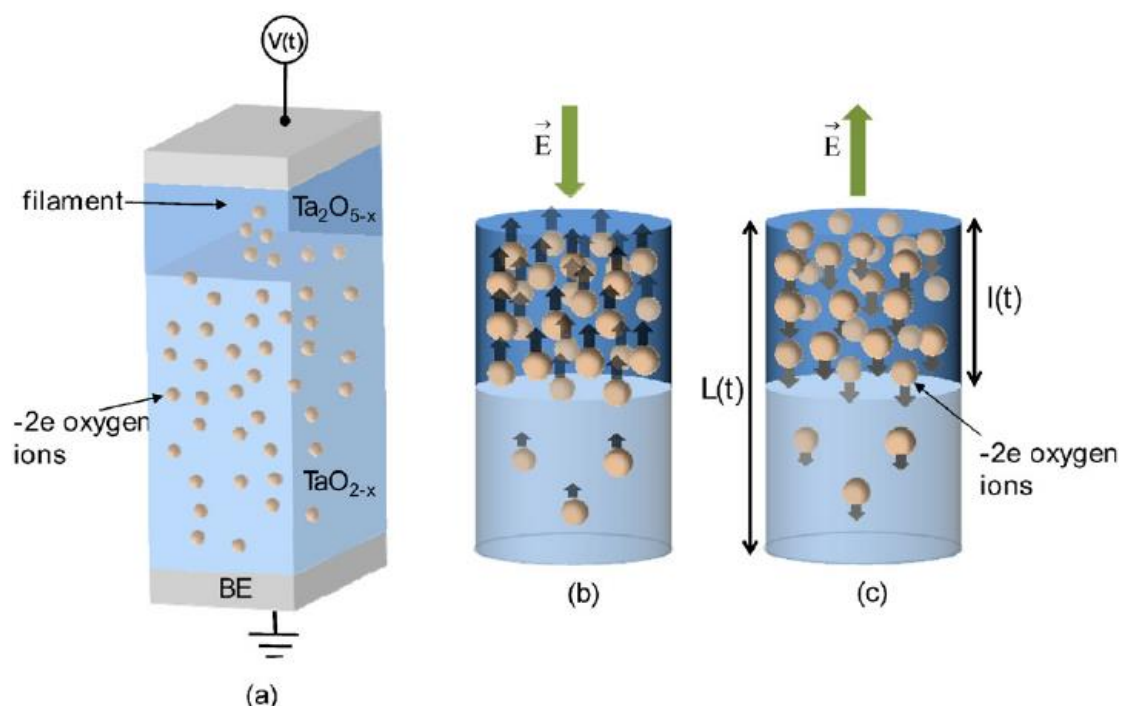
Οι μηχανισμοί που βασίζονται στις κινήσεις των ανιόντων αλλά και στο διπολικό (bipolar) switching, παρατηρούνται κυρίως στα οξειδωτικά υλικά κυρίως οξείδια μετάλλων (metal oxide, πχ WO_x , HfO_x , TaO_x , AlO_x) και έχουν μελετηθεί αρκετά λόγω της συμβατότητάς τους με την τεχνολογία CMOS.

Το οξείδιο πυριτίου, είναι ένα υποσχόμενο υλικό για εφαρμογές RE-RAM λόγω της μεθόδου κατασκευής του, της ικανοποιητικής φυσικής και χημικής σταθερότητας και την συμβατότητα με τα ενσωματωμένα κυκλώματα (IC).



Εικόνα 2: Σχηματική αναπαράσταση τομής TaO_x διάταξης που αποτελείται από μια λεπτή $\text{Ta}_2\text{O}_{5-x}$ μονωτική στρώση και μια στρώση βάσης TaO_{2-x} [3]



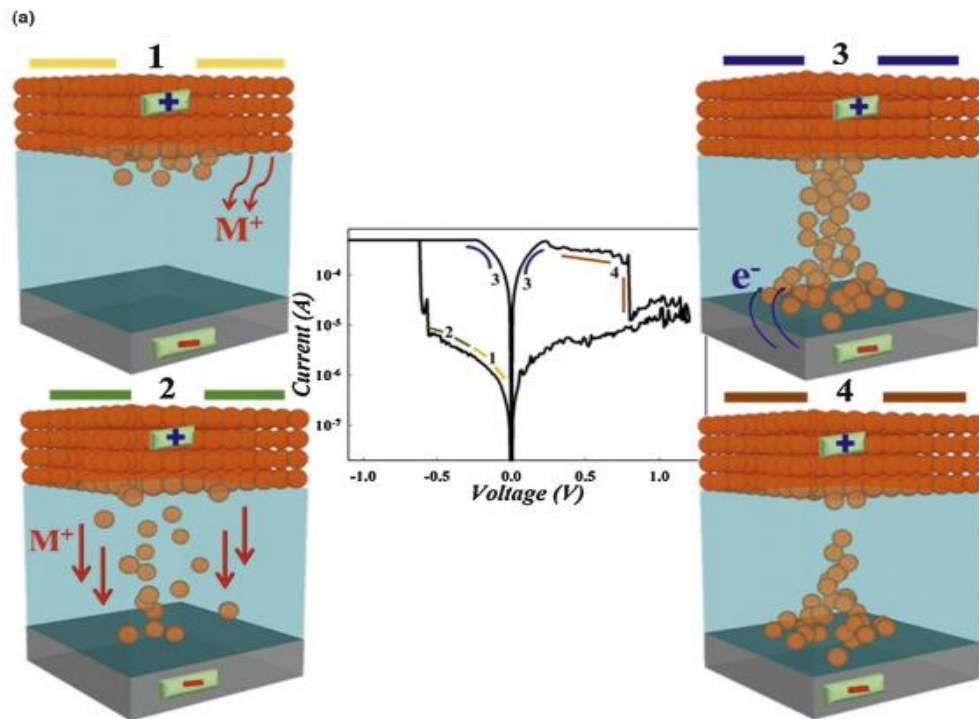


Εικόνα 3: Anion type reram, (Ji-Hyun Hur et al)

3.3 Cation-type RE-RAM

Ο μηχανισμός του resistance switching που γίνεται σύμφωνα με την redox reaction και μετακίνηση των μεταλλικών ιόντων, οδηγεί στον σχηματισμό και την διάλυση των μεταλλικών νημάτων (filaments) σε αυτόν τον τύπο RE-RAM. Επιπλέον, ο μηχανισμός του resistance switching μιας CBRAM οφείλεται σε μια ηλεκτροχημική αντίδραση με άργυρο ή χαλκό (Ag ή Cu) για τον σχηματισμό ή την διάλυση των μεταλλικών «νημάτων» (filaments).

Ας υποθέσουμε ότι έχουμε χαλκό (Cu). Τότε η διαδικασία του switching είναι η εξής. Αρχικά, ο χαλκός οξειδώνεται και ιόντα Cu^+ εγχέονται στις στιβάδες διηλεκτρικού υλικού από το ηλεκτρόδιο Cu. Έπειτα, τα ιόντα Cu (Cu^+) έλκονται από το πιο κάτω ηλεκτρόδιο. Η αντίδραση αναγωγής από HRS σε LRS πραγματοποιείται όταν τα ιόντα Cu^+ έρχονται σε επαφή με το ηλεκτρόδιο της βάσης μέσω του αγωγίμου δρόμου στο διηλεκτρικό υλικό. [3]

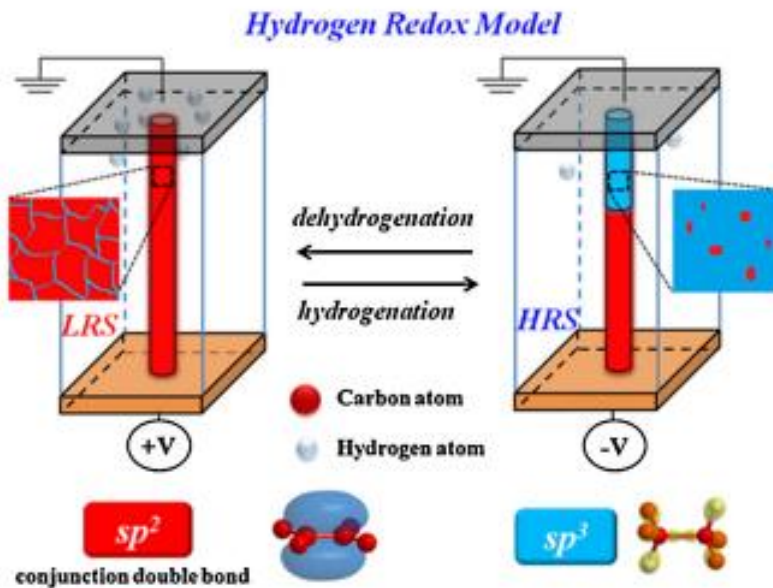


Εικόνα 4: Το σχηματικό διάγραμμα του μηχανισμού resistive switching μιας CBRAM.

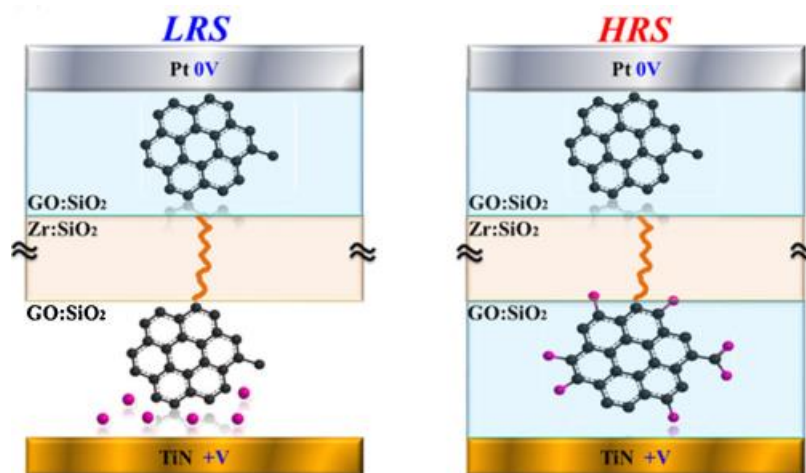
3.4 Carbon-based RE-RAM

Μέχρι πρότινος, η (C-RERAM) εμφανίζεται ως μια εναλλακτική λόγω των πλεονεκτημάτων της απλής χημικής σύνθεσης, του χαμηλού κόστους, της υψηλής σταθερότητας και της χαμηλής κατανάλωσης ισχύος.

Ο μηχανισμός του resistance switch λειτουργεί ως εξής. Η α -C:H στοιβάδα, ενώθηκε με άνθρακα και σχημάτισε έναν συζυγή διπλό δεσμό ο οποίος είναι ένας αγωγίμος δρόμος. Από αυτόν τον αγωγίμο δρόμο. Η κατάσταση LRS επιτεύχθηκε με την διαδικασία της αφυδρογόνωσης (dehydrogenation) όπως φαίνεται και στην παρακάτω εικόνα. Η HRS επιτεύχθηκε με αντίστροφο ηλεκτρικό πεδίο και την διαδικασία της υδρογονοποίησης (hydrogenation) [1]



Εικόνα 5: Μοντέλο Hydrogen redox μιας μνήμης $pt/a-C:H/TiN$ στις καταστάσεις LRS και HRS.



Εικόνα 6: Μοντέλο Resistance switching για διάταξη τριών στρώσεων.

3.5 Oxide-based electrode RE-RAM

Τα ηλεκτρονικά μόντουλ (modules) αναμένεται να ανανεωθούν προκειμένου να κατέχουν πολλαπλές λειτουργίες όπως ευκολία στην μεταφορά, ευελιξία, και φορητότητα. Το Indium tin oxide (ITO) είναι το πιο ευρέως διαδεδομένο λόγω των πλεονεκτημάτων της υψηλής αγωγιμότητας, της υψηλής εκπομπής σε ορατό φως, της χημικής ευστάθειας και την ιδιότητα της αντίστασης σε οξείδωση.

Ο μηχανισμός της δημιουργίας αγώγιμης οδού, του ITO κυριαρχείται από υψηλά ποσοστά οπών οξυγόνου. Χάρη στην παρουσία μεγάλου αριθμού οπών οξυγόνου, το ITO είναι μια άριστη δεξαμενή αποθήκευσης οξυγόνου. Σε αντίθεση με τα συνηθισμένα ενεργά ηλεκτρόδια οξυγόνου, που μπορούν να αποθηκεύσουν ιόντα οξυγόνου μόνο μέσω οξείδωσης στην επιφάνεια.

Κατά τον σχηματισμό αγώγιμης οδού, η κινούσα δύναμη του ιόντος του οξυγόνου κατευθύνεται μέσα στο ITO ηλεκτρόδιο επειδή υπάρχουν πολλές οπές οξυγόνου μέσα σε αυτό. Ενδυναμωμένο από το gradient της συγκέντρωσης μαζί με την ηλεκτρική κινούσα δύναμη, τα ιόντα οξυγόνου μπορούν πιο εύκολα να προωθηθούν στο ηλεκτρόδιο ITO και ως εκ τούτου να πάρουμε μια μικρή τάση. Στην κατάσταση LRS, ο ημιαγωγός ITO γύρω από την κορυφή της αγώγιμης οδού, και επαγόμενο από την υψηλή συγκέντρωση οξυγόνου, οδηγεί στον σχηματισμό ημιαγωγού παρόμοιο με το ITO. Όσον αφορά την διαδικασία του reset το αγώγιμο μονοπάτι, θα οξειδωθεί και θα διασπαστεί, οδηγώντας στην κατάσταση HRS.

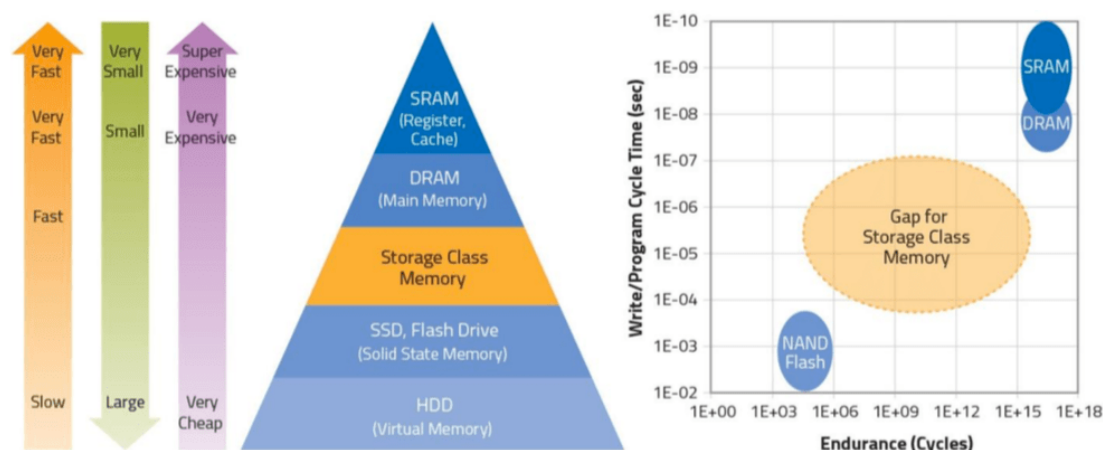
Κεφάλαιο 4 Εφαρμογές για In-Memory Computing

4.1 Τι είναι το In-Memory Computing

Στα σημερινά συστήματα, η παραδοσιακή ιεραρχία μνήμης/αποθήκευσης (memory/storage) είναι ευθύς (straightforward). Για αυτό, η SRAM είναι ενσωματωμένη στον επεξεργαστή του cache, ο οποίος μπορεί να έχει γρήγορη πρόσβαση σε προγράμματα που χρησιμοποιούνται συχνά [4].

Η μνήμη DRAM, η οποία χρησιμοποιείται ως κύρια μνήμη, είναι ξεχωριστή και βρίσκεται σε μια ηλεκτρονική διάταξη (dual in-line memory module - DIMM).

Και οι μονάδες δίσκου (disk drives) και οι μονάδες αποθήκευσης SSD με λογική NAND χρησιμοποιούνται για αποθήκευση.



Εικόνα 7: Ιεραρχία memory / storage. Πηγή Lam (ανακτήθηκε από <https://semiengineering.com/entities/lam-research/>)

Με βάση αυτή την ιεραρχία, τα συστήματα αντιμετωπίζουν πάρα πολλά δεδομένα που προέρχονται από το δίκτυο.

Εκτός από περισσότερα δεδομένα, υπάρχει η ανάγκη και για ταχύτερη μνήμη πχ στις εφαρμογές αναγνώρισης προσώπου και ελέγχου ταυτότητας που χρησιμοποιούν την τεχνητή νοημοσύνη.

Καθώς ο όγκος των δεδομένων αυξάνεται από δεκάδες σε εκατοντάδες terabyte μέσα σε έναν διακομιστή (server), αντιμετωπίζουμε πρόβλημα μεταφοράς δεδομένων από τον SSD στην CPU και αντίστροφα κάτι που αναμένεται να δημιουργήσει συμφόρηση

(bottleneck). Γι' αυτό δημιουργήθηκε η ανάγκη για κάποια ενδιάμεσα στάδια στην ιεραρχία της πυραμίδας για την αντιμετώπιση του προβλήματος. Ένας τρόπος για να λυθεί αυτό είναι η ενσωμάτωση επεξεργαστών (processors), μνήμης (memory) και άλλων συσκευών σε μια παραδοσιακή αρχιτεκτονική von Neumann. Η λύση αυτή, προσθέτει κόστος και πολυπλοκότητα. Εξαιτίας αυτών των προβλημάτων στραφήκαμε σε νέες προσεγγίσεις που είναι να προχωρήσουμε προς καινούργιες αρχιτεκτονικές μνήμης.

Η τάση για ανάπτυξη του in-memory computing και near-memory computing θα οδηγήσει σε καινοτόμες αρχιτεκτονικές και μνήμες. Παρακάτω αναλύεται τι ακριβώς είναι το in-memory computing.

Ο in-memory σχεδιασμός γίνεται χρησιμοποιώντας ήδη υπάρχουσα τεχνολογία κατασκευής μνημών. Εφαρμογές που χρησιμοποιείται η in-memory computing τεχνολογία:

- Βάσεις δεδομένων ως cache αλλά και άλλες εφαρμογές
- Chip τα οποία συμμετέχουν στην επεξεργασία δεδομένων στα νευρωνικά δίκτυα.

Για χρόνια, η μέθοδος in-memory computing, χρησιμοποιείται στις βάσεις δεδομένων. Μια βάση δεδομένων αποθηκεύεται σε έναν υπολογιστή. Σε μια παραδοσιακή βάση δεδομένων, τα δεδομένα αποθηκεύονται σε μια μονάδα δίσκου, αλλά η πρόσβαση στα δεδομένα από τη μονάδα δίσκου μπορεί να είναι μια αργή διαδικασία. Έτσι, οι προμηθευτές βάσεων δεδομένων έχουν αναπτύξει τρόπους επεξεργασίας των δεδομένων στην κύρια μνήμη (main memory) ενός server αντί της μονάδας δίσκου. Αυτό, με τη σειρά του, αυξάνει τις ταχύτητες των δεδομένων. Η προσέγγιση αυτή χρησιμοποιεί τον κλασικό τρόπο δηλαδή το μοντέλο αρχιτεκτονικής von Neumann. Αυτό που γίνεται με αυτόν τον τρόπο είναι η καλύτερη εναπόθεση των δεδομένων στην πυραμίδα, ώστε η ανταλλαγή δεδομένων να γίνει όσο πιο γρήγορη γίνεται.

Στον κόσμο των ημιαγωγών/συστημάτων, ο υπολογισμός στη μνήμη έχει την ίδια βασική αρχή με διαφορετική έννοια, ότι δηλαδή η μνήμη έρχεται ακόμα πιο κοντά ή

μέσα στις διαδικασίες επεξεργασίας δεδομένων. Στο παρελθόν, αυτή η τεχνολογία αναφερόταν μερικές φορές ως «processing in memory». Για χρόνια, οι πωλητές εισήγαγαν διάφορες συσκευές στο πεδίο, αλλά πολλές από αυτές τις προσπάθειες απέτυχαν. Πρόσφατα, αρκετές εταιρείες παρουσίασαν νέες και βελτιωμένες εκδόσεις αυτής της τεχνολογίας. Υπάρχουν διάφορες προσεγγίσεις που χρησιμοποιούν DRAM, flash και νέους τύπους μνήμης. Πολλές από αυτές έχουν τεχνολογία in-memory computing αλλά όπως προαναφέρθηκε δεν πρέπει να συγχέονται με την in – memory τεχνολογία στις βάσεις δεδομένων.

Πολλές από τις νέες in-memory αρχιτεκτονικές τσιπ έχουν σχεδιαστεί για τα νευρωνικά δίκτυα. Τα νευρωνικά δίκτυα αποτελούνται από πολλαπλούς νευρώνες και συνάψεις δηλαδή τις συνδέσεις των νευρώνων. Ένας νευρώνας θα μπορούσε να αποτελείται από ένα κύτταρο μνήμης (memory cell) με λογικές πύλες (logic gates).

4.2 Πράξεις πινάκων με in-memory computing

Τα νευρωνικά δίκτυα λειτουργούν υπολογίζοντας γινόμενα και αθροίσματα πινάκων (matrix products and sums). Αποτελείται από τρεις στιβάδες είσοδος (input), «κρυμμένο» (hidden) και έξοδος (output) [4].

Σε μια πράξη (operation), τα δεδομένα ή μοτίβα (patterns) αρχικά συμπληρώνονται στην στιβάδα εισόδου (input layer) και μεταδίδονται μέσω των hidden layer.

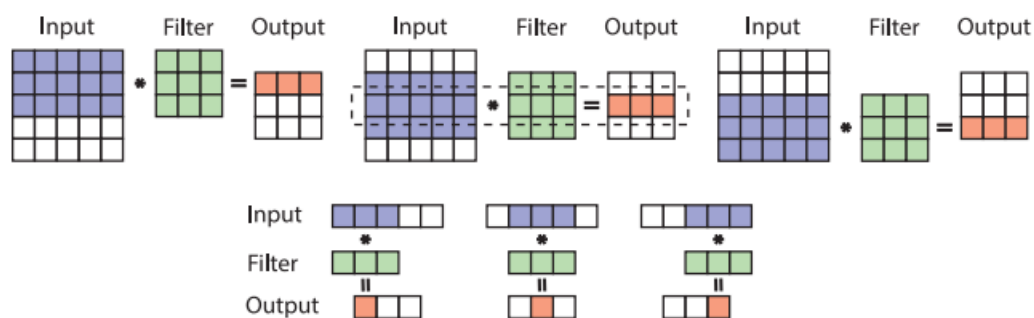
Μια ενδιαφέρουσα εφαρμογή που μπορεί να φανεί η χρήση της τεχνολογίας in-memory είναι στα συνελκτικά νευρωνικά δίκτυα (convolution neural networks – CNN) τα οποία είναι σχεδιασμένα για την αναγνώριση και ταξινόμηση εικόνων χρησιμοποιώντας τις αρχές της γραμμικής άλγεβρας [6].

4.2.1 Επαναχρησιμοποίηση μοτίβου δεδομένων στις στρώσεις συνέλιξης (convolution-layers)

Ο υπολογισμός της δισδιάστατης συνέλιξης φαίνεται στην παρακάτω εικόνα. Ένας πυρήνας συνέλιξης που ονομάζεται και φίλτρο, (convolution kernel, filter) ξεκινάει από την πάνω αριστερά γωνία των εισόδων (input), πηγαίνει προς το δεξιό τέρμα και

έπειτα μεταπηδάει στην επόμενη σειρά. Αυτή η διαδικασία, μπορεί να αναπαρασταθεί σε ιεραρχία δύο επιπέδων (two-level hierarchy)

- Το πρώτο επίπεδο, είναι ένα «παράθυρο» με πολλές σειρές, το οποίο κυλάει από πάνω προς τα κάτω ώστε να διασφαλίσει σε κάθε σειρά του πίνακα επαναχρησιμοποίηση δεδομένων
- το δεύτερο επίπεδο, είναι ένα «παράθυρο» πολλών στοιχείων που κυλάει από τα αριστερά προς τα δεξιά, για να διασφαλίσει επαναχρησιμοποίηση δεδομένων σε όλη την σειρά.



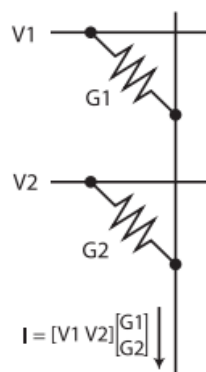
Εικόνα 8: Τα δύο παράθυρα κύλισης σε ένα 2D CNN.

4.2.2 RE-RAM based DNN Accelerators

Η παρακάτω εικόνα, δείχνει το βασικό λειτουργικό μέρος μιας RE-RAM σε εφαρμογή βαθιών νευρωνικών δικτύων (DNN). Τα στοιχεία, ενός τελεστή πίνακα, παριστάνονται ως αγωγιμότητες σε κάθε κελί RE-RAM στα σημεία διασταύρωσης των κελιών και ο τελεστής του διανύσματος είναι η είσοδος της τάσης στο κελί μνήμης. Σε αυτή την μορφή, ο πολλαπλασιασμός πινάκων μπορεί να γίνει μετρώντας το ρεύμα στις γραμμές ρεύματος (bitlines). Εφόσον η RE-RAM, υποστηρίζει πολλαπλασιασμό διανύσματος πίνακα, μια αφελής μέθοδο για να γίνει η συνέλιξη (convolution), σε μια RE-RAM, είναι η μετατροπή της συνέλιξης σε πολλαπλασιασμό πινάκων (ISAAC, PipeLayer).

Η ακρίβεια των δεδομένων σε κάθε κελί, περιορίζεται από τον αριθμό επιπέδων αντίστασης (resistance levels) των RE-RAM. Μια υψηλή ακρίβεια, μπορεί να

επιτευχθεί με τον συνδυασμό πολλών κελιών RE-RAM, τα οποία έχουν χαμηλή ακρίβεια δεδομένων σε κάθε κελί.



Εικόνα 9: Πολλαπλασιασμός διανύσματος πίνακα σε μια RRAM.

Κάθε νευρώνας αντιδρά στα δεδομένα. Χρησιμοποιώντας ένα σταθμισμένο σύστημα (weighted system) συνδέσεων, ένας νευρώνας στο δίκτυο αντιδρά πιο δυνατά όταν αισθάνεται ένα αντίστοιχο μοτίβο. Το αποτέλεσμα της πράξης των πινάκων, αποκαλύπτεται στο επίπεδο εξόδου (output layer).

Κεφάλαιο 5 Αξιολόγηση RE-RAM

5.1 Πλεονεκτήματα των RE-RAM

Στο παρόν κεφάλαιο αναλύονται τα σημαντικά πλεονεκτήματα των μνημών τυχαίας πρόσβασης με βάση την αντίσταση (RE-RAM). Οι μνήμες αυτές προσφέρουν ξεχωριστές ιδιότητες που επιτρέπουν την εφαρμογή τους σε πλήθος τεχνολογικών περιβαλλόντων και συσκευών. Μέσα από αυτή την ανάλυση, θα εξετάσουμε πώς η ενσωμάτωση των RE-RAM στα σύγχρονα συστήματα υπολογιστών βελτιώνει την απόδοση, μειώνει την κατανάλωση ενέργειας και προσφέρει ευελιξία στην αρχιτεκτονική μνήμης. Το κεφάλαιο αυτό θα παρουσιάσει επίσης τις διάφορες εφαρμογές των RE-RAM, τονίζοντας τη σημασία τους σε προηγμένες τεχνολογίες

όπως το cloud computing, η μηχανική μάθηση και η επεξεργασία big data. Ορισμένα από τα πλεονεκτήματα των RE-RAMS είναι τα εξής:

1. Ταχύτητα και απόδοση: Οι RE-RAM κάνουν γρηγορότερη ανάγνωση δεδομένων (read) και εγγραφή δεδομένων (write) συγκριτικά με τις κοινές NAND Flash memory. Με αυτό τον τρόπο βελτιώνεται η συνολική απόδοση.
2. Ενεργειακή απόδοση: Οι RE-RAM απαιτούν χαμηλότερες τάσεις λειτουργίας και άρα καταναλώνουν λιγότερη ηλεκτρική ισχύ και συνεπώς ενέργεια. Είναι πολύ σημαντικό ιδίως σε κέντρα επεξεργασίας δεδομένων και σε κινητές συσκευές όπου η κατανάλωση ηλεκτρικής ενέργειας είναι μείζονος σημασίας.
3. Ανθεκτικότητα: Με λιγότερους κύκλους εγγραφών (write cycles), σε σχέση με τις NAND Flash, οι RE-RAM έχουν μεγαλύτερη διάρκεια ζωής καθιστώντας τις καταλληλότερες ως προς την ανθεκτικότητα για την αποθήκευση δεδομένων.
4. Συμβατότητα και διαστάσεις: Οι RE-RAM είναι συμβατές με τα ενσωματωμένα κυκλώματα και μικρές σχετικά σε διαστάσεις καθιστώντας τις κατάλληλες για χρήση σε διάφορες εφαρμογές και πλακέτες.

5.2 Μειονεκτήματα των RE-RAM

Στην τρέχουσα ενότητα θα διερευνηθούν τα μειονεκτήματα των μνημών τυχαίας πρόσβασης με βάση την αντίσταση (RE-RAM). Παρά τα σημαντικά οφέλη που προσφέρουν, οι RE-RAM εμφανίζουν επίσης περιορισμούς που μπορούν να επηρεάσουν την ευρεία υιοθέτησή τους σε βιομηχανικές και εμπορικές εφαρμογές. Αυτά περιλαμβάνουν τεχνικές προκλήσεις όπως η αστάθεια κατά τη λειτουργία, οι περιορισμοί στην κλιμάκωση και οι δυσκολίες στη διαχείριση της διάρκειας ζωής των κυψελών. Το κεφάλαιο αυτό θα αναφέρει επίσης τους προβληματισμούς που υπάρχουν σχετικά με την αξιοπιστία και την οικονομική βιωσιμότητα των RE-RAM, αναλύοντας πώς αυτά τα ζητήματα επηρεάζουν την προοπτική τους στην αγορά. Ορισμένα από τα μειονεκτήματα των RE-RAMS είναι [11]:

1. Προκλήσεις στο αναλογικό πεδίο: Η εργασία σε αναλογικό πεδίο έχει αρκετές προκλήσεις όπως θόρυβος (noise), μη γραμμικότητες I-V (nonlinear I-V characteristics), κλπ. Επιπρόσθετα, χρησιμοποιώντας αναλογικό κύκλωμα, η επικοινωνία με τα ψηφιακά απαιτεί την χρήση ADCs/DACs, οι οποίοι, υποβαθμίζουν την ακρίβεια του σήματος. Για παράδειγμα, ένας μετατροπέας ADC/DAC μπορεί να χρειαστεί 85% έως 98% της συνολικής επιφάνειας ανά ισχύ ενός NCS (neuromorphic computing system). Η ψηφιακή μετάδοση σήματος αντιθέτως, επιτρέπει καλύτερο έλεγχο και λειτουργία υψηλής συχνότητας.
2. Προβλήματα αξιοπιστίας των RE-RAM: Ο υψηλός ρυθμός ελαττωμάτων και η διακύμανση των διαδικασιών (process variation - PV) οδηγούν σε ζητήματα αξιοπιστίας. Για παράδειγμα, λόγω "αστοχίας ενός bit", ένα στοιχείο μπορεί να κολλήσει σε υψηλή ή χαμηλή τιμή αγωγιμότητας. Ειδικά για μεγάλα NCS, μια εφαρμογή RE-RAM οδηγεί σε μεγάλη συμφόρηση καλωδίων και χαμηλή αξιοπιστία των λειτουργιών ανάγνωσης/εγγραφής λόγω πτώσης τάσης και process variation. Με την αύξηση του ποσοστού αστοχίας της συσκευής, η ακρίβεια του νευρωνικού δικτύου (NN) μειώνεται δραστικά.
3. Προκλήσεις για την επίτευξη υψηλής ακρίβειας και απόδοσης: Σε σύγκριση με τη SRAM, η RE-RAM έχει υψηλό λόγο ενέργειας εγγραφής (write) ως προς την ενέργεια μεταφοράς δεδομένων (latency) κάτι που αυξάνει τη συνολική κατανάλωση ενέργειας. Οι περιορισμοί της RE-RAM, π.χ. στην συνδεσμολογία αντιστάσεων σε σειρά (series resistance) και σε ανεπιθύμητες διαδρομές ηλεκτρικού ρεύματος (sneak path), μειώνουν περαιτέρω την απόδοση. Επιπλέον, κατά τη διάρκεια της εκπαίδευσης (training) ενός NN, ο ακριβής συντονισμός της RE-RAM απαιτεί συχνές ενημερώσεις βαρών και μεγάλο αριθμό επαναλήψεων εκπαίδευσης για σύγκλιση. Αυτό οδηγεί σε μεγάλο αριθμό εγγραφών και μεγάλη κατανάλωση ενέργειας. Τα μη ιδανικά χαρακτηριστικά της RE-RAM, π.χ., η PV και η απότομη συμπεριφορά κατά τη διαδικασία SET όπως περιεγράφηκε σε προηγούμενο κεφάλαιο, επιβαρύνουν περαιτέρω τον συντονισμό μιας RE-RAM. Αν και τα σφάλματα που οφείλονται σε αστοχίες των RE-RAM ή στην αναλογική λειτουργία μπορούν να ελαχιστοποιηθούν με αυξημένη προπόνηση (training), αυτό οδηγεί σε

“ποινή” καθυστέρησης (latency)/ενέργειας (energy) και επιδεινώνει τα προβλήματα αντοχής της RE-RAM. Επιπλέον, η επανεκπαίδευση μπορεί να μην είναι επαρκής σε περίπτωση υψηλού ποσοστού σφάλματος.

4. Περιορισμοί στα Νευρωνικά δίκτυα: Δεν μπορούν να υλοποιηθούν όλες οι αρχιτεκτονικές/επίπεδα (architectures / layers) νευρωνικών δικτύων χρησιμοποιώντας RE-RAM, π.χ. τα επίπεδα LRN (local response normalization) με crossbar (διαδικασία που έχει περιγράψει σε προηγούμενο κεφάλαιο).

Συμπεράσματα

Από την παρούσα θεωρητική μελέτη συμπεραίνουμε πως η χρήση των RE-RAM παρουσιάζει εξαιρετικές δυνατότητες για τη βελτίωση της απόδοσης και τη μείωση της κατανάλωσης ενέργειας στις σύγχρονες υπολογιστικές αρχιτεκτονικές. Η ενσωμάτωση υπολογιστικής δυνατότητας απευθείας στη μνήμη επιτρέπει την επεξεργασία δεδομένων με μεγαλύτερη ταχύτητα και μικρότερο ενεργειακό κόστος.

Επιπρόσθετα, η δυνατότητα των RE-RAM να λειτουργούν υπό διάφορες ηλεκτρικές συνθήκες και το μικρότερο μέγεθος τους σε σχέση με παραδοσιακές μνήμες, προσφέρουν σημαντικά πλεονεκτήματα για εφαρμογές σε συσκευές όπου ο χώρος και η κατανάλωση ενέργειας είναι κρίσιμοι παράγοντες, όπως τα φορητά και τα ενσωματωμένα συστήματα.

Αφήνεται προς διερεύνηση το κατά πόσον είναι γρήγορα εφικτή η χρήση των RE-RAM και κατά πόσο αποτελεσματικές είναι σε σύγκριση με τις υπάρχουσες RAM για τον υπολογισμό εφαρμογών στην γραμμική άλγεβρα όπως οι πράξεις πινάκων. Με αυτόν τον τρόπο επισημαίνεται στον μελλοντικό αναλυτή η μελέτη των τεχνικών χαρακτηριστικών ακρίβειας και απόδοσης των RERAM.

Τέλος, οι προκλήσεις που αντιμετωπίζουν οι RE-RAM, όπως η διατήρηση της αξιοπιστίας τους κατά τη διάρκεια μακροχρόνιων κύκλων λειτουργίας και οι τεχνολογικοί περιορισμοί στην κλιμάκωσή τους, είναι πεδία που απαιτούν περαιτέρω έρευνα και ανάπτυξη. Η επικέντρωση σε αυτούς τους τομείς μπορεί να οδηγήσει σε βελτιωμένες τεχνολογίες μνήμης που θα είναι καταλληλότερες για ευρεία υιοθέτηση στο μέλλον.

1 Βιβλιογραφία

[1] Yi-Jiun Chen, Hsin-Lu Chen, Kow-Ming Chang, Tai-Fa Young, Hydrogen induced redox mechanism in amorphous carbon resistive random access memory, January 2014, Nanoscale Research Letters 9(1):52, DOI:10.1186/1556-276X-9-52, Source PubMed License CC BY 2.0

[2] Amit Prakash, Hyunsang Hwang, Multilevel Cell Storage and Resistance Variability in Resistive Random Access Memory, From the journal Physical Sciences Reviews, <https://doi.org/10.1515/psr-2016-0010>

[3] Ji-Hyun Hur, Kyung Min Kim, Man Chang, Seung Ryul Lee, Dongsoo Lee, Chang Bum Lee, Myoung-Jae Lee, Young-Bae Kim, Chang-Jung Kim and U-In Chung, Modeling for multilevel switching inoxide-based bipolar resistive memory, Semiconductor Laboratory, Samsung Advanced Institute of Technology, Gyeonggi-Do 446-712, Korea, E-mail: jhhur@samsung.com and yb305.kim@samsung.com Received 15 February 2012, in final form 18 April 2012 Published 10 May 2012

[4] FEBRUARY 21ST, 2019 - BY: MARK LAPEDUS, New approaches are competing for attention as scaling benefits diminish. In-Memory Vs. Near-Memory Computing. Ανακτήθηκε από In-Memory Vs. Near-Memory Computing (semiengineering.com)

[5] Furqan Zahoor, Tun Zainal Azni Zulkifli & Farooq Ahmad Khanday, Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications, Nano Review, Published: 22 April 2020, Volume 15, article number 90, (2020)

[6] Ximing Qiao, Xiong Cao, Huanrui Yang, Linghao Song, and Hai Li Duke University, Durham, North Carolina, USA, AtomLayer: A Universal ReRAM-Based CNN Accelerator with Atomic Layer Computation. June 2018 DOI:[10.1145/3195970.3195998](https://doi.org/10.1145/3195970.3195998) Conference: the 55th Annual Design Automation Conference

[7] Resistive RAM (ReRAM/RRAM) Memory utilizing resistive hysteresis, Ανακτήθηκε από https://semiengineering.com/knowledge_centers/memory/non-volatile-memory-nvm/resistive-ram/

- [8] Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications, Nano Review, Published: 22 April 2020, Volume 15, article number 90, (2020), Furqan Zahoor, Tun Zainal Azni Zulkifli & Farooq Ahmad Khanday
- [9] University of Southampton, ανακτήθηκε από τον ιστότοπο, https://www.southampton.ac.uk/advancedmaterials/research_projects/reram.page
- [10] Revolutionizing Data Storage: The Promise of ReRAM (Resistive Random-Access Memory), Drisha Gupta, September 18, 2023. Ανακτήθηκε από τον ιστότοπο <https://www.linkedin.com/pulse/revolutionizing-data-storage-promise-reram-resistive-memory-gupta>
- [11] Ανακτήθηκε από τον ιστότοπο <https://www.researchgate.net/post/Why-resistive-switching-memory-devicesReRAM-have-not-become-popular-despite-showing-great-promises-over-conventional-RAMs>