



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΤΩΝ ΚΑΤΕΡΓΑΣΙΩΝ

Εφαρμογή μοντέλων νευρωνικών δικτύων
σε προσομοιώσεις μοριακής δυναμικής
για τον προσδιορισμό της ατομικής δομής
υλικών

Διπλωματική εργασία
του φοιτητή
Πλακούτση Χρήστου

Επιβλέπων: Επικ. Καθ. Α. Μαρκόπουλος

ΑΘΗΝΑ, ΟΚΤΩΒΡΙΟΣ 2018

Αφιερωμένο στη Βίκυ και τον Μιχάλη

Life can only be understood backwards; but it must be lived
forwards.

Søren Kierkegaard

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια των σπουδών μου στη σχολή Μηχανολόγων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Το γενικότερο θέμα της διπλωματικής είναι η εφαρμογή μοντέλων νευρωνικών δικτύων σε προσομοιώσεις μοριακής δυναμικής για τον προσδιορισμό της ατομικής δομής υλικών. Αποτελεί μια σύγχρονη και ενδιαφέρουσα εφαρμογή των νευρωνικών δικτύων και μπορεί να μειώσει κατά πολύ το χρόνο προσομοιώσεων.

Θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο διδάκτορα Νίκο Κάρκαλο για τη συνεχή, μεθοδική και ουσιαστική βοήθεια που μου παρείχε σε όλα τα στάδια δημιουργίας της εργασίας αυτής. Ακόμα, θα ήθελα να ευχαριστήσω τον κ. Μαρκόπουλο Άγγελο που μου έδωσε τη δυνατότητα ενασχόλησης με ένα ενδιαφέρον θέμα και για την εμπιστοσύνη που μου έδειξε στην ανάληψη του παραπάνω θέματος.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου, που μου συμπαραστάθηκαν σε όλα τα χρόνια φοίτησής μου.

ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας διπλωματικής εργασίας είναι η διερεύνηση της δυνατότητας εφαρμογής νευρωνικών δικτύων σε προσομοιώσεις μοριακής δυναμικής και συγκεκριμένα σε προσομοιώσεις προσδιορισμού της ατομικής δομής υλικών. Η εργασία ασχολείται με τη γεωμετρική βελτιστοποίηση ενός συστήματος σεμεντίτη (Fe_3C) 512 ατόμων. Χρησιμοποιώντας δεδομένα από προσομοιώσεις μοριακής δυναμικής επιχειρείται ο προσδιορισμός της ατομικής δομής με την ελάχιστη ενέργεια με χρήση νευρωνικών δικτύων. Προκειμένου να προσδιοριστεί το αποδοτικότερο μοντέλο νευρωνικών δικτύων επιχειρείται επιπλέον η βελτιστοποίηση των χαρακτηριστικών του.

Για την επίλυση του προβλήματος προσδιορισμού της αρχιτεκτονικής των νευρωνικών δικτύων χρησιμοποιήθηκε μέθοδος δοκιμής και πλάνης (trial and error) και γενετικοί αλγόριθμοι. Από τη διαδικασία δοκιμής και πλάνης παρατηρήθηκε ότι το MSE κατά τη διαδικασία μάθησης όπως και το ποσοστιαίο σφάλμα (MAPE) μειώνονται με την αύξηση των νευρώνων, όμως από κάποιο αριθμό νευρώνων και πάνω το δίκτυο μειώνει την ικανότητα γενίκευσής του. Στη συνέχεια αφού προσδιορίστηκαν οι βέλτιστες παράμετροι του γενετικού αλγορίθμου, διερευνήθηκαν οι βέλτιστες τιμές των παραμέτρων του νευρωνικού δικτύου με βάση το MSE, το MAPE καθώς και συνδυασμό τους. Με τους γενετικούς αλγόριθμους παρατηρήθηκε ότι, εκτός από τους αριθμούς των νευρώνων, σημαντικό ρόλο παίζουν και δευτερεύοντες παράγοντες όπως ο διαχωρισμός του συνόλου των δεδομένων σε training και testing για την ικανοποιητική απόδοση των νευρωνικών δικτύων. Τέλος, οι ατομικές δομές που προήλθαν από τα βέλτιστα νευρωνικά δίκτυα και χρησιμοποιήθηκαν ως είσοδοι στις νέες προσομοιώσεις μοριακής δυναμικής, επιτάχυναν τη διαδικασία γεωμετρικής βελτιστοποίησης σε ποσοστό πάνω από 33%.

ABSTRACT

The aim of the present diploma thesis is the investigation of the applicability of artificial neural networks in molecular dynamics simulations and specifically in the definition of materials' atomic structure. The thesis deals with the geometry optimization of a system of 512 atoms of cementite (Fe_3C). Using data from molecular dynamics simulations, the definition of the atomic structure with the least energy is attempted, using neural networks. In order to define the most efficient architecture of neural networks, the optimization of their features is also attempted.

To solve the problem of defining the architecture of the neural networks a trial and error method has been used, along with genetic algorithms. From the trial and error it has been observed that the MSE during training as well as the MAPE are reduced while the number of neurons are increased, but, from a number of neurons upwards, the network reduces its capability to generalize efficiently. Then, after defining the best parameters of the genetic algorithm, the optimum parameters of the neural network have been investigated using MSE, MAPE, and a combination of both as the genetic algorithm's objective function. Using genetic algorithms it has been observed that, apart from the number of neurons in the hidden layers of the network, an important role is being played by secondary parameters in the efficiency of the network, such as the division percentage of the dataset into training and testing datasets. Finally, the atomic structures created by the optimum neural networks and were used as inputs in new molecular dynamics simulations, reduced the process of geometry optimization by a percentage of at least 33%.

ΠΕΡΙΕΧΟΜΕΝΑ

1	ΕΙΣΑΓΩΓΗ	1
2	ΜΟΡΙΑΚΗ ΔΥΝΑΜΙΚΗ	2
	2.1.1 <i>Εισαγωγή</i>	2
	2.1.2 <i>Θεωρητικό υπόβαθρο</i>	4
	2.1.3 <i>Συναρτήσεις δυναμικού</i>	7
	2.1.4 <i>Τρόποι υπολογισμού αλληλεπιδράσεων</i>	13
	2.1.5 <i>Μέθοδοι ολοκλήρωσης</i>	15
	2.1.6 <i>Ιστορική αναδρομή</i>	18
2.2	ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	20
	2.2.1 <i>Εισαγωγή</i>	20
	2.2.2 <i>Ιστορική αναδρομή</i>	29
2.3	ΥΛΙΚΑ	31
	2.3.1 <i>Σεμεντίτης - Fe₃C</i>	31
2.4	ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ	34
	2.4.1 <i>Ιστορική Αναδρομή</i>	34
	2.4.2 <i>Εισαγωγή</i>	34
3	ΔΙΕΘΝΗΣ ΒΙΒΛΙΟΓΡΑΦΙΑ	40
4	ΠΕΡΙΓΡΑΦΗ ΠΡΟΒΛΗΜΑΤΟΣ	46
5	ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	49
5.1	ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΡΧΙΚΗΣ ΠΡΟΣΟΜΟΙΩΣΗΣ.....	49
5.2	ΑΠΟΤΕΛΕΣΜΑΤΑ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	50
	5.2.1 <i>3-5-X-3</i>	51
	5.2.2 <i>3-10-X-3</i>	54
	5.2.3 <i>3-15-X-3</i>	56
	5.2.4 <i>3-20-X-3</i>	57
5.3	ΑΠΟΤΕΛΕΣΜΑΤΑ ΓΕΝΕΤΙΚΩΝ ΑΛΓΟΡΙΘΜΩΝ.....	59
5.4	ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΕΛΙΚΩΝ ΠΡΟΣΟΜΟΙΩΣΕΩΝ	68
5.5	ΣΥΓΚΡΙΣΗ ΧΡΟΝΩΝ.....	71
6	ΣΥΜΠΕΡΑΣΜΑΤΑ	73
7	ΒΙΒΛΙΟΓΡΑΦΙΑ	75

Κεφάλαιο 1

1 ΕΙΣΑΓΩΓΗ

Στην παρούσα εργασία γίνεται μια μελέτη για την εφαρμογή των τεχνητών νευρωνικών δικτύων σε προσομοιώσεις μοριακής δυναμικής που έχουν ως στόχο τον προσδιορισμό της ατομικής δομής υλικών. Ο προσδιορισμός της ατομικής δομής υλικών είναι ιδιαίτερα σημαντικός ιδίως σε περιπτώσεις κραμάτων ή άμορφων υλικών. Η διαδικασία αυτή ανάλογα με τον αριθμό των ατόμων του συστήματος μπορεί να είναι αρκετά χρονοβόρα γι' αυτό η ύπαρξη μιας αρχικής δομής που προσεγγίζει ικανοποιητικά την τελική επιθυμητή δομή επιταχύνει σημαντικά την διαδικασία. Ένας τρόπος εντοπισμού μιας δομής που προσεγγίζει ικανοποιητικά την τελική δομή είναι η χρήση νευρωνικών δικτύων.

Η παρούσα εργασία αποτελείται από τα εξής κεφάλαια:

Στο **κεφάλαιο 2**, γίνεται παρουσίαση των βασικών χαρακτηριστικών της μεθόδου της μοριακής δυναμικής, των νευρωνικών δικτύων και των γενετικών αλγορίθμων.

Στο **κεφάλαιο 3**, γίνεται μια βιβλιογραφική ανασκόπηση σχετική με εργασίες που αφορούν την χρήση νευρωνικών δικτύων σε προσομοιώσεις μοριακής δυναμικής.

Στο **κεφάλαιο 4**, παρουσιάζεται η μεθοδολογία που αναπτύσσεται στην παρούσα διπλωματική εργασία σχετικά με τον προσδιορισμό της ατομικής δομής υλικού με τη χρήση νευρωνικών δικτύων εκπαιδευμένων με δεδομένα προερχόμενα από προσομοιώσεις μοριακής δυναμικής.

Στο **κεφάλαιο 5**, παρουσιάζονται τα αποτελέσματα των προσομοιώσεων μοριακής δυναμικής, των νευρωνικών δικτύων και των γενετικών αλγορίθμων που χρησιμοποιήθηκαν για τη βελτιστοποίηση των νευρωνικών δικτύων και γίνεται μια συνοπτική σύγκριση του χρόνου προσομοιώσεων.

Στο **κεφάλαιο 6** παρουσιάζονται τα συμπεράσματα της παρούσας διπλωματικής εργασίας και παρατίθενται προτάσεις για μελλοντική μελέτη.

Κεφάλαιο 2

2 Μοριακή δυναμική

2.1.1 Εισαγωγή

Η μοριακή δυναμική είναι μια μέθοδος υπολογιστικής προσομοίωσης και αφορά την αριθμητική επίλυση των κλασσικών εξισώσεων κίνησης για μια ομάδα ατόμων. Για να μελετηθεί ένα σύστημα σε ατομική κλίμακα, όπως συμβαίνει με την προσομοίωση σε νανοκλίμακα με τη μοριακή δυναμική, πρέπει να υπολογιστούν οι δυνάμεις που αναπτύσσονται μεταξύ των ατόμων, επειδή οι δυνάμεις αυτές καθορίζουν κατά κύριο λόγο τα φαινόμενα που λαμβάνουν χώρα.

Για μια προσομοίωση με τη μέθοδο της μοριακής δυναμικής, χρειάζονται τρία βασικά συστατικά :

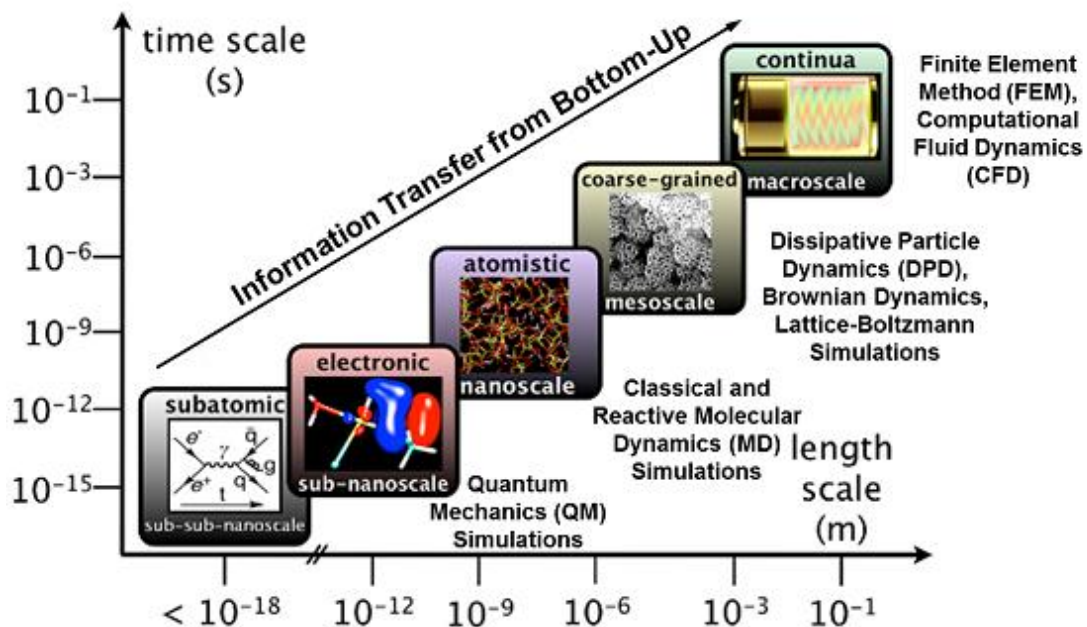
- Πρώτον, χρειάζεται ένας νόμος για την περιγραφή των αμοιβαίων αλληλεπιδράσεων μεταξύ των ατόμων στο σύστημα που μελετάται. Απο εκεί μπορούν να υπολογιστούν, έχοντας γνωστές τις θέσεις των ατόμων, οι δυνάμεις που επιδρούν στα άτομα και οι τάσεις στους τοίχους του δοχείου/κουτιού όπου βρίσκονται τα άτομα. Ο νόμος αυτός είναι στη γενική περίπτωση άγνωστος, αλλά μπορεί να προσεγγιστεί με διαφορετικούς βαθμούς ακρίβειας από ένα δυναμικό, ή μπορεί να μοντελοποιηθεί με υπολογισμούς ηλεκτρονιακής δομής (electronic structure calculations).
- Δεύτερον, χρειάζεται ένας αλγόριθμος για την αριθμητική ενσωμάτωση των εξισώσεων κίνησης των ατόμων του συστήματος.
- Τρίτον, για την επίλυση των εξισώσεων κίνησης, πρέπει να δοθούν ορισμένες έγκυρες αρχικές συνθήκες στον αλγόριθμο, δηλαδή έγκυρες αρχικές θέσεις και ταχύτητες για όλα τα άτομα στο σύστημα.

Με μια τέτοια προσομοίωση υπολογίζεται με προσεγγιστικό τρόπο η αληθινή δυναμική του μελετούμενου συστήματος, και είναι δυνατή η παρακολούθηση κάθε ατόμου με πολύ λεπτομερή τρόπο. Έτσι οι προσομοιώσεις MD βοηθούν στην παρατήρηση σημαντικών διεργασιών που λαμβάνουν χώρα σε ατομικό ή μοριακό επίπεδο, κάτι που συχνά είναι αδύνατο να παρατηρηθεί πειραματικά λόγω ανεπαρκούς ανάλυσης. Επιπλέον, σε μια προσομοίωση είναι εύκολο να προετοιμαστούν οι συνθήκες (θερμοκρασία, πίεση, ατομικές θέσεις) κατά βούληση, και υπάρχει ένα επίπεδο ελέγχου τους που είναι συχνά πολύ μεγαλύτερο από το αντίστοιχο πειραματικό.

Εκτός από τις αριθμητικές προσεγγίσεις κατά την ενσωμάτωση των εξισώσεων κίνησης, υπάρχουν δύο βασικές προσεγγίσεις εγγενείς στις προσομοιώσεις MD:

- Πρώτον, η υπόθεση ότι τα άτομα συμπεριφέρονται ως κλασσικά σώματα, δηλαδή ακολουθούν τις Νευτώνειες εξισώσεις κίνησης. Η ακρίβεια της προσέγγισης αυτής εξαρτάται από το μελετούμενο σύστημα, και στις συνθήκες προσομοίωσης. Στις περισσότερες περιπτώσεις οι κβαντικές επιδράσεις στη δυναμική των ατόμων είναι σχετικά μικρές, με εξαίρεση τα ελαφρά άτομα και τις χαμηλές θερμοκρασίες. Σε αυτές τις περιπτώσεις, και γενικά σε περιπτώσεις όπου οι κβαντικές επιδράσεις δεν μπορούν να παραβλεφθούν, συνίσταται η προσέγγιση Path Integral [1] ή κάποια παρόμοια μέθοδος.
- Δεύτερον, το μοντέλο που περιγράφει τις αλληλεπιδράσεις μεταξύ των ατόμων στο σύστημα. Μόνο μέσω μίας επαρκούς περιγραφής αυτών των αλληλεπιδράσεων είναι πιθανό να παρθούν χρήσιμες και αξιόπιστες πληροφορίες για τις ατομικές διεργασίες που συμβαίνουν στο σύστημα.

Tools: multi-scale computational framework that connects atomistic details (materials chemistry) to microstructures and overall system performance.



Σχήμα 1 : Σύγκριση προσομοιώσεων με βάση τη χρονική κλίμακα και το επίπεδο ακρίβειας

2.1.2 Θεωρητικό υπόβαθρο

2.1.2.1 Εξισώσεις κίνησης

Χρησιμοποιώντας μία κατάλληλη συνάρτηση δυναμικού και υπολογίζοντας τις δυνάμεις αλληλεπίδρασης των ατόμων, αφού αυτές συσχετιστούν με τις θέσεις των ατόμων, δημιουργείται ένα σύστημα διαφορικών εξισώσεων η επίλυση του οποίου δίνει την απόκριση του συστήματος. Για το λόγο αυτό χρειάζονται οι δυναμικές εξισώσεις του συστήματος. Η επιλογή των εξισώσεων αυτών γίνονται ανάλογα με την πολυπλοκότητα της μοντελοποίησης του προβλήματος. Στη νανοκλίμακα τον πρώτο λόγο έχει η κβαντική μηχανική και η εξίσωση Schrödinger, όμως είναι πολύπλοκη υπολογιστικά και βρίσκει εφαρμογή σε προσομοιώσεις μικρών συστημάτων. Στα πλαίσια της παρούσας μελέτης μπορούν να χρησιμοποιηθούν, θυσιάζοντας ένα κομμάτι ακρίβειας, οι εξισώσεις του Lagrange, του Hamilton και του Newton.

Μερικά προβλήματα στη στατιστική μηχανική είναι ακριβώς επιλύσιμα. Αυτό σημαίνει ότι ο ολοκληρωμένος προσδιορισμός των μικροσκοπικών ιδιοτήτων ενός συστήματος (π.χ. η Χαμιλτονιανή ενός ιδανικού μοντέλου όπως το τέλειο αέριο ή ο κρύσταλλος του Einstein) οδηγεί άμεσα σε ένα σύνολο από μακροσκοπικές ιδιότητες (π.χ. μια εξίσωση κατάστασης όπως η $PV = NRT$).

Η υπολογιστική προσομοίωση συσχετίζει άμεσα τις μικροσκοπικές λεπτομέρειες ενός συστήματος (τις μάζες των ατόμων, τις αλληλεπιδράσεις μεταξύ τους, τη μοριακή γεωμετρία κ.α) με μακροσκοπικές ιδιότητες πειραματικού ενδιαφέροντος (εξίσωση κατάστασης, συντελεστής μεταφοράς, κ.α.)

Η μικροσκοπική κατάσταση ενός συστήματος μπορεί να περιγραφεί ως συσχετισμός των θέσεων και των ορμών ενός συστατικού συνόλου σωματιδίων : των ατόμων και των μορίων. Εντός της προσέγγισης Born-Oppenheimer [2], είναι εφικτό να εκφραστεί η Χαμιλτονιανή ενός συστήματος ως συνάρτηση των πυρηνικών μεταβλητών. Με μια ακόμη προσέγγιση ότι η κλασσική περιγραφή είναι επαρκής, η Χαμιλτονιανή ενός συστήματος N μορίων γράφεται ως άθροισμα συναρτήσεων κινητικής και δυναμικής ενέργειας του συνόλου των συντεταγμένων q_i και ορμών p_i κάθε μορίου i . Υιοθετώντας το συμβολισμό

$$q = (q_1, q_2, \dots, q_N)$$
$$p = (p_1, p_2, \dots, p_N)$$

Ισχύει :

$$H(q, p) = H(p) + V(q),$$

Οι γενικευμένες συντεταγμένες q μπορούν απλά να είναι το σύνολο των καρτεσιανών συντεταγμένων r_i κάθε ατόμου (ή πυρήνα) στο σύστημα, όμως

μερικές φορές είναι χρήσιμο να αντιμετωπίζονται τα μόρια ως άκαμπτα σώματα, επομένως τότε το q θα συνίσταται από τις καρτεσιανές συντεταγμένες κάθε μοριακού κέντρου μάζας μαζί με ένα σύνολο μεταβλητών Ω_i που εκφράζουν το μοριακό προσανατολισμό. Σε κάθε περίπτωση το p είναι το κατάλληλο σύνολο συζυγών ορμών. Συνήθως, η κινητική ενέργεια H παίρνει τη μορφή

$$H = \sum_{i=1}^N \sum p_{ia}^2 / 2m_i$$

Όπου m_i είναι η μοριακή μάζα και ο δείκτης a συμπεριλαμβάνει τα διαφορετικά (x, y, z) στοιχεία της ορμής του μορίου i . [3]

Η δυναμική ενέργεια V περιέχει τις ενδιαφέρουσες πληροφορίες σχετικά με την ενδομοριακή αλληλεπίδραση: αν συμπεριφέρεται σχετικά σωστά, θα είναι δυνατό να δημιουργηθεί, από την H , μια εξίσωση κίνησης (σε Χαμιλτονιανή, Λαγκρανζιανή ή Νευτώνεια μορφή) που θα διέπει όλη τη χρονική εξέλιξη του συστήματος και όλες τις μηχανικές του ιδιότητες. Η λύση αυτής της εξίσωσης περιλαμβάνει τον υπολογισμό, από το V , των δυνάμεων f_i και των στροφορμών t_i που επιδρούν στα μόρια. [4]

Η δυναμική ενέργεια μπορεί να σπάσει σε όρους που εξαρτώνται από τις συντεταγμένες των ξεχωριστών ατόμων, ζευγών, τριπλετών, κ.ο.κ. :

$$V = \sum_i v_1(r_i) + \sum_i \sum_{j>i} v_2(r_i, r_j) + \sum_i \sum_{j>i} \sum_{k>j>i} u_3(r_i, r_j, r_k) + \dots$$

Ο πρώτος όρος αναπαριστά την επίδραση ενός εξωτερικού πεδίου (π.χ. τους τοίχους του κοντέινερ) στο σύστημα. Οι υπόλοιποι όροι αναπαριστούν σωματιδιακές αλληλεπιδράσεις. Ο δεύτερος όρος, το δυναμικό ζεύγους, είναι ο πιο σημαντικός. Το δυναμικό ζεύγους εξαρτάται μόνο από το μέγεθος της απόστασης $r_{ij} = |r_i - r_j|$ οπότε μπορεί να γραφεί $u_2(r_{ij})$.

2.1.2.2 Εξισώσεις Lagrange

Οι εξισώσεις Lagrange χρησιμοποιούνται για σύνθετα συστήματα και έχουν την μορφή

$$L = T - U$$

όπου T η κινητική και U η δυναμική ενέργεια ενός συστήματος

Η μέθοδος Lagrange θεμελιώνεται με το αξίωμα μεταβολών του Hamilton, δηλαδή ότι η τροχιά $q(t)$ ενός φυσικού συστήματος στο χώρο φάσεων αποτελεί στάσιμο σημείο του συναρτησιακού S , που ονομάζεται δράση, και είναι :

$$S = \int_{t_i}^{t_f} L\left(t, q(t), \frac{\partial q}{\partial t}(t)\right) dt$$

Με εφαρμογή του θεμελιώδους λήμματος του λογισμού των μεταβολών προκύπτει το σύστημα διαφορικών εξισώσεων Euler-Lagrange :

$$\nabla_x L(t, q(t), \dot{q}(t)) = \frac{d}{dt} \nabla_u L(t, q(t), \dot{q}(t))$$

με $x=q(t)$ και $u=\frac{\partial q}{\partial t}(t)$

2.1.2.3 Εξισώσεις Hamilton

Αν στα παραπάνω αντικαταστήσουμε τις γενικευμένες ταχύτητες με τις γενικευμένες ορμές, δηλαδή

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = \nabla \dot{q} L$$

και ορίσουμε τη Χαμιλτονιανή H ως :

$$H = \dot{q}^T p - L$$

τότε με εφαρμογή της αρχής Hamilton προκύπτουν οι δύο δυναμικές εξισώσεις πρώτης τάξης του συστήματος

$$\dot{q} = \nabla \dot{q} H$$

$$\dot{p} = -\nabla p H$$

2.1.2.4 Εξισώσεις Newton

Η πιο συνήθης σχέση δυναμικού και δύναμης είναι

$$F = -\nabla V_{ij}(r)$$

και οι πιο απλές εξισώσεις κίνησης προκύπτουν από το δεύτερο νόμο του Newton :

$$m \frac{\partial^2 r_i}{\partial t^2}(t) = F_i(t) = - \sum_{j=1, j \neq i}^{N_m} \nabla_x V_{ij}(\|r_{ij}(t)\|)$$

όπου r_i η θέση του ατόμου i , F_i η συνισταμένη δύναμη που δέχεται το άτομο i , F_{ij} η δύναμη που δέχεται το άτομο i από το άτομο j , M η μάζα του ατόμου i και N_m το πλήθος των ατόμων.

Με την ολοκλήρωση των παραπάνω εξισώσεων δίνονται οι τροχιές των ατόμων και λόγω του τρίτου νόμου του Newton ($F_{ij} = -F_{ji}$) αρκεί μόνο ένας υπολογισμός για την αλληλεπίδραση ενός ζεύγους ατόμων.

2.1.3 Συναρτήσεις δυναμικού

Οι συναρτήσεις δυναμικού μοντελοποιούν την ενέργεια που έχει ένα ζεύγος σωμάτων λόγω των δυνάμεων αλληλεπίδρασης. Η απώθησή τους όταν το ένα πλησιάζει το άλλο οφείλεται στην αλληλεπικάλυψη μη συζευγμένων ηλεκτρονιακών νεφών, ενώ η έλξη τους σε μεγάλες αποστάσεις δημιουργείται από τη συσχέτιση των ηλεκτρονίων.

Η ακρίβεια των δυναμικών που επιλέγονται για την προσομοίωση καθορίζει κατά πολύ την ποιότητα της προσομοίωσης, και γι' αυτό όπως φαίνεται παρακάτω υπάρχει μεγάλο ερευνητικό ενδιαφέρον για τη δημιουργία δυναμικών με πολύ μεγάλη ακρίβεια, διαδικασία η οποία μπορεί να πραγματοποιηθεί και με νευρωνικά δίκτυα.

Ένα δυναμικό που έχει δημιουργηθεί για συγκεκριμένη κατηγορία υλικών δεν είναι κατάλληλο για χρησιμοποίηση σε άλλες κατηγορίες, επομένως είναι αναγκαία η δημιουργία διαφορετικών δυναμικών για κάθε είδος υλικού. Τα περισσότερα γνωστά δυναμικά αφορούν μονοφασικά και μονοκρυσταλλικά υλικά με ένα είδος δεσμού μεταξύ των ατόμων. Για πολυκρυσταλλικά υλικά, υλικά μερικώς με ομοιοπολικούς και μερικώς με ιοντικούς δεσμούς, μεταλλικά κράματα, η δημιουργία κατάλληλων δυναμικών είναι αντικείμενο μελέτης και έρευνας. Υπάρχουν δύο προσεγγίσεις για τον υπολογισμό των συναρτήσεων δυναμικού, *ab initio* και εμπειρικές.

Με σκοπό τη διερεύνηση των γενικών ιδιοτήτων των υλικών και τη σύγκριση με τη θεωρία, μπορούν να χρησιμοποιηθούν εξιδανικευμένα μοντέλα δυναμικού ζεύγους. Παρακάτω παρουσιάζονται ορισμένα δυναμικά απλά και χρήσιμα για υπολογιστικές προσομοιώσεις.

Αυτές είναι :

- Hard-sphere potential :

$$U^{HS}(r) = \begin{cases} \infty, & r < \sigma \\ 0, & r \geq \sigma \end{cases}$$

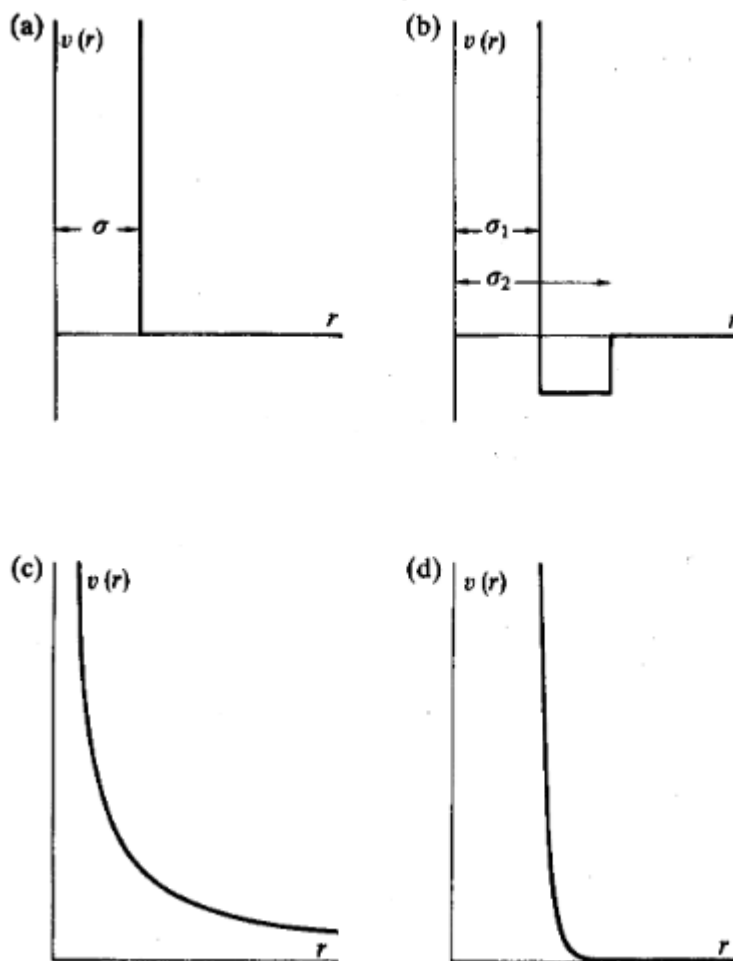
- Square-well potential

$$U^{SW}(r) = \begin{cases} \infty, & r < \sigma_1 \\ -\varepsilon, & \sigma_1 \leq r \leq \sigma_2 \\ 0, & r \geq \sigma_2 \end{cases}$$

- Soft-sphere potential

$$U^{SS}(r) = \varepsilon \left(\frac{\sigma}{r}\right)^v = ar^{-v}$$

όπου v μια παράμετρος που συνήθως επιλέγεται να είναι ακέραιος αριθμός. Όσο το v αυξάνεται, το soft-sphere potential τείνει να μοιάζει όλο και περισσότερο στο hard-sphere potential. Στον παρακάτω πίνακα παρουσιάζονται τα παραπάνω εξιδανικευμένα δυναμικά.

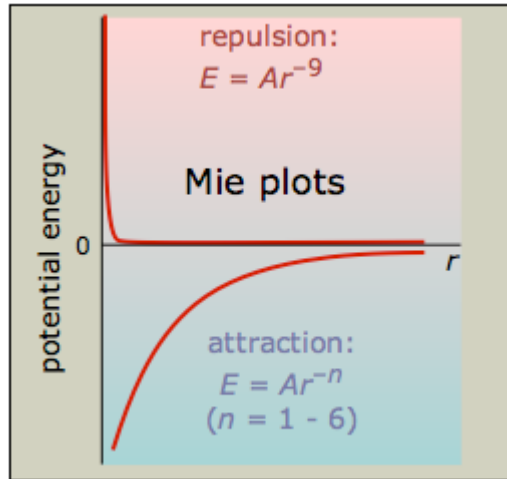


Σχήμα 2: Εξιδανικευμένα δυναμικά ζεύγους : a) hard sphere potential; b) square-well potential; c) soft-sphere potential with repulsion parameter $v=1$; d) soft-sphere potential with repulsion parameter $v=12$

- Δυναμικό Mie[5]

$$V_{ij} = \frac{n}{n-m} \left(\frac{n}{m}\right)^{\frac{m}{n-m}} \varepsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^n - \left(\frac{\sigma}{r_{ij}}\right)^m \right]$$

όπου η παράμετρος r_{ij} είναι η στιγμιαία απόσταση ανάμεσα στα άτομα i και j , σ είναι η απόσταση μηδενικού δυναμικού και εξαρτάται από τις ιδιότητες του υλικού, ε είναι η ενέργεια αποχωρισμού και n, m είναι σταθερές παράμετροι.

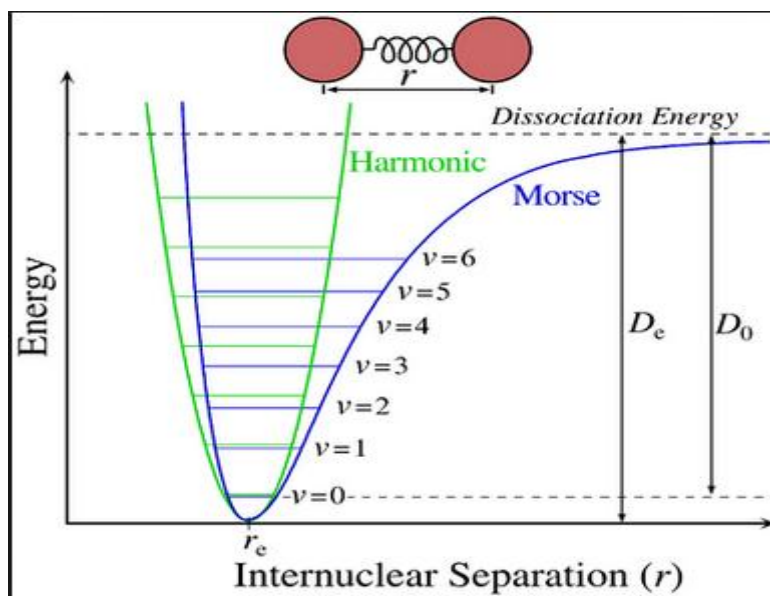


Σχήμα 3 : Δυναμικό Mie

- Δυναμικό Morse[6]

$$V_{ij} = D[e^{-2a(r_{ij}-r_e)} - 2e^{-a(r_{ij}-r_e)}]$$

όπου r_{ij} η απόσταση ανάμεσα στα σωματίδια i και j , r_e η απόσταση ισορροπίας ή ελάχιστου δυναμικού και D, a σταθερές που εξαρτώνται από τις ιδιότητες του υλικού.



Σχήμα 4 : Δυναμικό Morse

- Δυναμικό Lennard-Jones 12-6

$$U^{LJ}(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right)$$

που είναι ειδική περίπτωση του δυναμικού Mie για $n=12$ και $m=6$

- Δυναμικό Lennard-Jones 8-6

$$U^{LJ}(r) = \varepsilon\left(3\left(\frac{r_e}{r_{ij}}\right)^8 - 4\left(\frac{r_e}{r_{ij}}\right)^6\right)$$

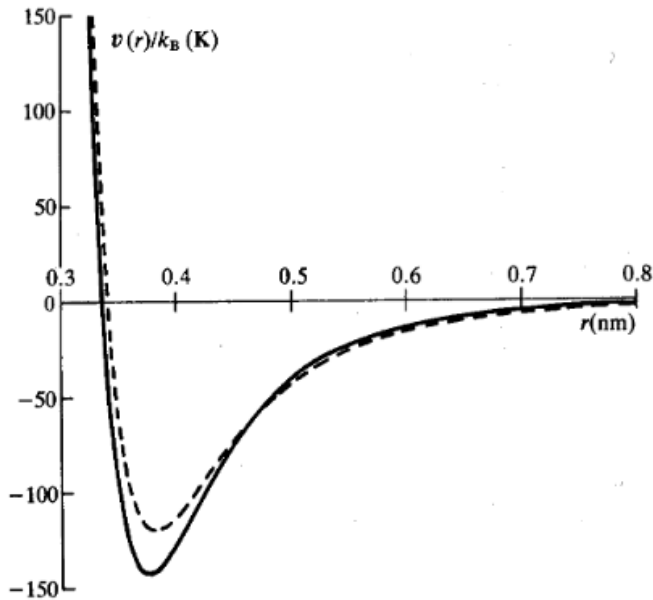
- Δυναμικό Lennard-Jones 9-6

$$U^{LJ}(r) = \varepsilon\left[2\left(\frac{r_e}{r_{ij}}\right)^9 - 3\left(\frac{r_e}{r_{ij}}\right)^6\right]$$

- Δυναμικό Lennard-Jones 9-3

$$U^{LJ}(r) = \frac{3\sqrt{3}}{2} \varepsilon \left[2 \left(\frac{\sigma}{r_{ij}} \right)^9 - 3 \left(\frac{\sigma}{r_{ij}} \right)^3 \right]$$

Στο παρακάτω σχήμα φαίνεται η διαφορά του θεωρητικού Lennard-Jones 12-6 μοντέλου από το πειραματικό και διευκολύνεται η σύγκριση με τα εξιδανικευμένα μοντέλα δυναμικού του σχήματος 1.

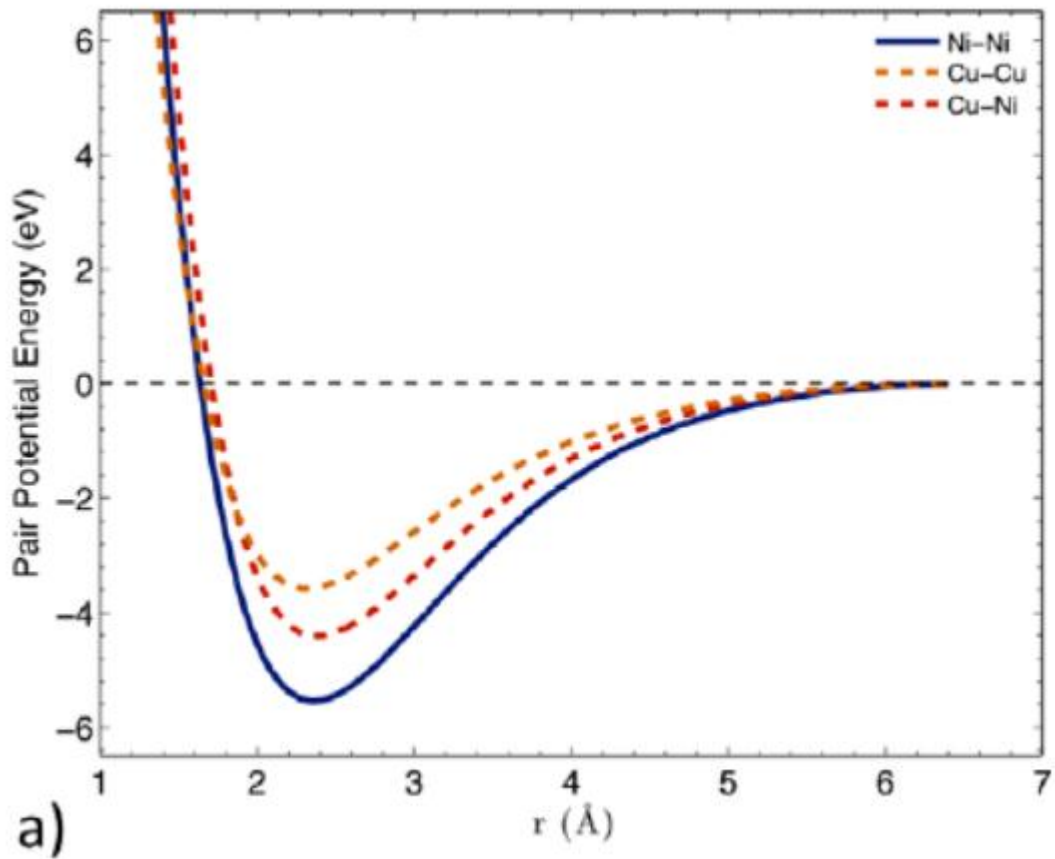


Σχήμα 5 : Δυναμικό ζεύγους για το στοιχείο αργό και σύγκριση με το θεωρητικό μοντέλο Lennard-Jones (διακεκομμένη).

- Δυναμικό EAM (embedded atom method)

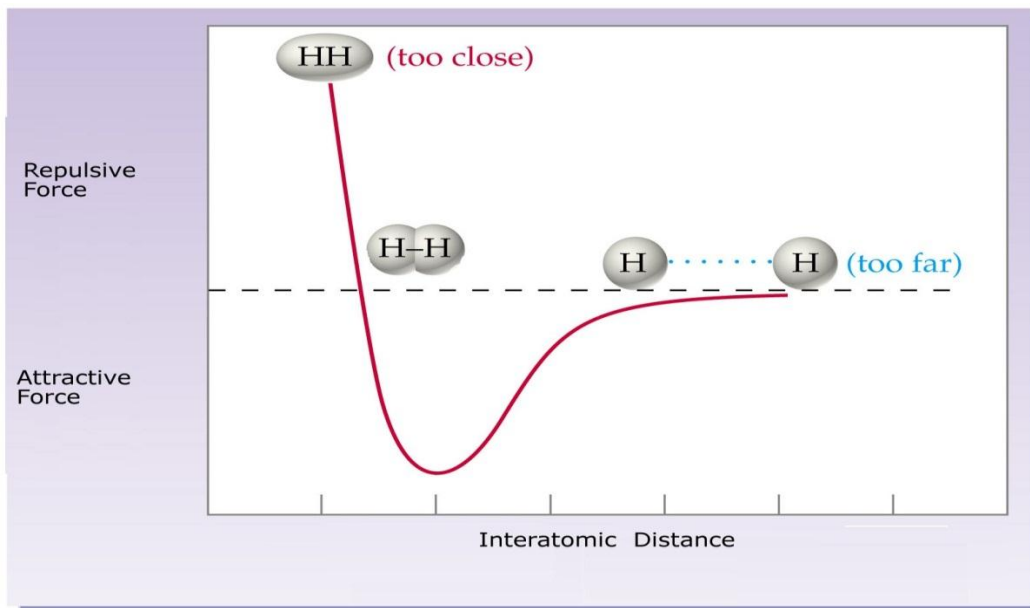
$$E_{ολ} = F_a \left(\sum_{i \neq j} \rho_{\beta}(r_{ij}) \right) + \frac{1}{2} \sum_{i \neq j} \Phi_{\alpha\beta}(r_{ij})$$

όπου r_{ij} η απόσταση μεταξύ των ατόμων i και j , $\Phi_{\alpha\beta}$ είναι μια συνάρτηση δυναμικού ζεύγους, ρ_{β} είναι η συνεισφορά στην πυκνότητα ηλεκτρονιακού φορτίου από το άτομο j τύπου β στη θέση του ατόμου i , και F είναι μια συνάρτηση ενσωμάτωσης που αναπαριστά την απαιτούμενη ενέργεια για την τοποθέτηση του ατόμου i τύπου α στο ηλεκτρονιακό νέφος. $E_{ολ}$ είναι η ολική και όχι η δυναμική ενέργεια.



Σχήμα 6: Δυναμικό EAM

Για την κατανόηση της επίδρασης του δυναμικού στα άτομα παρατίθεται το παρακάτω σχήμα :



Σχήμα 7 : Ερμηνεία του δυναμικού. Απωθητικές και ελκτικές δυνάμεις ως συνάρτηση της απόστασης

Μια από τις μεγαλύτερες δυσκολίες στις θεωρητικές προσεγγίσεις που προσπαθούν να περιγράψουν φυσικά και χημικά συστήματα είναι ο ακατάλληλος υπάρχων μαθηματικός εξοπλισμός για την επίλυση του προβλήματος N σωμάτων (many-body problem). Παρόλο που οι ιδιότητες ενός απομονωμένου σωματιδίου είναι θεμελιωμένες και οι στοιχειώδεις διαδικασίες που προκύπτουν όταν δύο τέτοια σωματίδια αλληλεπιδρούν περιγράφονται από γνωστούς νόμους, η συμπεριφορά συστημάτων με πολλά αλληλεπιδρώντα μόρια δεν μπορεί, στη γενική περίπτωση, να αντιμετωπιστεί θεωρητικά με ακριβή τρόπο. [7]

Ο αναλυτικός υπολογισμός της αλληλεπίδρασης ενός σχετικά μεγάλου αριθμού σωματιδίων συμπεριλαμβάνει την επίλυση είτε ολοκληρωμάτων πολλών διαστάσεων είτε διαφορικών εξισώσεων υψηλής τάξης. Οι μαθηματικές αυτές εκφράσεις μπορούν να αναχθούν σε διαχειρίσιμες εξισώσεις σε αραιά (dilute) συστήματα καθώς, σε αυτή την περίπτωση, η συμπεριφορά του συστήματος μπορεί να θεωρηθεί ως μια αλληλουχία δυαδικών αλληλεπιδράσεων. Στην περίπτωση μη αραιών (nondilute) συστημάτων, δηλαδή όταν η εμβέλεια των διασωματιδιακών δυνάμεων δεν είναι μικρή συγκρινόμενη με τη μέση διασωματιδιακή απόσταση, αναλυτικές θεωρίες έχουν αναπτυχθεί για να προσεγγίσουν το πρόβλημα N σωμάτων με διάφορους εвриματικούς τρόπους. Η πιο συνήθης στρατηγική είναι, αφού γίνει διαλογή ενός αντιπροσωπευτικού σωματιδίου, να αφηθεί υπό την επήρεια του δυναμικού των υπόλοιπων σωματιδίων, όταν αυτά διατηρούνται σταθερά σε μια μέση θέση. Το δυναμικό αυτό μπορεί να παρθεί από σαφή φυσικά μοντέλα ή με αυτόνομο τρόπο. [7]

Υπάρχουν πολλές διαφορετικές προσεγγίσεις για τον υπολογισμό του δυναμικού V . Σε αυτές, η κατανομή των ηλεκτρονίων στο σύστημα δεν μοντελοποιείται από ένα ισχύον δυναμικό, αλλά περιγράφεται από τη μέθοδο density functional theory [8]. Μια εκδοχή είναι η ηλεκτρονιακή πυκνότητα να αναπαρασταθεί από μια προέκταση του μοντέλου ελεύθερου ηλεκτρονίου (free electron model) [9]. Σε μια άλλη εκδοχή, ηλεκτρονιακοί βαθμοί ελευθερίας συμπεριλαμβάνονται στην περιγραφή, και τα ηλεκτρόνια αφήνονται να ηρεμήσουν κατά τη διάρκεια της προσομοίωσης με μια διαδικασία γνωστή ως τεχνητή απόπηση (simulated annealing). [10]

2.1.4 Τρόποι υπολογισμού αλληλεπιδράσεων

Μετά την επιλογή των κατάλληλων εξισώσεων κίνησης και συναρτήσεων δυναμικού, χρειάζεται η επιλογή του κατάλληλου τρόπου υπολογισμού των αλληλεπιδράσεων μεταξύ των ατόμων. Διακρίνονται τρεις κύριες μέθοδοι [11] :

- All pairs method

Με τη μέθοδο αυτή υπολογίζονται όλες οι N_m-1 δυνάμεις αλληλεπίδρασης για κάθε ένα από τα N_m άτομα, και είναι η πιο πολύπλοκη χρονικά και υπολογιστικά.

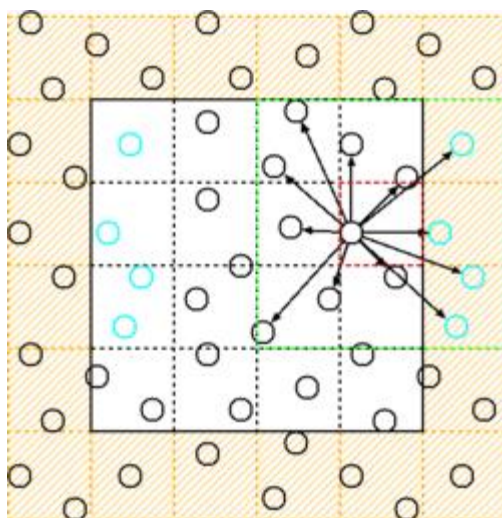
Η υπολογιστική πολυπλοκότητα της μεθόδου αυτής είναι

$$N_m \left(\frac{N_m - 1}{2} \right) = \theta(N_m)^2$$

αν ληφθεί υπόψη ο 3^{ος} νόμος του Νεύτωνα της δράσης-αντίδρασης.

- Cell subdivision-cell lists

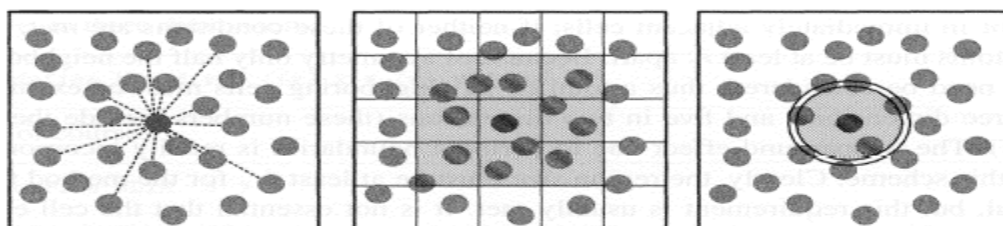
Στην περίπτωση αυτή εφαρμόζεται μία απόσταση αποκοπής r_c και υπολογίζονται οι δυνάμεις για κάθε άτομο μόνο μέσα στα όρια της απόστασης αυτής. Η περιοχή προσομοίωσης υποδιαιρείται σε κύτταρα, και κάθε κύτταρο, κυβικό ή τετραγωνικό, έχει διάσταση ακμής λίγο μεγαλύτερη από τη διάσταση αποκοπής. Άρα η αλληλεπίδραση περιορίζεται σε άτομα που βρίσκονται σε γειτονικά κύτταρα σε απόσταση μικρότερη της απόστασης αποκοπής. Λόγω συμμετρίας αρκεί να υπολογίζονται μόνο τα μισά γειτονικά κύτταρα συν το κύτταρο στο οποίο βρίσκεται το άτομο υπό εξέταση, μια μέθοδος που είναι χρήσιμη όταν η διάσταση της προσομοίωσης είναι μεγαλύτερη από $4r_c$.



Σχήμα 8 : Μέθοδος cell subdivision

- Neighbor-list method

Για να μειωθεί το υπολογιστικό κόστος της παραπάνω μεθόδου, ορίζεται μία λίστα γειτόνων εντός της περιοχής $d_{ij} < r_c + \Delta r$. Λόγω της σχετικά αργής μεταβολής του μικροσκοπικού περιβάλλοντος, η λίστα αυτή παραμένει ακριβής για αρκετές επαναλήψεις. Ο όρος Δr χρησιμοποιείται για τη διασφάλιση της μη εισόδου εντός εμβέλειας ατόμων που δεν περιλαμβάνονται στη λίστα μεταξύ των διαδοχικών ενημερώσεών της.



Σχήμα 9: Σύνοψη των τριών μεθόδων υπολογισμού της αλληλεπίδρασης

2.1.5 Μέθοδοι ολοκλήρωσης

Στην υπολογιστική υλοποίηση της μεθόδου της Μοριακής Δυναμικής, σημαντικό ρόλο παίζει ο τρόπος ολοκλήρωσης των διαφορικών εξισώσεων, προκειμένου να προκύψουν από αυτές τα επιθυμητά αποτελέσματα.

Από την αριθμητική ανάλυση είναι διαθέσιμες πολλές μέθοδοι ολοκλήρωσης διαφορικών εξισώσεων, η πλειοψηφία των οποίων όμως απορρίπτεται για χρήση στη μέθοδο της Μοριακής Δυναμικής. Ο λόγος που συμβαίνει κάτι τέτοιο είναι ότι οι μέθοδοι αυτές απαιτούν πάνω από έναν υπολογισμό των δυνάμεων αλληλεπίδρασης ανάμεσα στα άτομα σε κάθε χρονικό βήμα. Όπως φάνηκε και από τα προηγούμενα, ο υπολογισμός των αλληλεπιδράσεων αποτελεί το πιο πολύπλοκο από την άποψη του υπολογισμού τμήμα, επομένως χρήση των παραπάνω μεθόδων θα μεγάλωνε ακόμη περισσότερο την πολυπλοκότητα των υπολογισμών. Κάτι που θα βοηθούσε προκειμένου να μπορούσαν να χρησιμοποιηθούν κάποιες από αυτές τις μεθόδους, θα ήταν η αύξηση του χρονικού βήματος ολοκλήρωσης Δt , έτσι ώστε να μειωθεί ο αριθμός των αναγκαίων επαναλήψεων και επομένως και η πολυπλοκότητα του αλγορίθμου. Κάτι τέτοιο όμως δεν είναι δυνατό, διότι η τιμή του άνω ορίου του χρονικού βήματος ολοκλήρωσης είναι άνω φραγμένη, καθώς απαιτείται ακριβής υπολογισμός του δυναμικού στις μικρές αποστάσεις ανάμεσα στα άτομα, συγκεκριμένα στο απωθητικό του τμήμα. Μία μέθοδος που χρησιμοποιείται αρκετά συχνά λόγω αρκετών πλεονεκτημάτων είναι η μέθοδος Runge - Kutta τετάρτης τάξης. Στα πλεονεκτήματα της μεθόδου αυτής μπορεί κανείς να συμπεριλάβει το γεγονός ότι το τοπικό σφάλμα σε κάθε χρονική επανάληψη είναι της τάξης του h^5 , όπου h είναι το χρονικό βήμα επανάληψης, επομένως η μέθοδος προσφέρει αρκετά καλή ακρίβεια. Επιπλέον, γενικότερα η μέθοδος είναι σταθερή και εύκολη στον προγραμματισμό, όπως επίσης δεν απαιτεί γνώση πολλών παραμέτρων αρχικά. Δύο μειονεκτήματα που συνοδεύουν τη μέθοδο αυτή είναι ότι δεν παρέχει κάποια εκτίμηση της ακρίβειας που επιτυγχάνεται καθ' όλη τη διάρκεια του υπολογισμού και ότι 24N παράγωγοι πρέπει να υπολογίζονται σε κάθε χρονικό βήμα. [12]

2.1.5.1 Μέθοδος leapfrog

Η μέθοδος αυτή βασίζεται στο ανάπτυγμα Taylor της θέσης $x(t)$ συναρτήσει του χρόνου t , δηλαδή :

$$x(t + h) = x(t) + h\dot{x}(t) + \frac{h^2}{2}\ddot{x}(t) + O(h^3)$$

$$x(t - h) = x(t) - h\dot{x}(t) + \frac{h^2}{2}\ddot{x}(t) + O(h^3)$$

όπου $h=\Delta t$ είναι το χρονικό βήμα.

Στη μέθοδο αυτή, το ανάπτυγμα Taylor γράφεται :

$$x(t+h) = x(t) + h\dot{x}(t) + \frac{h}{2}\ddot{x}(t) + O(h^3) = x(t) + h\dot{x}\left(t + \frac{h}{2}\right)$$

$$x(t+h) = x(t) - h\dot{x}(t) + \frac{h}{2}\ddot{x}(t) + O(h^3) = x(t) - h\dot{x}\left(t - \frac{h}{2}\right)$$

Αφαιρώντας τις δύο τελευταίες εξισώσεις έχουμε :

$$x\left(t + \frac{h}{2}\right) = \dot{x}\left(t - \frac{h}{2}\right) + h\ddot{x}(t)$$

$$x(t+h) = x(t) + h\dot{x}\left(t + \frac{h}{2}\right)$$

που αποτελεί το αναδρομικό σχήμα της μεθόδου leapfrog για τον υπολογισμό των θέσεων και των ταχυτήτων των ατόμων.

Ο όρος leapfrog προέρχεται από το γεγονός ότι οι θέσεις και οι ταχύτητες υπολογίζονται για διαφορετικές χρονικές στιγμές. Για να αποφευχθεί η αποθήκευση των θέσεων και των ταχυτήτων σε άλλες χρονικές στιγμές οι υπολογισμοί χωρίζονται σε τρία μέρη :

Πριν υπολογιστούν οι επιταχύνσεις, υπολογίζονται οι ταχύτητες για μισό χρονικό βήμα, με τις παλιές τιμές των επιταχύνσεων, δηλαδή με τη σχέση :

$$\dot{x}\left(t + \frac{h}{2}\right) = \dot{x}(t) + \frac{h}{2}\ddot{x}(t)$$

Στη συνέχεια υπολογίζονται οι θέσεις για ένα ολόκληρο χρονικό βήμα, χρησιμοποιώντας τις παραπάνω τιμές της ταχύτητας :

$$x(t+h) = x(t) + h\dot{x}\left(t + \frac{h}{2}\right)$$

Τέλος, για να υπολογιστούν οι καινούριες τιμές των επιταχύνσεων, χρησιμοποιούνται οι καινούριες συντεταγμένες και στη συνέχεια υπολογίζονται οι ταχύτητες προσθέτοντας και το άλλο μισό χρονικό βήμα, ως εξής :

$$\dot{x}(t+h) = \dot{x}\left(t + \frac{h}{2}\right) + \frac{h}{2}\ddot{x}(t+h)$$

2.1.5.2 Μέθοδος Verlet

Ο αλγόριθμος Verlet βασίζεται πάλι στο ανάπτυγμα Taylor και είναι αλγεβρικά ισοδύναμος με τον αλγόριθμο leapfrog.

$$x(t+h) = x(t) + h\dot{x}(t) + \frac{h^2}{2}\ddot{x}(t) + O(h^3)$$

$$x(t-h) = x(t) - h\dot{x}(t) + \frac{h^2}{2}\ddot{x}(t) + O(h^3)$$

Με πρόσθεση των παραπάνω αναπτυγμάτων έχουμε :

$$x(t+h) = 2x(t) - x(t-h) + h^2\ddot{x}(t) + O(h^4)$$

όπου οι θέσεις $x(t-h)$ και $x(t)$ είναι γνωστές και άρα μπορούν να υπολογιστούν και οι επιταχύνσεις τους ως συνάρτηση των συντεταγμένων τους.

Ο υπολογισμός της ταχύτητας δεν εμπλέκεται στην επίλυση και για να τον υπολογισμό της χρησιμοποιείται η σχέση :

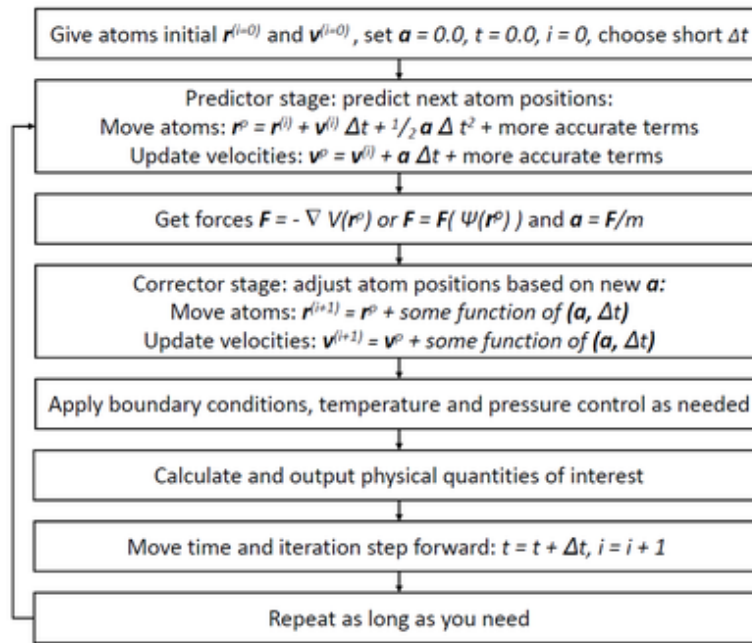
$$\dot{x}(t) = \frac{x(t+h) - x(t-h)}{2h} + O(h^2)$$

2.1.5.3 Μέθοδοι Prediction-Correction

Οι μέθοδοι πρόβλεψης - διόρθωσης (predictor - corrector (PC) methods) είναι μέθοδοι πολλαπλών τιμών, με την έννοια ότι χρησιμοποιούν πλήθος πληροφοριών που υπολογίζονται σε ένα ή περισσότερα προηγούμενα χρονικά βήματα. Ουσιαστικά, προηγούνται στην αριθμητική ολοκλήρωση, υπολογίζοντας αρχικά κάποια εκτίμηση του μεγέθους και βελτιώνοντας την τιμή αυτή, καθώς προχωράει ο υπολογισμός. Οι δύο πιο δημοφιλείς μορφές της μεθόδου διακρίνονται από το εάν χρησιμοποιούνται οι τιμές της επιτάχυνσης προηγούμενων χρονικών βημάτων (μέθοδος Adams πολλαπλών βημάτων), ή εάν χρησιμοποιούνται παράγωγοι της επιτάχυνσης της παρούσας χρονικής στιγμής (μέθοδος Nordsieck). Σε περιπτώσεις όπου η ακρίβεια είναι συγκεκριμένη ως προς δεδομένη ισχύ του βήματος ολοκλήρωσης Δt , οι δύο παραπάνω μορφές αποδεικνύεται ότι είναι αλγεβρικά ισοδύναμες.

Οι μέθοδοι είναι υψηλότερης τάξης από τη μέθοδο leapfrog, αλλά συνεπάγονται συγκεκριμένη ποσότητα από επιπλέον υπολογισμούς και απαιτούν μεγαλύτερη μνήμη για τις επιπλέον μεταβλητές που σχετίζονται με κάθε άτομο.

Simplified schematic of the molecular dynamics algorithm



Σχήμα 10: Διάγραμμα ροής μοριακής δυναμικής με τη μέθοδο prediction-correction

2.1.6 Ιστορική αναδρομή

Πάνε περίπου 60 χρόνια από τη δημοσίευση των πρώτων εργασιών από τους Alder και Wainright [7] που περιγράφουν τη χρήση της μεθόδου Molecular Dynamics (MD) και από τότε, μαζί με τις τεχνικές Monte Carlo, η MD έχει γίνει ένα βασικό εργαλείο για την θεωρητική μελέτη των ιδιοτήτων των υλικών σε σταθερές θερμοκρασίες. Αρχικά, η MD χρησιμοποιήθηκε σε συνδυασμό με απλά αλλά γενικά μοντέλα, όπως το μοντέλο σκληρών σφαιρών (hard spheres) ή τα μοντέλα υγρών Lennard-Jones, συστήματα τα οποία, μολονότι απλά, δεν ήταν υποκείμενα σε αναλυτική στατιστική αντιμετώπιση/χρήση. Σήμερα, όμως, η MD χρησιμοποιείται κυρίως σε συνδυασμό με αρκετά εξελιγμένα μοντέλα, καλύπτοντας το φάσμα από εμπειρικά δυναμικά πεδία (empirical force fields) μέχρι μεθόδους θεμελιωδών αρχών (first principles), με στόχο να περιγράψει με όσο το δυνατόν μεγαλύτερη ακρίβεια οποιοδήποτε υλικό χρειαστεί. Από υπολογιστική βοήθεια στη στατιστική μηχανική και τη φυσική πολλών σωμάτων (many-body physics), η MD έχει εξελιχθεί σε ένα ευρέως χρησιμοποιούμενο εργαλείο στη φυσικοχημεία (physical chemistry), στη φυσική της συμπυκνωμένης ύλης (condensed matter physics), στη βιολογία και στην επιστήμη των υλικών. [13] Παρακάτω θα μελετηθεί η φυσική σημασία της MD και θα παρατεθούν χρονολογικά οι πιο σημαντικές δημοσιεύσεις-σταθμοί στην ιστορία της με στόχο να γίνει κατανοητός ο τρόπος λειτουργίας της.

Στη μέθοδο MD, οι νευτώνειες εξισώσεις κίνησης ενός συνόλου N σωματιδίων σε όγκο V επιλύονται αριθμητικά. Η συνολική ενέργεια E του συστήματος διατηρείται σταθερή καθώς το σύστημα κινείται σε συγκεκριμένη τροχιά. Ο μέσος όρος κάθε ιδιότητας πάνω στην τροχιά είναι μια προσέγγιση της μετρούμενης τιμής αυτής της

ιδιότητας για τη θερμοδυναμική κατάσταση με τις εν λόγω τιμές N, V και E . Αυτός ο μέσος όρος είναι ισοδύναμος με ένα μέσο όρο σε μια μικροκανονική κατανομή (microcanonical ensemble/κατανομή NVE), αν η τροχιά περνά από όλα τα τμήματα φάσης-χώρου που έχουν τη δοσμένη ενέργεια. [...] Μια μετρήσιμη δομική ή θερμοδυναμική ιδιότητα F ενός συστήματος συσχετίζεται με μια συνάρτηση $F(rN, pN; V)$ της μηχανικής κατάστασης του συστήματος. Η συνήθης παραδοχή της στατιστικής θερμοδυναμικής είναι ότι η μετρήσιμη F είναι ίση με το μέσο όρο της κατανομής της συνάρτησης F πάνω σε μια κατάλληλα επιλεγμένη κατανομή καταστάσεων. Μελετώνται τέσσερις(4) κατανομές: η μικροκανονική(NVE), η κανονική(NVT), η ισοθερμοκρασιακή-ισοβαρής(NPT), και η ισοενθαλπική-ισοβαρής(NPH).

Σε μερικές περιπτώσεις είναι αναγκαίο να πραγματοποιηθούν προσομοιώσεις σε σταθερή πίεση ή/και θερμοκρασία.

Σε σταθερή πίεση, ο όγκος ενός συστήματος N σωματιδίων παρουσιάζει διακυμάνσεις. Για να περιγραφούν οι διακυμάνσεις αυτές, επινοήθηκε μια μέθοδος MD στην οποία ο όγκος είναι μια δυναμική μεταβλητή αντί για σταθερή ποσότητα. Το αποτέλεσμα είναι ένας τρόπος υπολογισμού τροχιών έτσι ώστε ο μέσος όρος της τροχιάς κάθε ιδιότητας να είναι ίσος με το μέσο όρο της NPH κατανομής της ιδιότητας αυτής. [14]

Σε σταθερή θερμοκρασία, η ενέργεια ενός συστήματος N σωματιδίων παρουσιάζει διακυμάνσεις. Για την προσομοίωση ενός τέτοιου συστήματος χρειάζεται ένας μηχανισμός εισαγωγής ενεργειακών διακυμάνσεων. Χρησιμοποιούνται στοχαστικές δυνάμεις που ενεργούν στα άτομα του δείγματος και αλλάζουν την κινητική τους ενέργεια. Το αποτέλεσμα είναι ένας τρόπος υπολογισμού τροχιών ώστε ο μέσος όρος της τροχιάς κάθε ιδιότητας να είναι ίσος με το μέσο όρο της NVT κατανομής της ιδιότητας αυτής. [14]

Σε σταθερή θερμοκρασία και πίεση, η ενέργεια, η πίεση(μήπως εννοεί όγκο??) και η ενθαλπία ενός συστήματος N σωματιδίων παρουσιάζουν διακυμάνσεις. Για να περιγραφεί ένα τέτοιο σύστημα χρειάζεται ένας τρόπος για την προσομοίωσή του. Αυτό μπορεί να γίνει εισάγοντας στοχαστικές συγκρούσεις στη μέθοδο MD σταθερής πίεσης. Το αποτέλεσμα είναι ένας τρόπος υπολογισμού τροχιών ώστε ο μέσος όρος της τροχιάς κάθε ιδιότητας να είναι ίσος με το μέσο όρο της NPT κατανομής της ιδιότητας αυτής. [14]

Η μέθοδος σταθερής πίεσης του Andersen γενικεύτηκε για τα στερεά από τους Parinello και Rahman [15] και η μέθοδος σταθερής θερμοκρασίας βελτιώθηκε από στοχαστική σε ντετερμινιστική από τον Nose [16] και έπειτα η χρονική κλίμακα εξαφανίστηκε από τον Hoover [17].

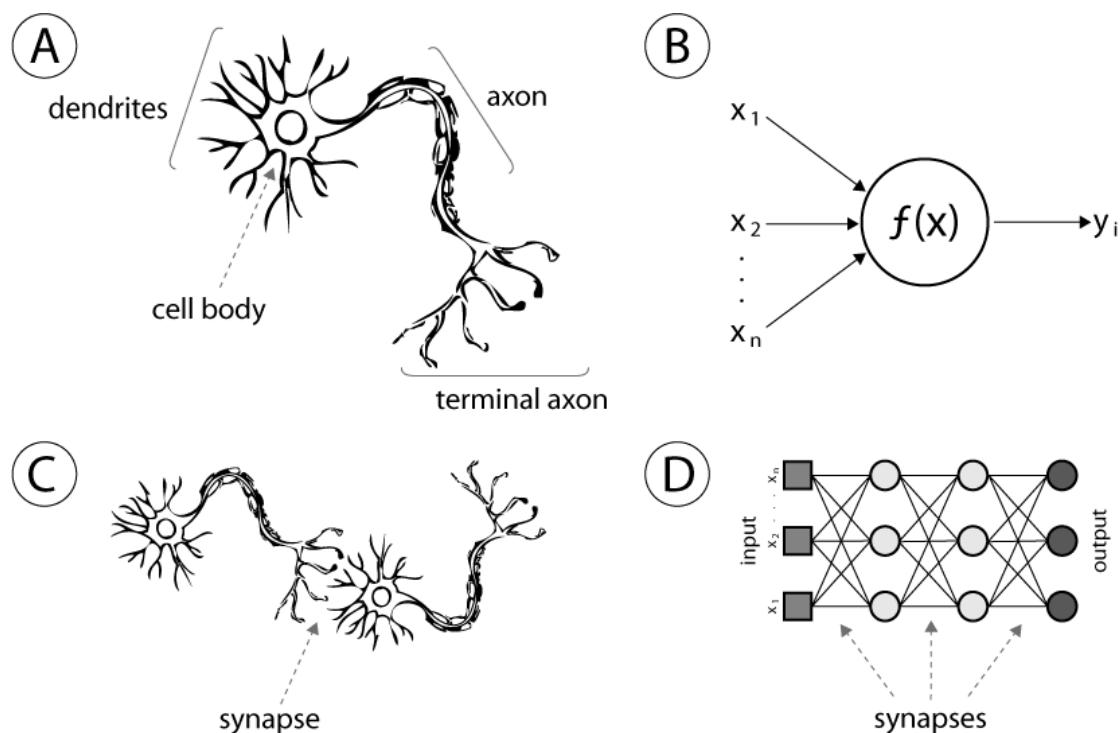
Το 1985 οι Car και Parinello [10] έδειξαν για πρώτη φορά ότι είναι δυνατό να πραγματοποιηθεί *ab initio* MD, δηλαδή MD στην οποία οι δυνάμεις στα άτομα δεν εξάγονται από εμπειρικά δυναμικά πεδία, αλλά από έναν ολοκληρωμένο υπολογισμό ηλεκτρονικής δομής με βάση θεμελιώδεις αρχές(first principles). Ο συνδυασμός των μεθόδων αυτών ονομάστηκε First Principles Molecular Dynamics (FPMD). Εν συντομία, η ιδέα των Car και Parinello ήταν να φέρουν ταυτόχρονα στο

προσκήνιο ηλεκτρόνια και ιόντα, αλλά με έναν ασυνήθιστο και εμπνευσμένο τρόπο. Όπως στη συνηθισμένη MD με δυναμικό πεδίο, τα ιόντα κινούνται σύμφωνα με τις δυνάμεις που τους ασκούνται, όμως οι δυνάμεις αυτές δεν προέρχονταν από εμπειρικό δυναμικό, αλλά από την αλληλεπίδραση μεταξύ τους με βάση τον Coulomb, και από την αλληλεπίδρασή τους με την πυκνότητα ηλεκτρονίου σθένους (valence electron density) γύρω τους. Το επίτευγμα των Car και Parinello κατάφερε να πείσει την επιστημονική κοινότητα ότι η FPMD είναι εφικτή, και σύντομα πολλές ομάδες ανα τον κόσμο άρχισαν να πραγματοποιούν FPMD προσομοιώσεις χρησιμοποιώντας είτε το σχήμα που πρότειναν οι Car/Parinello είτε εναλλακτικά σχήματα.

2.2 Τεχνητά νευρωνικά δίκτυα

2.2.1 Εισαγωγή

Η ονομασία τεχνητά νευρωνικά δίκτυα (artificial neural networks-ANN) για συγκεκριμένη ομάδα μαθηματικών συναρτήσεων οφείλεται στην ομοιότητα των γραφικών τους παραστάσεων με το δίκτυο νευρώνων του εγκεφάλου μας. Αυτή η αναλογία έχει οδηγήσει σε βιολογικά εμπνευσμένη ονοματολογία. Οι βιολογικοί νευρώνες είναι διασυνδεδεμένοι σε μια δομή δικτύου, όπου ο κάθε νευρώνας δέχεται ηλεκτρικά σήματα εισόδου και μεταδίδει μια απάντηση όταν η συνολική είσοδος υπερβαίνει ένα συγκεκριμένο όριο ενεργοποίησης. Κάθε ξεχωριστός νευρώνας μπορεί λοιπόν να θεωρηθεί ως ένας επεξεργαστής σημάτων με μια βηματική συνάρτηση Heaviside ως συνάρτηση ενεργοποίησης.



Σχήμα 11 : Πραγματικά και τεχνητά νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα ουσιαστικά προσπαθούν να μιμηθούν τα βιολογικά νευρικά δίκτυα αναπαριστώντας νευρώνες με βηματοειδείς συναρτήσεις

επεξεργασίας σημάτων. Για αριθμητικούς λόγους, η ασυνέχεια της βηματικής συνάρτησης αποφεύγεται στην πράξη, και σιγμοειδείς συναρτήσεις, όπως η υπερβολική εφαπτομένη ή η λογιστική συνάρτηση χρησιμοποιούνται ως συναρτήσεις ενεργοποίησης στα ANN.

Μερικές από τις πιο βασικές συναρτήσεις ενεργοποίησης είναι οι εξής :

Γραμμική συνάρτηση $f_a^1(x) = x$

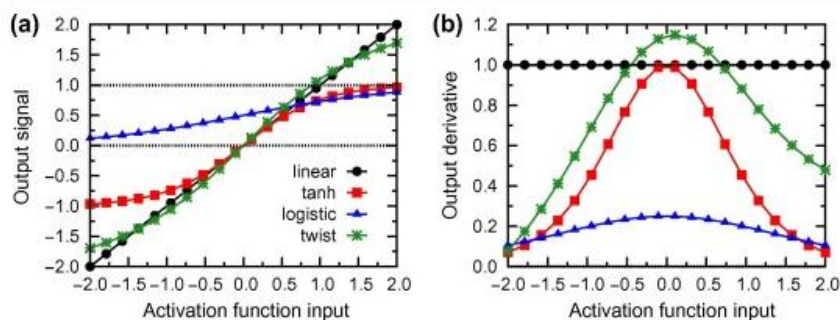
Υπερβολική εφαπτομένη $f_a^2(x) = \tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$

Λογιστική συνάρτηση $f_a^3(x) = \frac{1}{1+e^{-x}}$

Τanh με γραμμική περιστροφή $f_a^4(x) = 1.7159 \tanh\left(\frac{2}{3}x\right) + ax$

Λογαριθμική-σιγμοειδής $f_a^5(x) = \frac{1}{1+e^{-n}}$

Εφαπτομενική-σιγμοειδής $f_a^6(x) = \frac{2}{1+e^{-2n}} - 1$



Σχήμα 12: α) Γραφική παράσταση των συναρτήσεων ενεργοποίησης για τιμές από -2 έως 2
β) Παράγωγοι των συναρτήσεων ενεργοποίησης από -2 έως 2

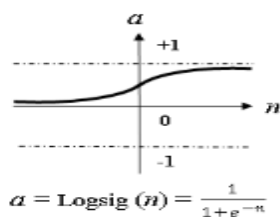


Fig. 2a: Log-Sigmoid

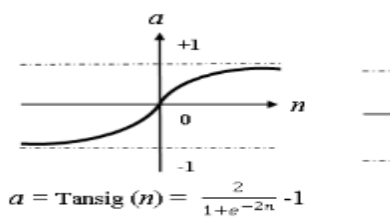


Fig. 2b: Tan-Sigmoid

Σχήμα 13 : Γραφική παράσταση της λογαριθμικής σιγμοειδούς και της εφαπτομενικής σιγμοειδούς

Τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν για τους παρακάτω σκοπούς[18]:

- Προσέγγιση συναρτήσεων ή ανάλυση παλινδρόμησης (ανάλυση χρονικής σειράς, βελτίωση χρόνου βελτιστοποίησης, κ.α.)
- Κατηγοριοποίηση μοτίβου(αναγνώριση μοτίβου ή ακολουθίας, κ.α.)
- Επεξεργασία δεδομένων(φιλτράρισμα, ομαδοποίηση, διαλογή, συμπίεση, κ.α.)
- Πρόβλεψη/πρόγνωση
- Βελτιστοποίηση
- Απομνημόνευση
- Έλεγχος(cnc κ.α.)

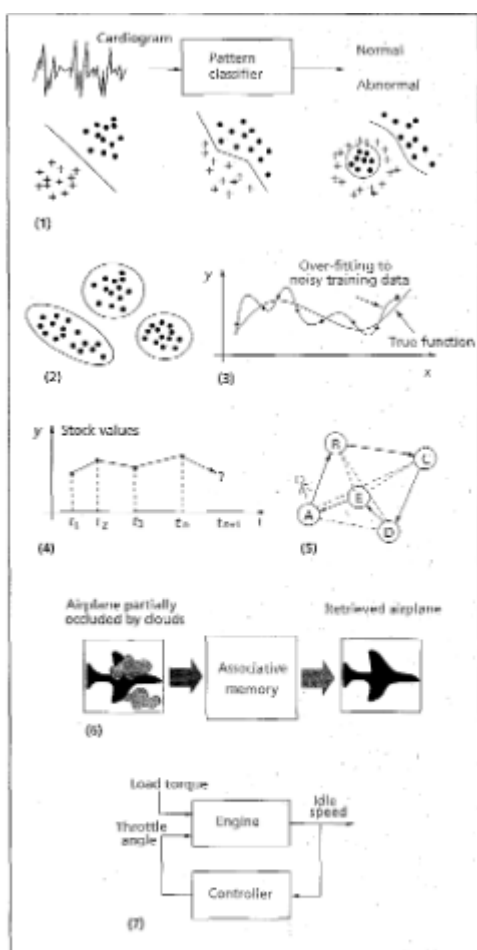
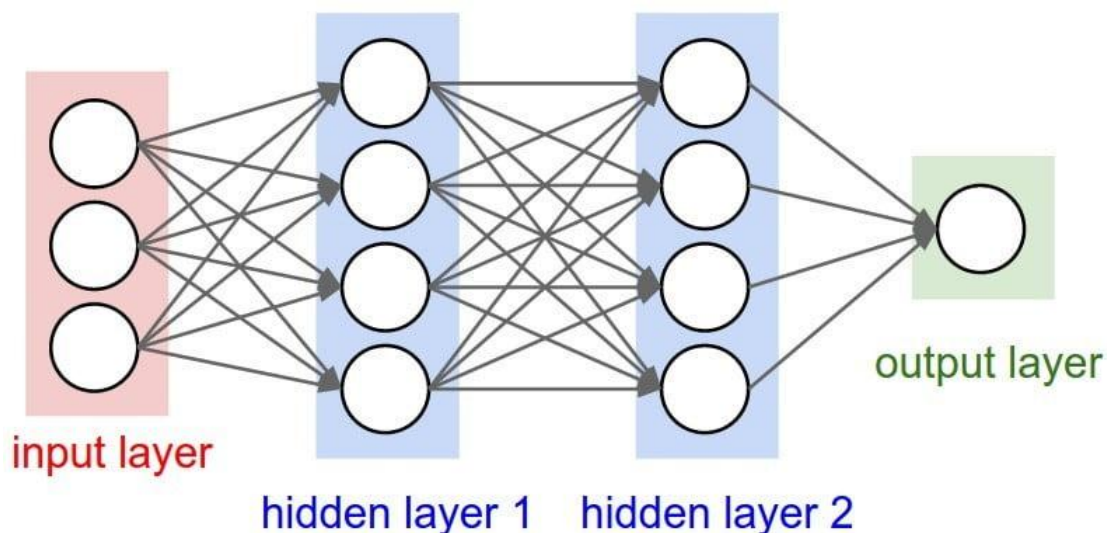


Figure A. Tasks that neural networks can perform: (1) pattern classification; (2) clustering/categorization; (3) function approximation; (4) prediction/forecasting; (5) optimization (a TSP problem example); (6) retrieval by content; and (7) control (engine idle speed). (Adapted from DARPA Neural Network Study')

Σχήμα 14 : Εφαρμογές τεχνητών νευρωνικών δικτύων

Λόγω της ικανότητας των ANN να μοντελοποιούν και να αναπαράγουν μη-γραμμικές διεργασίες, έχουν βρει εφαρμογές σε διάφορους κλάδους :

- Αναγνώριση συστήματος και έλεγχος (έλεγχος οχημάτων, πρόβλεψη τροχιάς [19], έλεγχος διεργασιών, διαχείριση φυσικών πόρων κ.α.)
- Παίγνια και λήψη αποφάσεων(τάβλι, σκάκι, πόκερ κ.α.)
- Αναγνώριση μοτίβου(συστήματα ραντάρ, εξακρίβωση προσώπου, κατηγοριοποίηση σημάτων [20], αναγνώριση αντικειμένων κ.α.)
- Αναγνώριση ακολουθίας(χειρονομιών, ομιλίας, αναγνώριση κειμένων χειρόγραφων ή τυπωμένων κ.α.)
- Ιατρική διάγνωση(καρκίνου του πνεύμονα [21],καρκίνου του προστάτη,διάκριση γραμμών επιθετικών κελιών καρκίνου από λιγότερο επιθετικά [22] [23] κ.α.)
- Χρηματοοικονομικά(αυτοματοποιημένα συστήματα συναλλαγών κ.α.)
- Εξόρυξη δεδομένων
- Ανάλυση αξιοπιστίας([24] [25])
- Γεωεπιστήμη(υδρολογία [26], παράκτια μηχανική [27] [28], γεωμορφολογία [29], μοντελοποίηση του ωκεανού κ.α.)



Σχήμα 15 : Ένα τυπικό τεχνητό νευρωνικό δίκτυο με 3 εισόδους, 2 κρυφές στρώσεις με 4 νευρώνες η κάθε μία και 1 έξοδο(3-4-4-1)

Για τη χρησιμοποίηση των νευρωνικών δικτύων χρειάζεται η επιλογή ορισμένων παραμέτρων:

- Αριθμός κρυφών στρώσεων μεταξύ των στρώσεων εισόδου και εξόδου
- Αριθμός νευρώνων σε κάθε κρυφή στρώση
- Συνάρτηση ενεργοποίησης για κάθε νευρώνα μιάς στρώσης(συνήθως σιγμοειδής)
- Αλγόριθμος εκμάθησης(π.χ. conjugate gradient, Levenberg-Marquardt)
- Ποσοστά του dataset για training/testing/validation
- Συνάρτηση κόστους ή αντικειμενική συνάρτηση

Ο αριθμός κρυφών στρώσεων είναι σημαντικός γιατί, αν χρησιμοποιηθούν μη-γραμμικές συναρτήσεις ενεργοποίησης επιτρέπουν στο δίκτυο να μάθει μη-γραμμικές σχέσεις μεταξύ διανυσμάτων εξόδου και εισόδου.

Η αύξηση του αριθμού των νευρώνων σε κάθε κρυφή στρώση οδηγεί σε μείωση του σφάλματος στην περίπτωση της ανάλυσης παλινδρόμησης, αλλά και σε αύξηση του υπολογιστικού χρόνου/κόστους.

Η συνάρτηση ενεργοποίησης καθορίζει τη μετάβαση από τη μιά στρώση στην άλλη και δημιουργεί την απαιτούμενη πολυπλοκότητα μεταξύ των εισόδων και των εξόδων του δικτύου. Στην περίπτωση της ανάλυσης παλινδρόμησης επιλέγεται συνήθως η γραμμική συνάρτηση για τη στρώση εξόδου.

Ο αλγόριθμος εκμάθησης καθορίζει την αναλογία ακρίβειας/υπολογιστικού κόστους, και ανάλογα με την περίπτωση επιλέγεται ο καταλληλότερος.

Το αρχικό dataset χωρίζεται σε τρία κομμάτια. Το κομμάτι του training χρησιμοποιείται για την εκμάθηση του νευρωνικού δικτύου. Χρησιμοποιείται για τον υπολογισμό της κλίσης και για την ενημέρωση των βαρών και του συστηματικού σφάλματος (bias). Το κομμάτι του validation χρησιμοποιείται για την αποφυγή του overfitting, δηλαδή της υπεραπόδοσης του δικτύου. Τυπικά και τα δύο σφάλματα, του training και του validation μικραίνουν κατά το αρχικό στάδιο της εκμάθησης. Όταν το δίκτυο ξεκινά να υπεραποδίδει, το σφάλμα του validation κομματιού αρχίζει να μεγαλώνει, παρόλο που το σφάλμα του training μικραίνει. Τα βάρη και τα συστηματικά σφάλματα του δικτύου αποθηκεύονται στο ελάχιστο σφάλμα του validation ώστε το δίκτυο να είναι αποδοτικότερο. Το τελευταίο κομμάτι, του testing, δεν χρησιμοποιείται κατά την εκμάθηση, αλλά χρησιμοποιείται για να συγκρίνει διαφορετικά μοντέλα.

Η αντικειμενική συνάρτηση πρέπει να βελτιστοποιηθεί από το νευρωνικό δίκτυο. Καθώς στην ανάλυση παλινδρόμησης είναι επιθυμητό να μειωθεί το σφάλμα μεταξύ των στόχων και των εξόδων του δικτύου, υπάρχει ένα πρόβλημα ελαχιστοποίησης σφάλματος.

Βάση της αρχιτεκτονικής τους, τα ANN κατηγοριοποιούνται ως εξής :

- Feed-forward δίκτυα, στα οποία δεν υπάρχουν κυκλικές συνδέσεις μεταξύ νευρώνων
- Feedback (ή recurrent) δίκτυα, στα οποία υπάρχουν κυκλικές συνδέσεις

Υπάρχουν τέσσερα είδη ή παραδείγματα [30] μάθησης :

- Επιβλεπόμενη μάθηση (supervised learning)

Στην επιβλεπόμενη μάθηση παρέχεται στο δίκτυο μια σωστή απάντηση (έξοδος) για κάθε μοτίβο εισόδου. Τα βάρη καθορίζονται με τέτοιο τρόπο ώστε να επιτρέψουν στο δίκτυο να παράξει απαντήσεις όσο το δυνατόν πιο κοντά στις γνωστές σωστές απαντήσεις.

- Ενισχυτική μάθηση (reinforcement learning)

Η ενισχυτική μάθηση είναι μια παραλλαγή της επιβλεπόμενης μάθησης στην οποία στο δίκτυο παρέχεται μόνο μια αξιολόγηση για την ορθότητα των εξόδων του δικτύου, και όχι οι σωστές εξοδοι αυτούσιες.

- Μη-επιβλεπόμενη μάθηση (unsupervised learning)

Στην μη-επιβλεπόμενη μάθηση δεν απαιτείται μιά σωστή έξοδος για κάθε είσοδο του δικτύου. Αντίθετα, εξερευνάται η υποβόσκουσα δομή των δεδομένων ή ο συσχετισμός μεταξύ μοτίβων των δεδομένων, και οργανώνονται τα μοτίβα σε κατηγορίες μέσω αυτών των συσχετισμών.

- Υβριδική μάθηση (hybrid learning)

Συνδυασμός επιβλεπόμενης και μη-επιβλεπόμενης μάθησης. Ένα μέρος των βαρών καθορίζεται από επιβλεπόμενη μάθηση, ενώ ένα άλλο μέρος από μη-επιβλεπόμενη.

Υπάρχουν τρία βασικά και πρακτικά προβλήματα που πρέπει να αντιμετωπιστούν κατά τη διάρκεια της μάθησης :

- Χωρητικότητα (capacity)

Η χωρητικότητα αφορά το πόσα μοτίβα μπορούν να αποθηκευτούν και ποιές συναρτήσεις και όρια αποφάσεων μπορεί να σχηματίσει ένα δίκτυο.

- Πολυπλοκότητα του δείγματος (sample complexity)

Η πολυπλοκότητα του δείγματος καθορίζει τον αριθμό των μοτίβων μάθησης που χρειάζονται έτσι ώστε το δίκτυο να μπορεί να κάνει μια έγκυρη γενίκευση. Πολύ λίγα μοτίβα μπορούν να δημιουργήσουν over-fitting, δηλαδή το δίκτυο να αποδίδει ικανοποιητικά στο training dataset αλλά μη ικανοποιητικά στο testing dataset, δηλαδή σε ξεχωριστά μοτίβα που προέρχονται από την ίδια κατανομή με το training dataset

- Υπολογιστική πολυπλοκότητα (computational complexity)

Ο αναγκαίος χρόνος ενός αλγορίθμου για τον υπολογισμό μιάς λύσης μέσω των μοτίβων εκμάθησης.

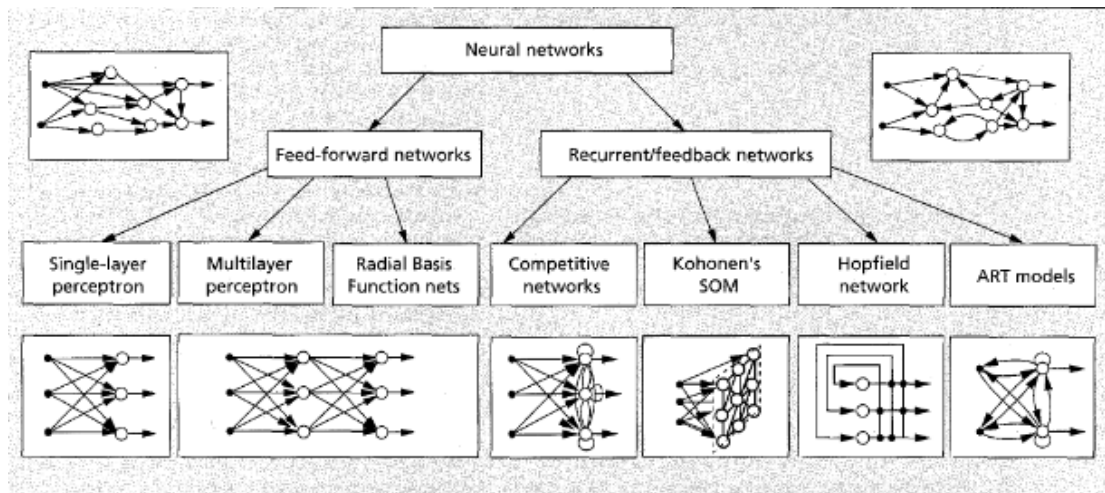


Figure 4. A taxonomy of feed-forward and recurrent/feedback network architectures.

Σχήμα 16 : Κατηγοριοποίηση των νευρωνικών δικτύων και σχηματική αναπαράσταση

Paradigm	Learning rule	Architecture	Learning algorithm	Task
Supervised	Error-correction	Single- or multilayer perceptron	Perceptron learning algorithms	Pattern classification
			Back-propagation Adaline and Madaline	Function approximation Prediction, control
	Boltzmann	Recurrent	Boltzmann learning algorithm	Pattern classification
	Competitive	Competitive	Linear discriminant analysis	Data analysis Pattern classification
Learning vector quantization			Within-class categorization Data compression	
Unsupervised	Error-correction	Multilayer feed-forward	ART network	Pattern classification Within-class categorization
			ARTMap	Within-class categorization
	Hebbian	Feed-forward or competitive	Sammon's projection	Data analysis
			Principal component analysis	Data analysis Data compression
	Competitive	Competitive	Hopfield Network	Associative memory learning
			Vector quantization	Associative memory Categorization Data compression
Hybrid	Error-correction and competitive	Kohonen's SOM	Kohonen's SOM	Categorization Data analysis
		ART networks	ART1, ART2	Categorization
		RBF network	RBF learning algorithm	Pattern classification Function approximation Prediction, control

Σχήμα 17 : Αναλυτική κατηγοριοποίηση των νευρωνικών δικτύων με βάση τη μέθοδο μάθησης, την αρχιτεκτονική, τον αλγόριθμο μάθησης και την εφαρμογή τους.

Στα multilayer perceptrons (MLPs), οι τεχνητοί νευρώνες οργανώνονται σε στρώσεις, και μόνο οι νευρώνες σε γειτονικές στρώσεις είναι συνδεδεμένοι. Η αλγεβρική έκφραση του j νευρώνα στην i στρώση ενός MLP με N στρώσεις είναι :

$$x_{i,j}(\{x_{i-1,k}\}) = f_a^{i,j} \left(\sum_k w_{k,j}^i x_{i-1,k} \right), \quad \mu\epsilon \ i = 1, \dots, (N - 1)$$

Όπου $\{x_{i-1,k}\}$ είναι τα σήματα εισόδου από την προηγούμενη στρώση ($i-1$), $w_{k,j}^i$ είναι το «βάρος» του σήματος από τον k νευρώνα της ($i-1$) στρώσης και $f_a^{i,j}$ είναι η συνάρτηση ενεργοποίησης. Ισοδύναμα, το διάνυσμα όλων των νευρώνων στη στρώση i δίνεται από τη σχέση :

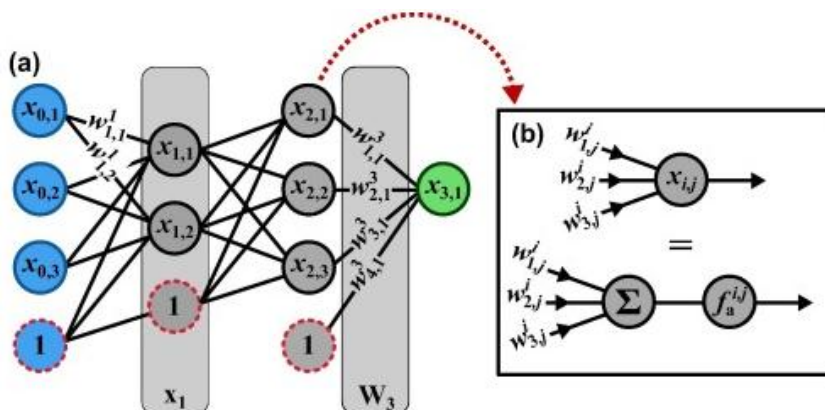
$$x_i(x_{i-1}) = f_a^i(W_i x_{i-1}), \quad \mu\epsilon \ i = 1, \dots, (N - 1)$$

Το W_i είναι ο πίνακας «βάρους» για τη μετάδοση σήματος από την i στην $i-1$ στρώση με $(W_i)_{k,j} = w_{k,j}^i$.

Στη γραφική παράσταση των MLP, οι νευρώνες είναι οι κόμβοι, και $w_{k,j}^i$ είναι τα βάρη των συνδέσεων, δηλαδή οι άκρες του γραφήματος. Ένα παράδειγμα ενός MLP και η γραφική παράσταση ενός νευρώνα φαίνονται παρακάτω. Η αλγεβρική έκφραση (συνάρτηση δικτύου) του MLP είναι :

$$N(x_0; \{W_i\}) = f_a^3 \{W_3 f_a^2 [W_2 f_a^1 (W_1 x_0)]\} = x_3$$

Το διάνυσμα x_0 είναι η είσοδος (στρώση εισόδου, μπλέ κόμβοι στο σχήμα) της συνάρτησης δικτύου N , και x_3 είναι η τιμή της συνάρτησης (στρώση εξόδου, πράσινοι κόμβοι στο σχήμα). Οι στρώσεις μεταξύ των στρώσεων εισόδου και εξόδου (γκρι κόμβοι στο σχήμα) ονομάζονται κρυφές στρώσεις, καθώς δεν έχουν εύληπτη ερμηνεία. Τα στοιχεία των πινάκων βάρους W_i , οι παράμετροι βάρους, είναι οι παράμετροι του μοντέλου που πρέπει να καθοριστούν κατά τη διάρκεια της διαδικασίας εκμάθησης. Ο αριθμός των στρώσεων σε ένα MLP και ο αριθμός των νευρώνων ανά στρώση ορίζουν την αρχιτεκτονική του δικτύου. Το MLP του παραδείγματος έχει αρχιτεκτονική 3-2-3-1.



Σχήμα 18 : a) Ένα τεχνητό νευρωνικό δίκτυο. Μπλε - είσοδοι, πράσινο- έξοδος, κόκκινη διακεκομμένη - bias neurons. b) Ένας τεχνητός νευρώνας που προσθέτει τα σήματα εισόδου και μεταδίδει το προϊόν μιας συνάρτησης ενεργοποίησης

Κάθε στρώση του MLP πριν την στρώση εξόδου περιλαμβάνει έναν κόμβο που μεταδίδει ένα σταθερό σήμα ίσο με 1. Αυτοί οι σταθεροί νευρώνες (bias neurons-με κόκκινη διακεκομμένη διαγράμμιση στο σχήμα) βοηθούν μια αλλαγή σταθερών με τη βαρύτητα των παραμέτρων βάρους τους.

Οι παράμετροι βάρους για όλες τις στρώσεις του δικτύου καθορίζονται κατά τη διαδικασία εκπαίδευσης. Η ανάλυση παλινδρόμησης είναι ένα πρόβλημα supervised learning, δηλαδή οι παράμετροι βάρους πρέπει να ρυθμιστούν έτσι ώστε ένα σύνολο από δείγματα αναφοράς εισόδου-εξόδου $\{(\mathbf{x}_{0,n}, \mathbf{y}_n)\}$ να αναπαράγεται από τη συνάρτηση δικτύου N με όσο το δυνατόν μεγαλύτερη ακρίβεια. Μαθηματικά αυτό σημαίνει ότι η απόλυτη τιμή του σφάλματος e_n κάθε δείγματος αναφοράς

$$e_n(\mathbf{x}_{0,n}, \mathbf{y}_n; \{\mathbf{W}_l\}) = N(\mathbf{x}_{0,n}; \{\mathbf{W}_l\}) - \mathbf{y}_n$$

πρέπει να ελαχιστοποιηθεί. Επομένως, η εκπαίδευση του ANN μπορεί να εκφραστεί ως πρόβλημα βελτιστοποίησης πολλών διαστάσεων :

$$\{\mathbf{W}_l^{opt}\} = \underset{\{\mathbf{W}_l\}}{\operatorname{arg\,min}} \varepsilon(\{\mathbf{W}_l\}), \text{ με } \varepsilon(\{\mathbf{W}_l\}) = \frac{1}{2} \sum_n^{samples} e_n^2$$

όπου ε είναι η τετραγωνική συνάρτηση σφάλματος του ANN και $\{\mathbf{W}_l^{opt}\}$ είναι το σύνολο των βέλτιστων παραμέτρων βάρους.

Για να εκτιμηθεί η συνολική συνάρτηση σφάλματος της παραπάνω εξίσωσης, το σφάλμα e_n κάθε δείγματος n στο σύνολο αναφοράς πρέπει να υπολογιστεί. Οι μέθοδοι βελτιστοποίησης που χρειάζονται ολόκληρη τη συνάρτηση σφάλματος λέγονται μέθοδοι εκπαίδευσης ομάδας/παρτίδας (batch training). Εναλλακτικά, κάθε δείγμα αναφοράς μπορεί να παρουσιάζεται ξεχωριστά στο ANN, και μια διόρθωση βάρους μπορεί να υπολογιστεί βασιζόμενη στο σφάλμα e_n κάθε ξεχωριστού δείγματος αναφοράς (online training).

Παρακάτω δίδονται τρεις διαφορετικές μέθοδοι εκπαίδευσης που μπορούν να χρησιμοποιηθούν για τη βελτιστοποίηση των βαρών ενός ANN.

1) Gradient descent(GD). Ο απλούτερος αλγόριθμος βελτιστοποίησης που υπάρχει είναι ο αλγόριθμος gradient descent, που προσπαθεί την άμεση επαναληπτική ελαχιστοποίηση της κλίσης της συνάρτησης σφάλματος. Σε κάθε επανάληψη l της μεθόδου GD, οι παράμετροι βάρους επικαιροποιούνται/ανανεώνονται σύμφωνα με τη σχέση :

$$\mathbf{w}^{(l+1)} = \mathbf{w}^l + \Delta \mathbf{w}^{GD,(l+1)}, \text{ με } \Delta \mathbf{w}^{GD,(l+1)} = -\gamma \nabla \varepsilon,$$

όπου \mathbf{w}^l είναι το διάνυσμα όλων των παραμέτρων βάρους, γ ένας εμπειρικός ρυθμός εκμάθησης, και η παράγωγος/κλίση της συνάρτησης σφάλματος, $\nabla \varepsilon$, δίνεται σε προηγούμενη εξίσωση. Η μέθοδος αυτή χρησιμοποιείται κυρίως για online training.

2) Limited memory BFGS (L-BFGS).

Οι Σχεδόν-Νευτώνειες (Quasi-Newton) μέθοδοι βασίζονται σε δεύτερης τάξης επέκταση της αντικειμενικής συνάρτησης και οι επαναληπτικές ανανεώσεις των βαρών παίρνουν τη μορφή:

$$\Delta \mathbf{w}^{QN,(I+1)} = -(\mathbf{H}^{(I)})^{-1} \nabla \varepsilon^{(I)},$$

όπου \mathbf{H} είναι μια προσέγγιση του Εσσιανού πίνακα, δηλαδή οι δεύτερες παράγωγοι της συνάρτησης δικτύου ως προς τα βάρη. Μια δημοφιλή σχεδον-Νευτώνεια μέθοδος είναι η Broyden-Fletcher-Goldfarb-Shanno (BFGS) μέθοδος στην οποία το \mathbf{H} ανανεώνεται σε κάθε επανάληψη[4 refs]. Για την εκπαίδευση των ANN με μεγάλο αριθμό παραμέτρων βάρους, η L-BFGS είναι ιδιαίτερα χρήσιμη, καθώς δεν απαιτεί την αποθήκευση ολόκληρου του Εσσιανού πίνακα.

3) Levenberg - Marquardt (LM). Η μέθοδος Levenberg-Marquardt είναι μια συμβατική μέθοδος για το ταίριαγμα (fitting) ελαχίστων τετραγώνων που χρησιμοποιείται συχνά για την εκπαίδευση των ANN. Το διάνυσμα ανανέωσης των LM βαρών δίνεται από τη σχέση :

$$\Delta \mathbf{w}^{LM,(I+1)} = -(\mathbf{J}^{T,(I)} \mathbf{J}^{(I)} + \lambda \mathbf{I})^{-1} \mathbf{J}^{T,(I)} \mathbf{e}^{(I)},$$

όπου \mathbf{J} είναι ο Ιακωβιανός (Jacobian) πίνακας της εξίσωσης με τη συνάρτηση σφάλματος, \mathbf{I} είναι ο ταυτοτικός πίνακας, και λ είναι ένας ρυθμός εκμάθησης. Υπάρχει ομοιότητα της ανανέωσης βαρών της μεθόδου LM με την αντίστοιχη έκφραση της Quasi-Newton καθώς $\mathbf{J}^T \mathbf{J}$ είναι η προσέγγιση Gauss-Newton για τον Εσσιανό πίνακα. Για $\lambda = 0$ η μέθοδος είναι όμοια με τον αλγόριθμο Gauss-Newton, αλλά για μεγάλα λ πλησιάζει τον αλγόριθμο Gradient Descent με μικρό ρυθμό εκμάθησης. Η ιδέα πίσω από τη μέθοδο LM είναι η εκκίνηση με την πιο ισχυρή μέθοδο gradient descent, αλλά, όταν τα βάρη πλησιάσουν αρκετά στο βέλτιστο, να γίνει αλλαγή προς την γρηγορότερα συγκλίνουσα μέθοδο Gauss-Newton. Για να επιτευχθεί αυτό, η τιμή του λ μειώνεται μετά από κάθε επανάληψη που κατέληξε σε μείωση του σφάλματος, και αυξάνεται όταν η ανανέωση βαρών θα ανέβαζε το σφάλμα. [31]

2.2.2 Ιστορική αναδρομή

Το 1943 ο νευροφυσιολόγος Warren McCulloch και ο μαθηματικός Walter Pitts δημιούργησαν ένα υπολογιστικό μοντέλο[32] για τα νευρωνικά δίκτυα βασισμένο στη μαθηματική λογική και τους αλγόριθμους που ονομάστηκε λογική κατωφλίου. Μοντελοποίησαν ένα απλό νευρωνικό δίκτυο με ηλεκτρικά κυκλώματα. Το μοντέλο αυτό άνοιξε το δρόμο για τη διχοτόμηση της έρευνας πάνω στα νευρωνικά δίκτυα σε βιολογικές διεργασίες στον εγκέφαλο και σε εφαρμογές των νευρωνικών δικτύων στην τεχνητή νοημοσύνη. Η δουλειά αυτή οδήγησε σε εξελίξεις στα νευρικά δίκτυα και κατα προέκταση στα πεπερασμένα αυτόματα (finite automata) [33]. Το 1949 ο Donald Hebb[34] δημιούργησε ένα μοντέλο εκμάθησης βασισμένο στη νευρική πλαστικότητα, δηλαδή την ικανότητα του εγκεφάλου να αλλάζει κατά τη

διάρκεια της ζωής ενός ατόμου, που ονομάστηκε εκμάθηση του Hebb(Hebbian learning) και είναι μη-επιβλεπόμενη εκμάθηση. Αυτό εξελίχθηκε σε μοντέλα για τη μακροχρόνια ενδυνάμωση(long term potentiation) των συνάψεων, δηλαδή μοτίβα που παράγουν μια μακροχρόνια αύξηση στη μετάδοση σήματος μεταξύ δύο νευρώνων. Ερευνητές εφάρμοσαν αυτές τις ιδέες σε υπολογιστικά μοντέλα με τις μηχανές Turing B-τύπου [35]. Οι Farley και Clark πρωτοχρησιμοποίησαν υπολογιστικές μηχανές για να προσομοιώσουν ένα δίκτυο του Hebb[36] και οι Rochester, Holland, Habit και Duda δημιούργησαν υπολογιστικές μηχανές για νευρωνικά δίκτυα [37]. Το 1958 ο Rosenblatt δημιούργησε το perceptron [38], έναν αλγόριθμο για την αναγνώριση προτύπων (pattern recognition). Το 1959 οι Hubel και Wiesel πρότειναν ένα βιολογικό μοντέλο που βασιζόταν στην ανακάλυψή τους δύο τύπων κελιών στον πρωτεύοντα οπτικό φλοιό, απλών και σύνθετων κελιών [39].Επίσης το 1959 οι Widrow και Hoff ανέπτυξαν μοντέλα που ονόμασαν ADALINE και MADALINE(Multiple Adaptive Linear Elements)και το MADALINE ήταν το πρώτο νευρωνικό δίκτυο που εφαρμόστηκε σε ένα αληθινό πρόβλημα, καθώς είναι ένα προσαρμοζόμενο φίλτρο που εξαλείφει την ηχώ στις τηλεφωνικές γραμμές. Τα πρώτα λειτουργικά δίκτυα με πολλές στρώσεις δημοσιεύτηκαν το 1965 από τους Ivakhnenko και Lapa [40].

Η έρευνα πάνω στα νευρωνικά δίκτυα έπεσε σε τέλμα μετά από την έρευνα των Minsky και Papert το 1969 [41], που ανακάλυψαν δύο σημαντικά προβλήματα με τις υπολογιστικές μηχανές που επεξεργάζονταν νευρωνικά δίκτυα. Πρώτον, ότι τα βασικά perceptron δεν μπορούσαν να επεξεργαστούν το κύκλωμα αποκλειστικής διάζευξης(exclusive-or), και δεύτερον ότι οι υπολογιστές δεν είχαν αρκετή επεξεργαστική ικανότητα για να φέρουν εις πέρας ικανοποιητικά την απαιτούμενη δουλειά που απαιτούσαν τα μεγάλα νευρωνικά δίκτυα.

Το 1975 ο Werbos δημιούργησε έναν αλγόριθμο backpropagation [42]που μετέφερε τον όρο του σφάλματος προς τα πίσω, στις στρώσεις με τους νευρώνες, έτσι ώστε να ανανεωθούν τα βάρη σε κάθε κόμβο και να μειωθεί περαιτέρω το σφάλμα, λύνοντας ουσιαστικά το πρόβλημα της αποκλειστικής διάζευξης και κάνοντας την εκμάθηση δικτύων με πολλές στρώσεις εφικτή και αποδοτική.

Το 1988 οι Qian και Sejnowski δημοσίευσαν μια μέθοδο για την πρόβλεψη της δευτερεύουσας δομής πρωτεϊνών [43] και το 1993 οι Rost και Sander έκαναν το ίδιο [44], με τον κλάδο αυτό να αλλάζει μέσω των νευρωνικών δικτύων.

Το 1992 ο Schmidhuber χρησιμοποίησε μια πολυεπίπεδη ιεραρχία δικτύων με κάθε δίκτυο να εκπαιδεύεται ξεχωριστά από πριν με μη-επιβλεπόμενη μάθηση και backpropagation [45]για να λύσει το πρόβλημα εξαφανιζόμενης κλίσης [46], που διατύπωσε ο Hochreiter το 1991 και αφορά τη μείωση του σφάλματος κατά τη μεταφορά του από τη μία στρώση νευρώνων στην άλλη, που εμποδίζει τη βελτίωση των βαρών στους κόμβους. Για την αντιμετώπιση του ίδιου προβλήματος, ο Behnke το 2003 [47] βασίστηκε στο πρόσημο της κλίσης με τον αλγόριθμο resilient propagation [48] που αναπτύχθηκε το 1992.

Το 2006 ο Hinton et al. [49]πρότεινε μια αναπαράσταση μεγάλου επιπέδου χρησιμοποιώντας διαδοχικές στρώσεις νευρώνων απο δυαδικές ή πραγματικές

λανθάνουσες μεταβλητές [50] με μια περιορισμένη μηχανή Boltzmann (restricted Boltzmann machine) [51] για τη μοντελοποίηση κάθε στρώσης.

Το 2012 οι Le, Ng και Dean [52] δημιούργησαν ένα δίκτυο που κατάφερε να αναγνωρίζει έννοιες μεγαλύτερου επιπέδου, όπως γάτες, χρησιμοποιώντας μη κατηγοριοποιημένες εικόνες από βίντεο στο Youtube.

2.3 Υλικά

2.3.1 Σεμεντίτης - Fe₃C

Τα άτομα σεμεντίτη που χρησιμοποιήθηκαν στην παρούσα διπλωματική [53] ανήκουν στο space group Pnma, στο Strukturbericht σύμβολο D0₁₁, και στο Pearson σύμβολο oP16. Τα σύμβολα Strukturbericht είναι μια μερικώς συστηματική μέθοδος για τον καθορισμό της δομής ενός κρυστάλλου. Επομένως οι A δομές είναι μονοατομικές, οι B δομές διατομικές με ίσους αριθμούς ατόμων κάθε τύπου, οι C διατομικές με 2-1 αναλογία κλπ. Το σύμβολο Pearson καθορίζει την κρυσταλλική συμμετρία και τον αριθμό των ατόμων στην κυψελίδα [54]. Για παράδειγμα, το NaCl έχει δομή face - centered (F) cubic (c) με 8 άτομα στην κυψελίδα, άρα το σύμβολο Pearson που έχει είναι cF8. Τα space groups είναι συνολικά 230 και υπάρχουν αναλυτικά στους Διεθνείς Πίνακες για την Κρυσταλλογραφία [55].

Ο σεμεντίτης, με βάση το space group του Pnma, ανήκει στο απλό ορθορομβικό σύστημα. Στο ορθορομβικό σύστημα, η συμβατική κυψελίδα είναι ένα παραλληλεπίπεδο που ορίζεται από τρία αμοιβαία ορθογώνια διανύσματα με άνισο μήκος :

$$A_1 = a\hat{x}$$

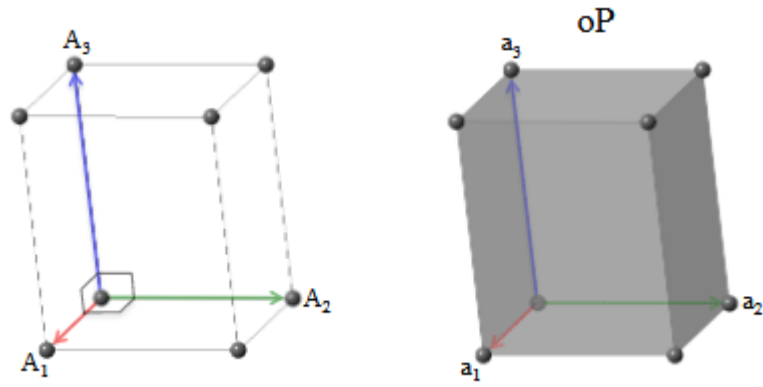
$$A_2 = b\hat{y}$$

$$A_3 = c\hat{z}$$

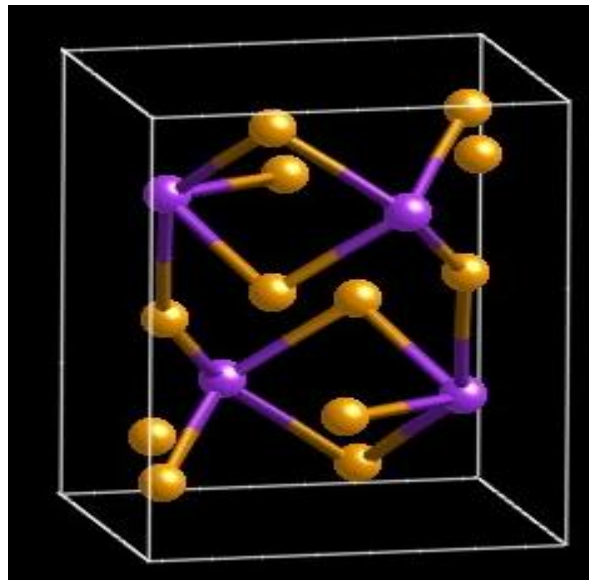
Έτσι ώστε $a \neq b \neq c$ αλλά $\alpha = \beta = \gamma = \pi/2$. Ο όγκος της συμβατικής κυψελίδας είναι

$$V = abc$$

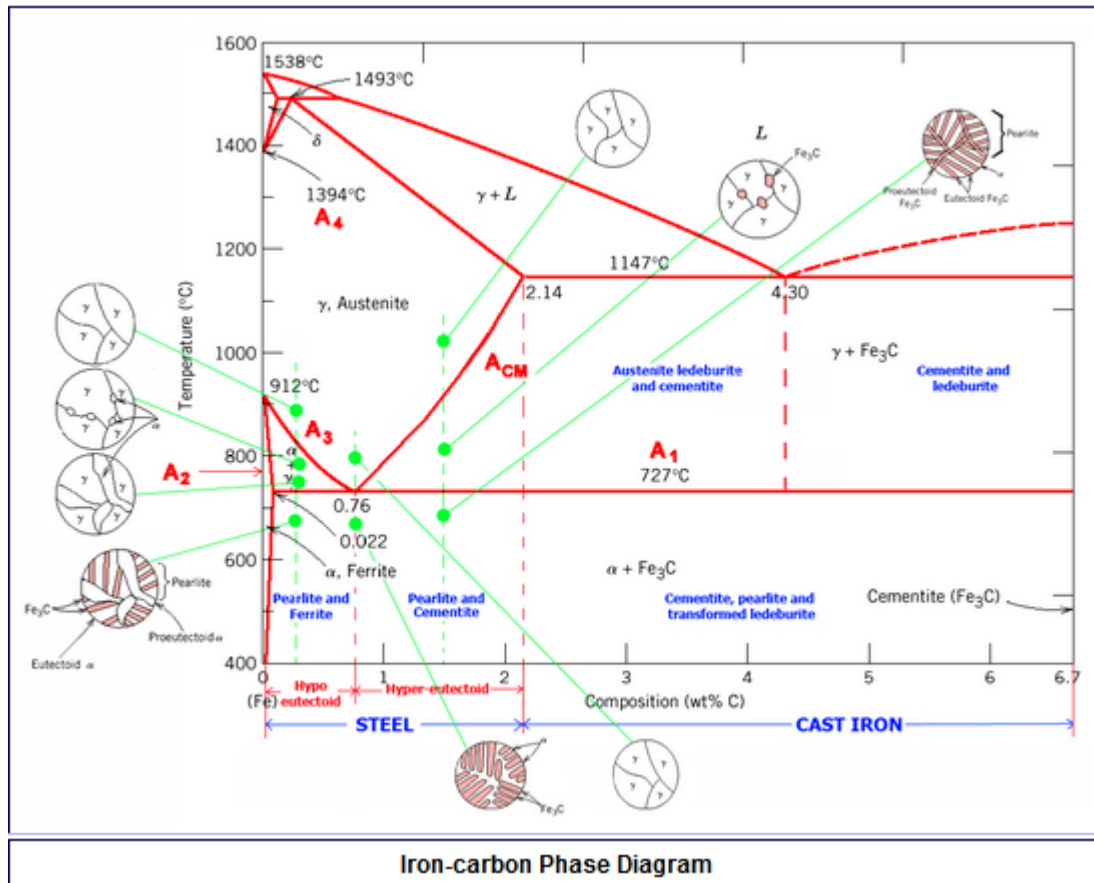
Το απλό ορθορομβικό πλέγμα είναι ταυτόσημο με τη συμβατική κυψελίδα, δηλαδή ισχύουν τα παραπάνω. Υπάρχουν δώδεκα άτομα σιδήρου και τέσσερα άτομα άνθρακα ανά κυψελίδα, όπως δείχνει και η φόρμουλα Fe₃C.



Σχήμα 19 : Συμβατική ορθορομβική κυψελίδα (αριστερά) και απλό ορθορομβικό πλέγμα (δεξιά)



Σχήμα 20 : Κρυσταλλική δομή του σεμεντίτη. Τα μωβ άτομα αναπαριστούν τον άνθρακα. Κάθε άτομο άνθρακα περιστοιχίζεται από οκτώ άτομα σιδήρου. Κάθε άτομο σιδήρου είναι συνδεδεμένο με τρία άτομα άνθρακα. <http://www.phasetrans.msm.cam.ac.uk/2003/Lattices/cementite.html>



Σχήμα 21 : Το διάγραμμα φάσης του σιδήρου-άνθρακα.
https://www.tf.unikiel.de/matwis/amat/iss/kap_6/illustr/s6_1_2.html

- A_1 : Το άνω όριο του πεδίου φάσης του φερρίτη/σεμεντίτη(οριζόντια γραμμή που διασχίζει το ευτηκτοειδές σημείο
- A_2 : Η θερμοκρασία όπου ο σίδηρος απομαγνητίζεται(θερμοκρασία Curie). Ο καθαρός σίδηρος βρίσκεται ακόμα στην α -φάση
- A_3 : Το σύνορο μεταξύ του γ -ωστενίτη και του πεδίου ωστενίτη/φερρίτη
- A_4 : Το σημείο όπου σε υψηλές θερμοκρασίες η α -φάση γίνεται δ -φάση
- A_{CM} : Το όριο μεταξύ του γ -ωστενίτη και του πεδίου ωστενίτη/σεμεντίτη

Ο σεμεντίτης είναι ένα ενδιάμεσο μεταβατικό μέταλλο, είναι ένα σκληρό, ψαθυρό υλικό[56] που κανονικά κατηγοριοποιείται ως κεραμικό στην καθαρή μορφή του, και είναι ένα σημαντικό συστατικό στην μεταλλουργία με σίδηρο, ενώ είναι παρών στους περισσότερους χάλυβες και χυτοσίδηρους[57].

Στο σύστημα σιδήρου-άνθρακα,δηλαδή στους απλούς ανθρακοχάλυβες και χυτοσίδηρους, είναι ένα κοινό συστατικό καθώς ο φερρίτης μπορεί να περιέχει στο μέγιστο 0,02% κατά βάρος άνθρακα[58]. Έτσι, στους ανθρακοχάλυβες και χυτοσίδηρους που κρυστώνουν αργά, ένα κομμάτι του άνθρακα είναι στη μορφή του σεμεντίτη[56]. Ο σεμεντίτης σχηματίζεται κατευθείαν από την τήξη στην περίπτωση του άσπρου χυτοσιδήρου. Στους ανθρακοχάλυβες, ο σεμεντίτης προκαλείται απο

ιζηματοποίηση του ωστενίτη όταν ο ωστενίτης μετατρέπεται σε φερρίτη σε αργό κρύωμα (slow cooling), ή από το μαρτενσίτη κατά τη διάρκεια επαναφοράς (tempering). Ενώ ο σεμεντίτης είναι θερμοδυναμικά ασταθής, μετατρέπόμενος τελικά σε ωστενίτη (μικρή περιεκτικότητα άνθρακα) και γραφίτη (μεγάλη περιεκτικότητα άνθρακα) σε μεγαλύτερες θερμοκρασίες, δεν αποσυντίθεται από θέρμανση σε θερμοκρασίες κάτω από την ευτηκτοειδή θερμοκρασία (περίπου 727° C).

2.4 Γενετικοί αλγόριθμοι

2.4.1 Ιστορική Αναδρομή

Τη δεκαετία του 50' και του 60' αρκετοί επιστήμονες μελέτησαν εξελικτικά συστήματα με την σκέψη ότι η εξέλιξη θα μπορούσε να χρησιμοποιηθεί σαν ένα εργαλείο βελτιστοποίησης σε προβλήματα μηχανικής. Η ιδέα πίσω από όλα αυτά τα συστήματα ήταν η δημιουργία και εξέλιξη ενός αριθμού/πληθυσμού απο υποψήφιες λύσεις για ένα δοθέν πρόβλημα, με τη χρήση τελεστών εμπνευσμένων από τη φυσική γενετική μεταβολή και τη φυσική επιλογή. [59]

Το 1966 οι Fogel, Owens, και Walsh [60] ανέπτυξαν τον εξελικτικό προγραμματισμό (evolutionary programming), μιά τεχνική στην οποία οι υποψήφιες λύσεις αναπαριστώνταν ως πεπερασμένα αυτόματα [33], που εξελίσσονταν με τυχαία μετάλλαξη των διαγραμμάτων της αλλαγής κατάστασής τους, και διαλέγοντας τις καταλληλότερες.

Το 1973 ο Rechenberg [61] ανέπτυξε τις εξελικτικές στρατηγικές (evolution strategies), μιά μέθοδο για τη βελτιστοποίηση πραγματικών παραμέτρων για μηχανήματα όπως αεροτομές.

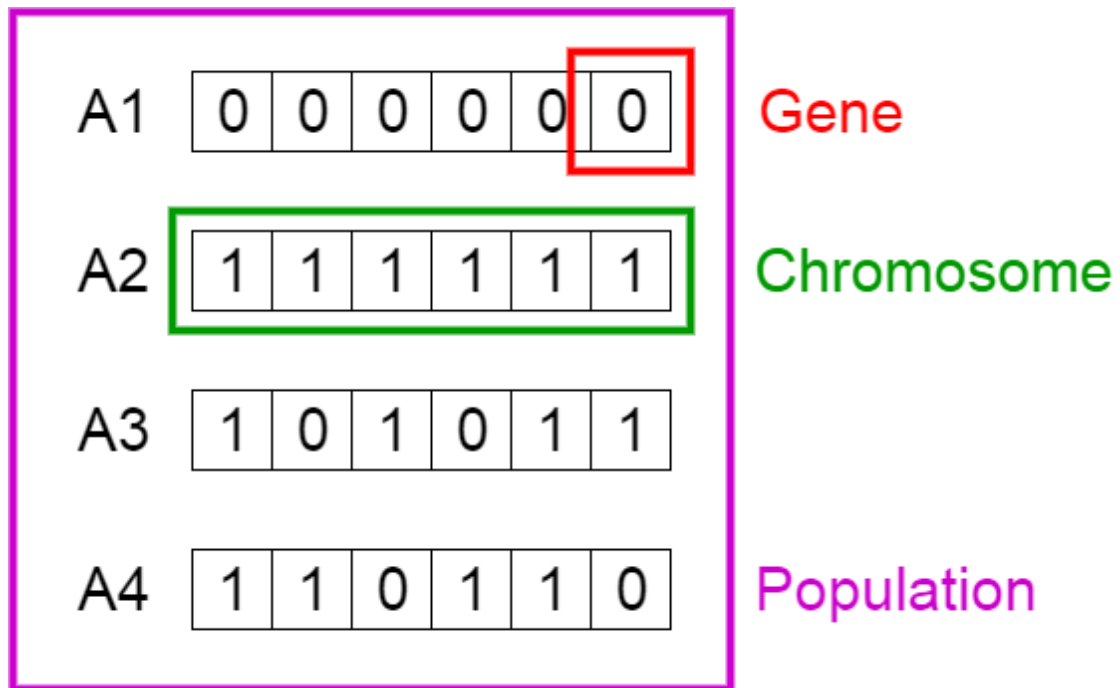
Το 1975 ο Holland [62], έχοντας μελετήσει το φαινόμενο της προσαρμοστικότητας όπως συμβαίνει στη φύση, παρουσίασε τους γενετικούς αλγόριθμους ως προσέγγιση της βιολογικής εξέλιξης και ανέπτυξε ένα θεωρητικό υπόβαθρο για την προσαρμοστικότητα στο πλαίσιο αυτό.

Οι τρεις αυτές κατηγορίες είναι ο κορμός της εξελικτικής υπολογιστικής θεωρίας.

2.4.2 Εισαγωγή

Καθώς οι γενετικοί αλγόριθμοι σχεδιάστηκαν με πρότυπο τη φύση, η ονοματολογία των βασικών τους στοιχείων είναι βιολογική. Τα βασικά στοιχεία σχεδόν κάθε εξελικτικού αλγόριθμου είναι τα παρακάτω :

- Μια συνάρτηση καταλληλότητας (fitness function) προς βελτιστοποίηση
- Ένας πληθυσμός χρωμοσωμάτων (chromosomes)



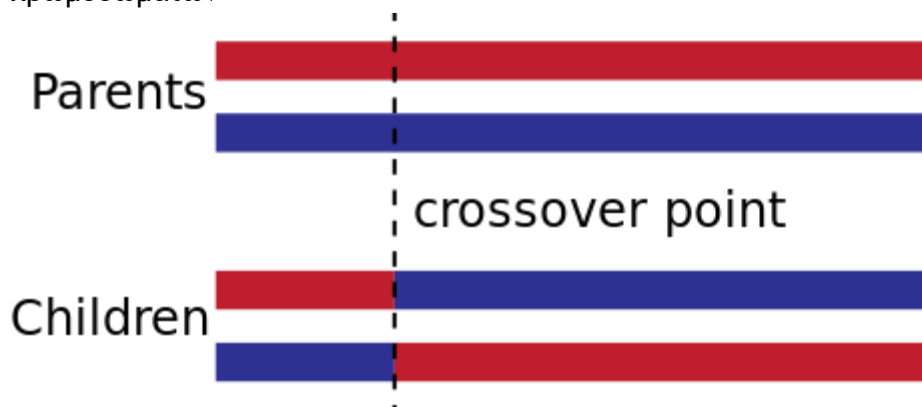
Σχήμα 22: Εξήγηση του γονιδίου, του χρωμοσώματος και του πληθυσμού σε δυαδική μορφή

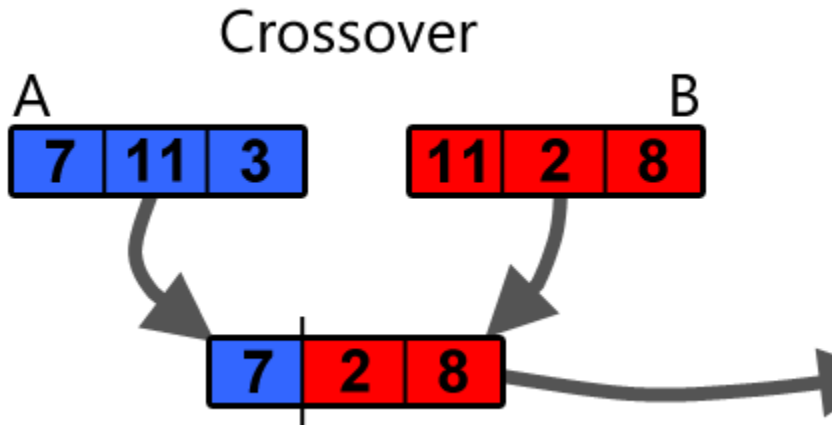
- Επιλογή(selection) για το ποιά από τα χρωμοσώματα θα αναπαράγουν

Fitness	Initial Population		
22	101010100111110101	Selection	Selected parent string one 110011010101011100
9	110011010101011100		
8	111110101111010101	Selection	Selected parent string two 111001111100001001
70	111001111100001001		
19	110011010101011100		
48	101110101111001001		
23	110011010101011100		
38	111001111100001001		

Σχήμα 23 : Επιλογή μεταξύ του πληθυσμού

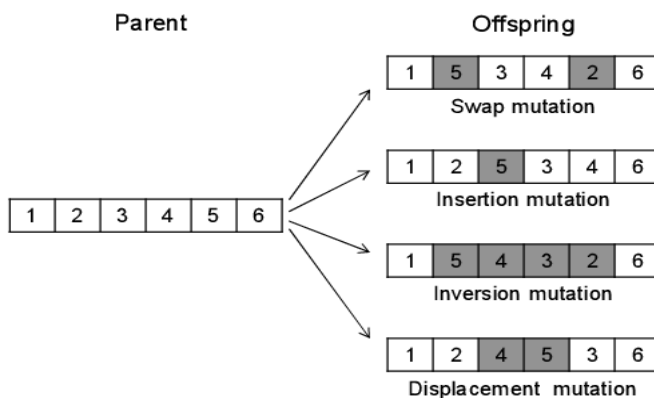
- Διασταύρωση(crossover) με σκοπό την παραγωγή νέας γενιάς χρωμοσωμάτων





Σχήμα 24 : a) Εξήγηση του μηχανισμού διασταύρωσης και b) παράδειγμα διασταύρωσης

- Τυχαία μετάλλαξη (mutation) των χρωμοσωμάτων στη νέα γενιά



Σχήμα 25: Εξήγηση 4 διαφορετικών ειδών μεταλλάξεων

Φυσικά πρέπει να γίνει μια αναλογία με τη μαθηματική ονοματολογία.

Συνάρτηση καταλληλότητας είναι η αντικειμενική συνάρτηση, δηλαδή η συνάρτηση που πρέπει να βελτιστοποιηθεί. [63] Ο όρος καταλληλότητα δανείστηκε από την εξελικτική θεωρία, και χρησιμοποιείται γιατί η συνάρτηση αυτή ελέγχει και ποσοτικοποιεί πόσο κατάλληλη είναι κάθε πιθανή λύση και είναι από τους πιο καθοριστικούς παράγοντες του αλγόριθμου.

Ο όρος χρωμόσωμα αναφέρεται σε μια αριθμητική τιμή ή τιμές που αναπαριστούν μια υποψήφια λύση στο πρόβλημα. Κάθε υποψήφια λύση κωδικοποιείται σαν μια σειρά από παραμετρικές τιμές. Αν ένα πρόβλημα έχει N διαστάσεις, τυπικά κάθε χρωμόσωμα κωδικοποιείται σαν μια σειρά από N τιμές. Μια συνηθισμένη διαδικασία είναι η μετατροπή κάθε παραμετρικής τιμής σε μια σειρά από 0 και 1 (bits), και η ιστορική της σημασία οφείλεται στο ότι τα χρωμοσώματα κωδικοποιούνταν με τον ίδιο τρόπο. Οι σύγχρονοι υπολογιστές επιτρέπουν συνδυασμούς, πραγματικούς αριθμούς κ.α.

Ένας γενετικός αλγόριθμος ξεκινά με μια τυχαία συλλογή από χρωμοσώματα που χρησιμοποιούνται ως ο αρχικός πληθυσμός, η πρώτη γενιά, και μετά κάθε

χρωμόσωμα του πληθυσμού αξιολογείται από τη συνάρτηση καταλληλότητας για να ελεγχθεί πόσο καλά λύνει το πρόβλημα.

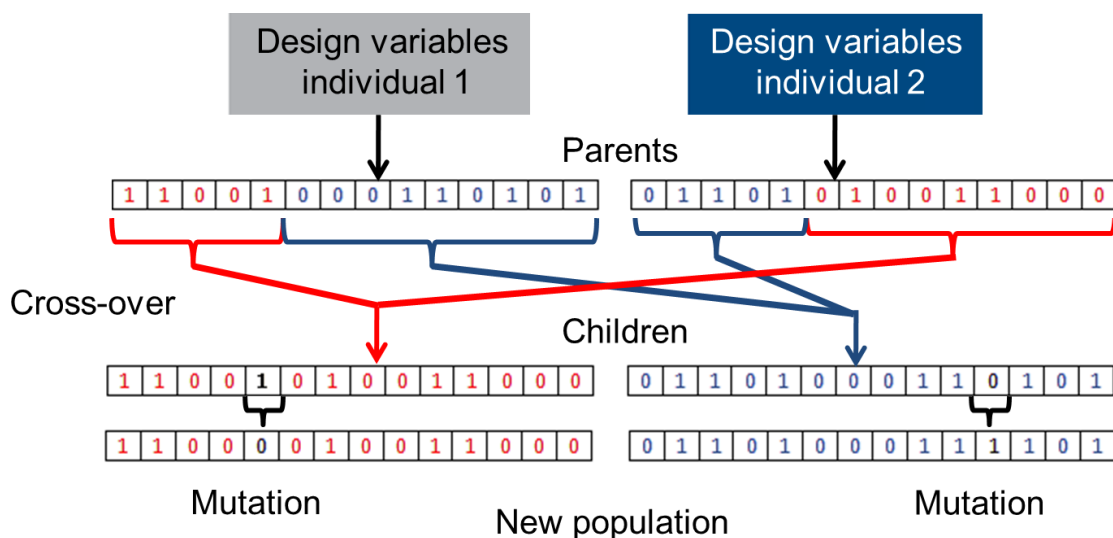
Έπειτα ο τελεστής επιλογής διαλέγει μερικά από τα χρωμοσώματα για αναπαραγωγή, βασιζόμενος σε μια κατανομή πιθανότητας που ορίζει ο χρήστης. Όσο πιο κατάλληλο είναι ένα χρωμόσωμα, τόσο πιο πιθανό είναι να επιλεγεί. Ο τελεστής επιλογής διαλέγει τα χρωμοσώματα με αντικατάσταση, που σημαίνει ότι ένα χρωμόσωμα μπορεί να επιλεγεί παραπάνω από μία φορά.

Ο τελεστής διασταύρωσης μοντελοποιεί τη βιολογική διασταύρωση και τον ανασυνδυασμό των χρωμοσωμάτων στη (βιολογική) μείωση, δηλαδή την κυτταρική διαίρεση κατά τη διάρκεια της αναπαραγωγής. Για παράδειγμα αν 010 και 001 είναι τα δύο γονεϊκά χρωμοσώματα και διασταυρωθούν μετά το δεύτερο bit, τα χρωμοσώματα των απογόνων θα είναι 011 και 000.

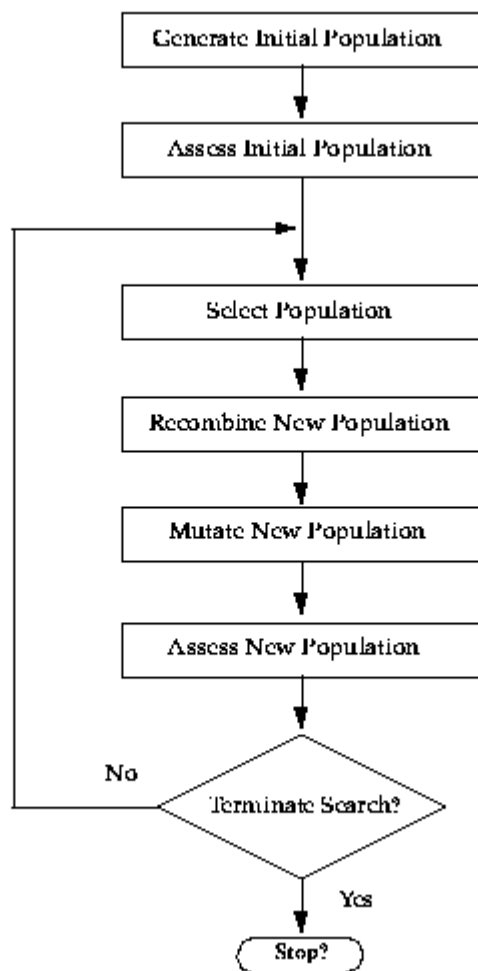
Ο τελεστής μετάλλαξης αλλάζει τυχαία κάποια από τα bits στα νέα χρωμοσώματα (μετατρέπει το 0 σε 1 και αντίστροφα), τυπικά με πολύ μικρή πιθανότητα, της τάξης του 0.01.

Ο τελεστής αυτός παίζει έναν σημαντικό ρόλο, παρόλο που είναι δευτερεύων σε σχέση με τους τελεστές επιλογής και διασταύρωσης. Οι δύο προαναφερθέντες τελεστές διατηρούν τη γενετική πληροφορία των καταλληλότερων χρωμοσωμάτων, αλλά τα χρωμοσώματα αυτά είναι τα καταλληλότερα μόνο σχετικά με την τρέχουσα γενιά. Αυτό μπορεί να οδηγήσει τον αλγόριθμο σε πολύ γρήγορη σύγκλιση και να χάσει «δυναμικά χρήσιμο γενετικό υλικό (0 ή 1 σε συγκεκριμένες θέσεις) [64]. Με άλλα λόγια, ο αλγόριθμος μπορεί να εγκλωβιστεί σε τοπικά βέλτιστα προτού βρει το ολικό μέγιστο. Ο τελεστής μετάλλαξης βοηθά στην επίλυση αυτού του προβλήματος, αλλά μπορεί να οδηγήσει τον αλγόριθμο σε αργότερη σύγκλιση.

Η παραπάνω διαδικασία επαναλαμβάνεται και οι καλύτερες λύσεις κάθε γενιάς αποθηκεύονται, ώσπου για συνεχόμενα τρεξίματα η καλύτερη λύση να μένει σταθερή, οπότε και τερματίζεται ο αλγόριθμος.



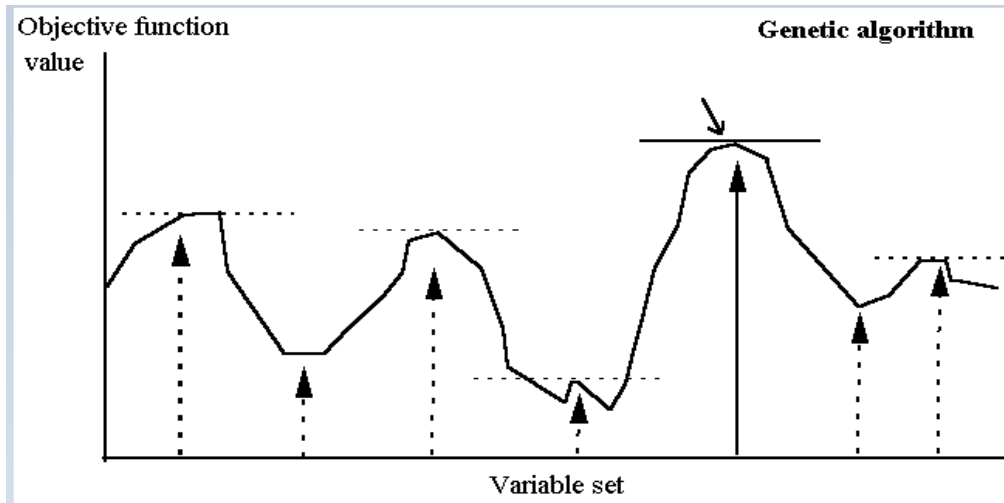
Σχήμα 26 : Εξήγηση αναπαραγωγής



Σχήμα 27 : Διάγραμμα ροής γενετικού αλγορίθμου

Συγκριτικά με τις παραδοσιακές μεθόδους βελτιστοποίησης, παρατηρούνται οι εξής διαφορές[65] :

1. Οι γενετικοί αλγόριθμοι μεταχειρίζονται κωδικοποιημένες εκδοχές των παραμέτρων του προβλήματος αντί για τις ίδιες τις παραμέτρους
2. Ενώ σχεδόν όλες οι συμβατικές μέθοδοι ξεκινάνε την αναζήτηση από ένα σημείο, οι γενετικοί αλγόριθμοι πάντα λειτουργούν με έναν πληθυσμό από σημεία(σειρές) και αυτό συνεισφέρει σημαντικά στη στιβαρότητά τους. Βελτιώνεται έτσι η πιθανότητα εύρεσης του ολικού βέλτιστου, και ,αντίθετα, μειώνεται ο κίνδυνος εγκλωβισμού σε τοπικά βέλτιστα.
3. Οι τυπικοί γενετικοί αλγόριθμοι δεν χρησιμοποιούν καμία βοηθητική πληροφορία σχετικά με την τιμή της αντικειμενικής συνάρτησης, όπως την παράγωγο. Επομένως μπορούν να εφαρμοστούν σε κάθε είδος συνεχούς ή διακριτού προβλήματος βελτιστοποίησης. Το μόνο πράγμα που πρέπει να γίνει είναι ο ορισμός μιας ικανοποιητικής συνάρτησης αποκωδικοποίησης.
4. Οι γενετικοί αλγόριθμοι χρησιμοποιούν πιθανολογικούς τελεστές μετάβασης ενώ οι συμβατικές μέθοδοι για συνεχή βελτιστοποίηση εφαρμόζουν ντετερμινιστικούς τελεστές μετάβασης. Πιο ειδικά, ο τρόπος που υπολογίζεται μια νέα γενιά από την τρέχουσα έχει ορισμένα τυχαία στοιχεία.



Σχήμα 28 : Τοπικά και ολικά βέλτιστα

Κεφάλαιο 3

3 Διεθνής Βιβλιογραφία

Όπως αναλύθηκε παραπάνω, οι εφαρμογές των νευρωνικών δικτύων αγγίζουν πολλούς και διαφορετικούς κλάδους. Παρακάτω θα αναφερθούν ορισμένες δημοσιεύσεις που συνδέουν τη μοριακή δυναμική με τα νευρωνικά δίκτυα.

1. Στις περισσότερες περιπτώσεις, τα νευρωνικά δίκτυα χρησιμοποιούνται για την ανάπτυξη μοντέλων δυναμικού χρησιμοποιώντας υπολογισμούς ηλεκτρονιακής δομής.

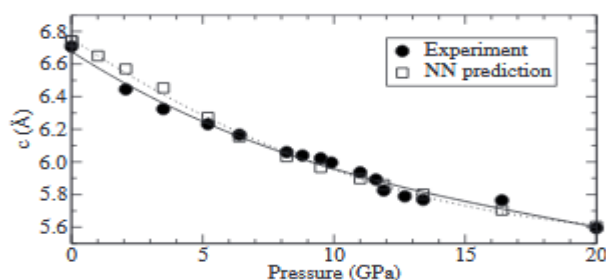
Το 1999 ο Hobday [66] χρησιμοποίησε νευρωνικά δίκτυα για να αναπτύξει μια επιφάνεια δυναμικής ενέργειας για το σύστημα C-H. Οι είσοδοι των νευρωνικών δικτύων αποτελούνταν από όρους ζεύγους όπως συνημίτονα κατεύθυνσης, μήκη δεσμών και συναρτήσεις cut-off ($0 \leq f \leq 1$), πληροφορίες πρώτης γειτνίασης όπως αποστάσεις και γωνίες, και πληροφορίες δεύτερης γειτνίασης όπως ο αριθμός και το είδος των ατόμων. Η έξοδος ήταν η δυναμική ενέργεια ενός ατόμου. Το δίκτυο είχε 5 εισόδους και μία κρυφή στρώση με 6 νευρώνες. Τα αριθμητικά αποτελέσματα δείχνουν ότι η επιφάνεια δυναμικής ενέργειας που δημιουργήθηκε, αν και 60-80% πιο αργή από το δυναμικό Brenner [67], έχει μικρότερο υπολογιστικό κόστος από ab initio υπολογισμούς.

Το 2004 ο Lorenz [68] ανέπτυξε μια επιφάνεια δυναμικής ενέργειας χρησιμοποιώντας νευρωνικά δίκτυα στα οποία η εκμάθηση έγινε με ab initio δεδομένα. Η πιθανότητα προσκόλλησης του H₂ σε μία επιφάνεια παλλαδίου (Pd) καθορίστηκε χρησιμοποιώντας υπολογισμούς μοριακής δυναμικής στην επιφάνεια δυναμικής ενέργειας που δημιουργήθηκε με τα νευρωνικά δίκτυα. Από τις 659 ab initio ενέργειες, 619 χρησιμοποιήθηκαν για training και 40 για testing. Το δίκτυο είχε δύο κρυφές στρώσεις με 24 και 18 νευρώνες αντίστοιχα, 8 εισόδους και 1 έξοδο. Το 99,6% του training σετ και το 95% του testing σετ είχε σφάλμα κάτω από 0.1 eV. Το στατιστικό σφάλμα της πιθανότητας τελικά ήταν 2,2%.

Το 2005 ο Raff [69] χρησιμοποίησε νευρωνικά δίκτυα για να υπολογίσει με παρεμβολή επιφάνειες δυναμικής ενέργειας για τα Si₅ και C₂H₃Br. Για το C₂H₃Br χρησιμοποιήθηκε δομή με 12 εισόδους, μία κρυφή στρώση με 20 νευρώνες και 1 έξοδο. Για το Si₅, τα δεδομένα χωρίστηκαν σε περίπου 80% training, 10% testing και 10% validation. Το δίκτυο είχε 9 εισόδους, μία κρυφή στρώση με 45 νευρώνες, και μία έξοδο. Και στις δύο περιπτώσεις για τη μάθηση του δικτύου χρησιμοποιήθηκε ο αλγόριθμος Levenberg-Marquardt.

Το 2010 οι Khaliullin et al [70] χρησιμοποίησαν νευρωνικά δίκτυα για να αναπαράξουν την ab initio επιφάνεια δυναμικής ενέργειας για τις φάσεις διαμαντιού και γραφίτη του άνθρακα. Η αρχιτεκτονική του δικτύου περιλάμβανε 2 κρυφές στρώσεις με 25 νευρώνες η κάθε μία, και 10% των συνολικών δομών για testing. Το RMSE του training set ήταν 4 meV και του testing 4.9 meV. Έπειτα

πραγματοποιήθηκαν προσομοιώσεις MD για 512 άτομα κυβικού διαμαντιού και 960 άτομα εξαγωνικού γραφίτη με το δυναμικό αυτό με πολύ καλή ακρίβεια.



Σχήμα 29 : Σύγκριση πειραματικών αποτελεσμάτων με την πρόβλεψη των νευρωνικών δικτύων

Επίσης το 2010 οι Eshet et al [71] με τη βοήθεια νευρωνικών δικτύων δημιούργησαν ένα δυναμικό για το Νάτριο με σκοπό να αναπαραχθούν οι ιδιότητες των κρυσταλλικών και υγρών φάσεων υψηλής πίεσης και υψηλής θερμοκρασίας. Το δίκτυο είχε 2 κρυφές στρώσεις με 25 νευρώνες ανά στρώση, και 10% των δεδομένων χρησιμοποιήθηκαν για testing. Το RMSE για το training και το testing σετ αντίστοιχα ήταν 0.72 meV/atom και 0.91 meV/atom, και το Mean Absolute Error 6,18 meV/atom και 7.62 meV/atom.

Το 2015 οι Li και Guo [72] ανέπτυξαν μια επιφάνεια δυναμικής ενέργειας 15 διαστάσεων για την αντίδραση $OH + CH_4 \rightarrow H_2O + CH_3$ χρησιμοποιώντας νευρωνικά δίκτυα με 2 κρυφές στρώσεις, 40 νευρώνες στην πρώτη και 100 στη δεύτερη στρώση, με μέσο σφάλμα 3.9 meV.

Η δημιουργία δυναμικών μέσω νευρωνικών δικτύων για χρήση σε προσομοιώσεις MD δεν περιορίζεται στα μέταλλα, αλλά και στο νερό [73] [74] και σε υλικά αλλαγής φάσης (phase change materials) όπως το GeTe [75]

2. Μια άλλη κατηγορία όπου η μοριακή δυναμική και τα νευρωνικά δίκτυα συνυφαίνονται είναι η βελτιστοποίηση γεωμετρίας (geometry optimization) που ανήκει στην κατηγορία ανάλυσης διακύμανσης (regression analysis).

Το 2016 οι Aci και Ancic [76] χρησιμοποίησαν νευρωνικά δίκτυα με σκοπό πρώτον την πρόβλεψη των ατομικών συντεταγμένων ισορροπίας νανοσωλήνων άνθρακα (carbon nanotubes) χωρίς τη χρήση κάποιου λογισμικού προσομοίωσης και δεύτερον τη χρήση των αποτελεσμάτων των νευρωνικών δικτύων για επιτάχυνση των επαναλήψεων στην νέα υπολογιστική διαδικασία. Το δίκτυο είχε 5 εισόδους, τις 3 συντεταγμένες x,y,z, 2 ακέραιους u και w που καθόριζαν τα διανύσματα ασυμμετρίας (chiral vectors), μία κρυφή στρώση με 10 νευρώνες σε δύο περιπτώσεις και 20 σε μία άλλη (οι οποίες περιπτώσεις συγκρίνονται με βάση τα σφάλματα τους), και 3 έξοδοι, οι τελικές συντεταγμένες της κατάστασης ισορροπίας. Το σετ δεδομένων χωρίστηκε σε 70% για training, 15% για validation και 15% για testing. Το μοντέλο ANFIS που χρησιμοποιήθηκε

Το 2017 οι Aci και Ancic [77] χρησιμοποίησαν ένα μοντέλο ANFIS [78] για να μειώσουν το χρόνο υπολογισμού προσομοιώσεων DFT σε carbon nanotubes. Η αρχιτεκτονική

του δικτύου και ο διαχωρισμός του dataset σε training, testing, και validation έγινε όπως στο προηγούμενο τους άρθρο.

Το 2018 πάλι οι Acı και Avcı [79] συνέκριναν 7 διαφορετικά είδη νευρωνικών δικτύων και μια μέθοδο MLR(multiple linear regression) για τη γεωμετρική βελτιστοποίηση carbon nanotubes με σκοπό τα αποτελέσματά τους είτε να χρησιμοποιηθούν σε φυσικούς υπολογισμούς χωρίς τη χρήση κάποιου λογισμικού προσομοίωσης είτε να χρησιμοποιηθούν σε κάποιο λογισμικό προσομοίωσης, ως αρχικές θέσεις, για την μείωση του υπολογιστικού κόστους.

3. Ακόμα μία κατηγορία όπου τα νευρωνικά δίκτυα χρησιμοποιούνται στις υπολογιστικές προσομοιώσεις είναι η αναγνώριση ενεργειών και δυνάμεων με βάση την ατομική διάταξη του συστήματος.

Το 2012 οι Artrith και Behler [80] χρησιμοποίησαν νευρωνικά δίκτυα με αρχιτεκτονική 51-30-30-1, συναρτήσεις ενεργοποίησης υπερβολικής εφραπτομένης στους κόμβους των δύο κρυφών στρώσεων και γραμμική συνάρτηση ενεργοποίησης στην έξοδο για να προβλέψουν την ατομική ενέργεια και κατά προέκταση τις ατομικές δυνάμεις ενός συστήματος χαλκού, με προοπτική για επέκταση σε μεγαλύτερα συστήματα. Το RMSE ήταν 3.6 meV και 3.9 meV για το training και το testing σε αντίστοιχα. Το αντίστοιχο Mean Absolute Error(MAE) των ενεργειών ήταν 2.09 meV/atom και 2.22 meV/atom και των δυνάμεων 29.3 και 29.4 meV/Bohr. Επίσης μελετήθηκαν και εναλλακτικές αρχιτεκτονικές για τα νευρωνικά δίκτυα τα σφάλματα των ο Το 2012 οι Artrith και Behler [80] χρησιμοποίησαν νευρωνικά δίκτυα με αρχιτεκτονική 51-30-30-1, συναρτήσεις ενεργοποίησης υπερβολικής εφραπτομένης στους κόμβους των δύο κρυφών στρώσεων και γραμμική συνάρτηση ενεργοποίησης στην έξοδο για να προβλέψουν την ατομική ενέργεια και κατά προέκταση τις ατομικές δυνάμεις ενός συστήματος χαλκού, με προοπτική για επέκταση σε μεγαλύτερα συστήματα. Το RMSE ήταν 3.6 meV και 3.9 meV για το training και το testing σε αντίστοιχα. Το αντίστοιχο Mean Absolute Error(MAE) των ενεργειών ήταν 2.09 meV/atom και 2.22 meV/atom και των δυνάμεων 29.3 και 29.4 meV/Bohr. Επίσης μελετήθηκαν και εναλλακτικές αρχιτεκτονικές για τα νευρωνικά δίκτυα τα σφάλματα των οποίων φαίνονται στο παρακάτω σχήμα :

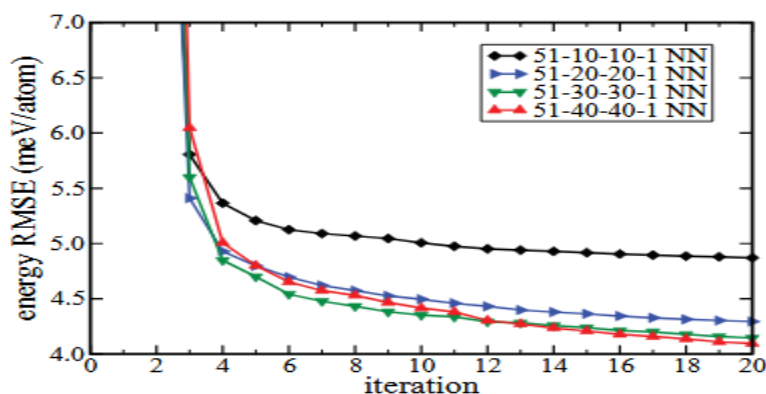


FIG. 4. (Color online) Comparison of the fitting errors of the training set for several neural network (NN) architectures.

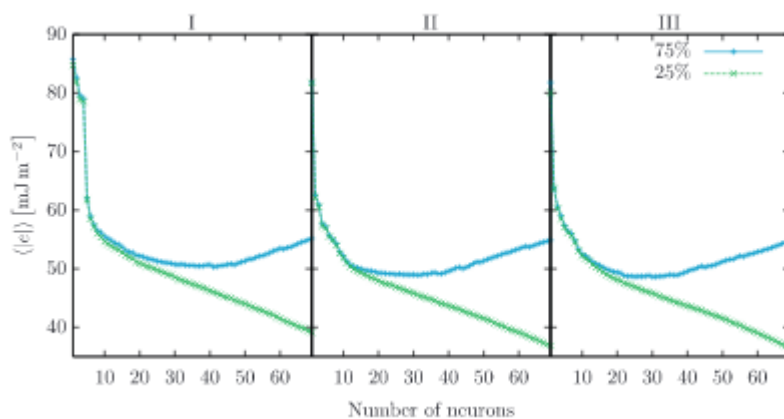
Σχήμα 30 : Σύγκριση σφάλματος training για διάφορες αρχιτεκτονικές νευρωνικών δικτύων

Το 2014 οι Restrepo, Giraldo και Thijssse [81] μελέτησαν την ενέργεια ορίου κόκκου (grain boundary energy) [82] με τη βοήθεια νευρωνικών δικτύων. Από τα 68705 δεδομένα που είχαν, το 25% χρησιμοποιήθηκε για την μάθηση του δικτύου(training) και το 75% για testing, αφού πρώτα χωρίστηκαν σε 3 σετ δεδομένων. Η αρχιτεκτονική του cascade νευρωνικού δικτύου [83] ήταν 5 εισόδοι, οι γωνίες του ορίου κόκκου, 1 έξοδος, η ενέργεια, και έγινε βελτιστοποίηση όσον αφορά τον αριθμό των νευρώνων της κρυφής στρώσης, με τον τελικό βέλτιστο αριθμό των νευρώνων να είναι 41,32 και 30 για τα τρία σετ δεδομένων. Τα σφάλματα του training και του testing για τα τρία dataset, όπως και το μέσο απόλυτο σφάλμα ως συνάρτηση του αριθμού των νευρώνων στην κρυφή στρώση φαίνονται στα παρακάτω σχήματα.

Table 2
Best fitted ANNs for datasets I, II and III. $\langle |e_{Test}| \rangle$ and $\langle |e_{Train}| \rangle$ are the average absolute error for the prediction and training dataset, respectively. The values of the error are presented in millijoules per square metre (mJ m^{-2}) and, within parentheses, as a percentage of the average GB energy.

Dataset	Number of neurons	$\langle e_{Test} \rangle$	$\langle e_{Train} \rangle$
I	41	50.3 (4.14)	46.0 (3.79)
II	32	48.9 (4.03)	45.3 (3.73)
III	30	48.6 (4.00)	45.8 (3.77)

Σχήμα 31 : Mean Absolute Error και Mean Absolute Percentage Error ως συνάρτηση του αριθμού των νευρώνων



Σχήμα 32 : Μέσο απόλυτο σφάλμα ως συνάρτηση του αριθμού των νευρώνων

Το 2014 οι Castin, Fernandez και Pasianot [84] χρησιμοποίησαν νευρωνικά δίκτυα για να προβλέψουν την ενέργεια μετάστασης (migration energy) των ατόμων υλικού bcc FeCr. Η αρχιτεκτονική του cascade νευρωνικού δικτύου περιλάμβανε 3 εισόδους σχετικές με τη γεωμετρία του συστήματος, δύο κρυφές στρώσεις με 1 νευρώνα η καθεμιά, και 1 έξοδο, τη migration energy. Τα συνολικά δεδομένα ήταν 62000 από τα οποία τα 40300 χρησιμοποιήθηκαν για τη μάθηση του δικτύου και τα 21700 για testing. Το μέσο σφάλμα που παρατηρήθηκε ήταν 100 meV (ως άθροισμα των επιμέρους ενεργειών ανά άτομο).

Το 2015 οι Artrith και Kolpak [85] σε ένα ενδιαφέρον άρθρο χρησιμοποίησαν δυναμικό κατασκευασμένο μέσω νευρωνικών δικτύων για να πραγματοποιήσουν προσομοιώσεις Molecular Dynamics και Monte Carlo [86] σε νανοκράμα (nanoalloy) χαλκού-χρυσού (Cu-Au). Η αρχιτεκτονική του βέλτιστου σε υπολογιστικό κόστος

δικτύου ήταν 148-10-10-1, δηλαδή 148 είσοδοι(που έχουν να κάνουν με τη γεωμετρία του συστήματος), 2 κρυφές στρώσεις με 10 νευρώνες ανά στρώση και μία έξοδος, η ενέργεια/άτομο. Τα 24995 δεδομένα που πάρθηκαν από DFT μοιράστηκαν σε 90% για training και 10% για testing. Μια σύγκριση διαφόρων αρχιτεκτονικών του δικτύου φαίνονται στο παρακάτω σχήμα. Έπειτα το δυναμικό δοκιμάστηκε με προσομοίωση MD και παράλληλα εξάχθηκαν οι ατομικές συντεταγμένες ώστε να υπολογιστούν οι ενέργειες με DFT, σε ένα σύμπλεγμα 55 ατόμων Cu-Au για να επιβεβαιωθεί η ακρίβεια του δυναμικού. Η διαφορά της ενέργειας του νευρωνικού δικτύου με αυτή της DFT ήταν συστηματικά -11.8 meV/άτομο και η τυπική απόκλιση 7.6 meV/άτομο. Όσον αφορά το αρχικό σφάλμα (RMSE και MAE)του δικτύου, αυτό φαίνεται αναλυτικά στον παρακάτω πίνακα.

Network architecture	Weights per element	RMSE (meV/atom)		MAE (meV/atom)	
		Training set	Test set	Training set	Test set
148-2-2-1	307	21.32	23.36	16.06	17.09
148-5-5-1	781	7.14	7.05	4.32	4.85
148-10-10-1	1611	6.81	7.55	4.15	4.34
148-15-15-1	2491	6.35	6.37	3.60	3.80
148-20-20-1	3421	6.44	8.89	3.89	4.31

Σχήμα 33 : Σύγκριση του RMSE,MAE και του αριθμού των βαρών ανάλογα με την αρχιτεκτονική του δικτύου

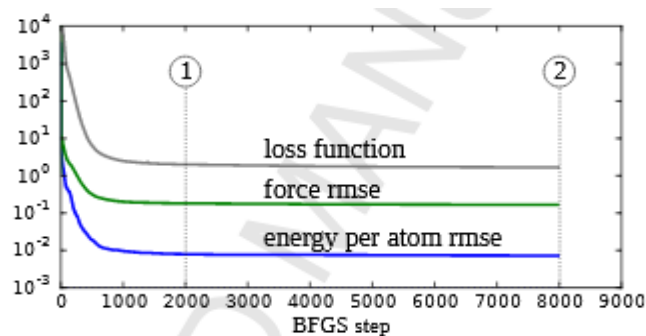
Το 2016 οι Khorsidi και Peterson [87]χρησιμοποίησαν νευρωνικά δίκτυα για να προβλέψουν την ενέργεια fcc χρυσού (Au) και προσροφητών (adsorbates) μονοξειδίου του άνθρακα (CO) και υδρογόνου (H) σε επιφάνεια μολυβδαινίου (Mo).

Το δίκτυο πρόβλεψης της ενέργειας του χρυσού είχε 2 κρυφές στρώσεις με 1 νευρώνα ανά στρώση, ως εισόδους τις ατομικές διαμορφώσεις, και 1 έξοδο, την ενέργεια ανά άτομο. Για τη μάθηση του δικτύου χρησιμοποιήθηκαν 100 ατομικές εικόνες, από τις οποίες οι 60 για training και οι 40 για testing. Το τελικό σφάλμα στις εναπομείνουσες ενέργειες ήταν λιγότερο από 0.1 meV/atom.

Το σύστημα με το μολυβδαίνιο είχε 3 κρυφές στρώσεις με 10 νευρώνες η καθεμιά, συνάρτηση ενεργοποίησης υπερβολική εφαπτομένη και αλγόριθμο μάθησης Broyden-Fletcher-Goldfarb-Shanno(BFGS). Ως είσοδοι χρησιμοποιήθηκαν οι ατομικές διαμορφώσεις 1977 ατόμων, εκ των οποίων 1497 για training και 480 για testing. Μετά από 600 ώρες CPU με συνολική μνήμη 125 GB το RMSE της ενέργειας/άτομο και της δύναμης/άτομο έφτασε τα 7.884 meV/atom και 0.182eV/A αντίστοιχα(σημείο 1 στο διάγραμμα). Μια μεγαλύτερη μείωση του RMSE σε 7.153 meV/atom και 0.166 eV/A (σημείο 2 στο διάγραμμα) πήρε 1000 ώρες παραπάνω.

Το 2007 οι Behler και Parinello [88]ανέπτυξαν με νευρωνικά δίκτυα μια αναπαράσταση μιάς επιφάνειας δυναμικής ενέργειας, που παρέχει την ενέργεια και τις δυνάμεις ως συνάρτηση των ατομικών συντεταγμένων του συστήματος, και είναι 5 τάξεις μεγέθους πιο γρήγορη από την μέθοδο DFT. Το παραπάνω μοντέλο εφαρμόστηκε στο πυρίτιο (Si) και αποδείχθηκε γρηγορότερη από την DFT. Περίπου 9000 ενέργειες υπολογίστηκαν από τις οποίες οι 8200 χρησιμοποιήθηκαν για training και οι 800 για testing του νευρωνικού δικτύου.Το δίκτυο είχε 48 εισόδους

και 2 κρυφές στρώσεις με περίπου 40 νευρώνες ανά στρώση. Το root mean squared error (RMSE) για το training set ήταν 4-5 meV και για το testing set 5-6 meV. Όσον αφορά τις ατομικές δυνάμεις, είχαν RMSE 0.2 eV/Å σε σχέση με τη DFT.



Σχήμα 34 : Αναπαράσταση των σφαλμάτων δυνάμεων και ενέργειας/άτομο ανάλογα με το iteration

4. Συγκεκριμένα παραδείγματα που δείχνουν ότι τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν για τη μελέτη ιδιοτήτων των υλικών.

Το 2008 οι Behler, Martonak, Donadio και Parinello [89] μελέτησαν τις μεταβάσεις του πυριτίου από φάση σε φάση με τα νευρωνικά δίκτυα της παραπάνω μεθόδου και σύγκριναν τα αποτελέσματα με τα αντίστοιχα πειραματικά. Τα νευρωνικά δίκτυα πρόβλεψαν ικανοποιητικά όλες τις αλλαγές φάσης του πυριτίου και τα αντίστοιχα φαινόμενα αλλαγής φάσης εκτός από την τελευταία, δηλαδή τη μετάβαση από hcp σε fcc, λόγω του ότι το νευρωνικό δίκτυο είχε δεδομένα μάθησης από 0 έως 100 GPa και η τελευταία μετάβαση γίνεται στα 104 GPa.

Η παραπάνω μέθοδος επεκτάθηκε για παραπάνω από ένα υλικά, προσθέτοντας τις ηλεκτροστατικές αλληλεπιδράσεις, το 2011 από τους Artrith, Morawietz και Behler [90] και εφαρμόστηκε στο μονοξείδιο του ψευδαργύρου (ZnO). Πάλι το σετ δεδομένων διαχωρίστηκε σε 90% για training και 10% για testing, και χρησιμοποιήθηκαν νευρωνικά δίκτυα με 2 κρυφές στρώσεις και 20 νευρώνες ανά στρώση. Ακόμα χρησιμοποιήθηκε ο αλγόριθμος μάθησης conjugate gradient και τα τελικά σφάλματα RMSE των φορτίων ήταν 0.0088 e/atom στο training και 0.0090 e/atom στο testing.

Κεφάλαιο 4

4 Περιγραφή προβλήματος

Η κεντρική ιδέα της παρούσας διπλωματικής είναι η επιτάχυνση του χρόνου πραγματοποίησης προσομοιώσεων μοριακής δυναμικής, μέσω της χρήσης αρχικών συντεταγμένων υπολογισμένων από τεχνητά νευρωνικά δίκτυα. Παρουσιάζεται η επιτάχυνση της ενεργειακής χαλάρωσης μιάς δομής σεμεντίτη Fe₃C με 512 άτομα στο ελεύθερο λογισμικό προσομοιώσεων μοριακής δυναμικής LAMMPS, ύστερα από χρήση νευρωνικών δικτύων η αρχιτεκτονική των οποίων βελτιστοποιήθηκε με γενετικούς αλγόριθμους. Τα βήματα που ακολουθήθηκαν για την παραπάνω διαδικασία είναι τα εξής:

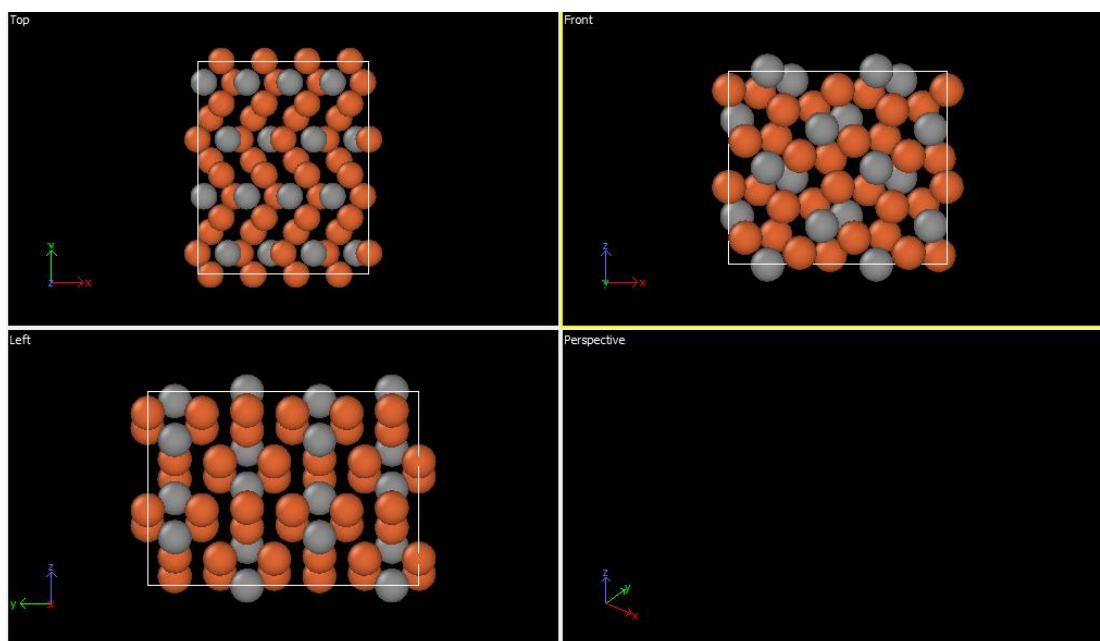
1. Επιλογή ενός κατάλληλου υλικού για μελέτη, η μελέτη του οποίου έχει ενδιαφέρον στο πλαίσιο του τομέα των κατεργασιών, στην περίπτωσή μας του σεμεντίτη
2. Εύρεση μιάς κατάλληλης αρχικής δομής για το σεμεντίτη και επέκτασή της στις δύο διαστάσεις.
3. Βελτιστοποίηση γεωμετρίας της παραπάνω διευρυμένης δομής μέσω του λογισμικού Lammps ώστε η ενέργεια και η πίεση του συστήματος να ελαχιστοποιούνται, και μέτρηση του απαιτούμενου χρόνου
4. Εξαγωγή των αρχικών συντεταγμένων που χρησιμοποιήθηκαν στη βελτιστοποίηση γεωμετρίας και των τελικών συντεταγμένων ισορροπίας του συστήματος
5. Δημιουργία νευρωνικών δικτύων στη Matlab με σκοπό τον υπολογισμό της σχέσης μεταξύ αρχικών και τελικών συντεταγμένων
6. Προσδιορισμός της αρχιτεκτονικής του δικτύου με δοκιμές (trial and error) και υπολογισμός του Mean Squared Error, του Mean Absolute Percentage Error, του Regression, και των σφαλμάτων MSE του training, testing και validation
7. Προσδιορισμός της βέλτιστης αρχιτεκτονικής του δικτύου με βάση το training MSE και το MAPE με τη χρήση γενετικών αλγορίθμων και διαλογή των δύο καλύτερων δικτύων για χρήση στο Lammps.
8. Χρησιμοποίηση των δύο βέλτιστοποιημένων νευρωνικών δικτύων για εύρεση των συντεταγμένων που θα χρησιμοποιηθούν ως νέες είσοδοι στο λογισμικό Lammps
9. Εισαγωγή των συντεταγμένων εξόδου των νευρωνικών δικτύων στο λογισμικό Lammps και μέτρηση του απαιτούμενου χρόνου γεωμετρικής βελτιστοποίησης
10. Σύγκριση του αρχικού και του τελικού χρόνου και εξαγωγή συμπερασμάτων

Η ιδέα μας βασίστηκε στο προαναφερθέν άρθρο των Aci και Avcı [76] οι οποίοι παρουσίασαν μια μέθοδο που επιταχύνει τις προσομοιώσεις DFT στο εμπορικό λογισμικό CASTEP από μέρες (για μεγάλα συστήματα) σε ώρες ή λεπτά. Παρόλο που οι προσομοιώσεις μοριακής δυναμικής ολοκληρώνονται πολύ πιο γρήγορα από τις

προσομοιώσεις DFT, παρουσιάζεται σαφής μείωση του υπολογιστικού κόστους στην περίπτωση ενεργειακής χαλάρωσης 512 ατόμων σεμεντίτη.

Τα παραπάνω 10 βήματα αναλύονται διεξοδικά παρακάτω :

1. Ο σεμεντίτης είναι ένα υλικό που, όπως φαίνεται και από το διάγραμμα φάσης του σιδήρου άνθρακα που παρατίθεται παραπάνω, εμφανίζεται πολύ συχνά κατά την κατεργασία του ανθρακοχάλυβα και του χυτοσιδήρου, άρα η μελέτη του έχει ενδιαφέρον για τον τομέα των κατεργασιών.
2. Μετά την επιλογή του υλικού, είναι αναγκαίο να βρεθεί μία κατάλληλη ατομική διάταξη προς μελέτη. Για το σκοπό αυτό χρησιμοποιήθηκε μια δομή [54] 136 ατόμων σεμεντίτη που φαίνεται στο παρακάτω σχήμα με τη βοήθεια του OVITO.



Σχήμα 35 : Αρχική δομή 136 ατόμων σεμεντίτη

Ύστερα, η παραπάνω δομή επεκτάθηκε 2 φορές κατά x και 2 φορές κατά y έτσι ώστε το πρόβλημα να γίνει πιο σύνθετο και χρήσιμο για περαιτέρω μελέτη, με την τελική δομή μετά την επέκταση να περιλαμβάνει 512 άτομα.

3. Αφού επεκτάθηκε η δομή, χρησιμοποιήθηκε το λογισμικό LAMMPS [91] για να πραγματοποιήσουμε μία προσομοίωση γεωμετρικής βελτιστοποίησης, δηλαδή υπολογισμού της διάταξης με την ελάχιστη δυνατή ενέργεια, για την δομή με τα 512 άτομα σεμεντίτη. Το λογισμικό LAMMPS διανέμεται ελεύθερα και λειτουργεί στο λειτουργικό σύστημα Linux. Ύστερα από τη δημιουργία ενός σχετικά απλού κώδικα, εισήχθησαν τα δεδομένα στο πρόγραμμα με ένα αρχείο `read_data` και εξήχθησαν τα αποτελέσματα σε μορφή ατομικών συντεταγμένων με την εντολή `dump`. Η ελαχιστοποίηση ενέργειας έγινε με τον αλγόριθμο `conjugate gradient` και οι εναπομείνουσες τάσεις μετά την ελαχιστοποίηση είχαν τιμές της τάξης του 10^{-3} . Τα αποτελέσματα της αρχικής προσομοίωσης αναλύονται σε επόμενο κεφάλαιο.

4. Οι εισοδοι και οι έξοδοι του προγράμματος αποθηκεύτηκαν με σκοπό να χρησιμοποιηθούν ως inputs και targets στη Matlab για τη μάθηση(training) των νευρωνικών δικτύων
5. Τα παραπάνω δεδομένα μεταφέρθηκαν στη Matlab όπου και δημιουργήθηκε ένας κώδικας για την δοκιμή διαφόρων νευρωνικών δικτύων.
6. Στην πραγματικότητα δεν υπάρχει κάποιος κανόνας που ορίζει τον αριθμό των νευρώνων που προσεγγίζουν ικανοποιητικά ένα δεδομένο πρόβλημα, ή άλλων μεταβλητών όπως το ποσοστό που αναλογεί στο training, το testing και το validation του σετ δεδομένων. Για το λόγο αυτό, πραγματοποιήσαμε 16 δοκιμές για τον αριθμό των νευρώνων, αυξάνοντάς τους κατα 5 σε κάθε κρυφή στρώση και μελετώντας τους συνδυασμούς στο διάστημα [5,20]. Ακόμα, επειδή κάθε φορά που εκκινείται η διαδικασία τα βάρη σε κάθε κόμβο αρχικοποιούνται τυχαία, για κάθε συνδυασμό τρέξαμε το πρόγραμμα 10 φορές για να βεβαιωθούμε ότι τα σφάλματα δεν οφείλονταν σε τυχαίο προσδιορισμό των αρχικών βαρών. Τα αποτελέσματα της διαδικασίας αναλύονται σε επόμενο κεφάλαιο.
7. Για τη συστηματική βελτιστοποίηση του παραπάνω προβλήματος χρησιμοποιήθηκαν γενετικοί αλγόριθμοι στη Matlab. Για τη βελτιστοποίηση του MSE του training δοκιμάστηκαν 9 διαφορετικοί συνδυασμοί crossover fraction και elite count κρατώντας σταθερά πληθυσμό 50 και 500 γενιές, και οι ίδιοι συνδυασμοί crossover fraction και elite count με 20 πληθυσμό και 500 γενιές. Για το MAPE δοκιμάστηκαν οι δύο καλύτεροι από τους παραπάνω 9 συνδυασμούς για πληθυσμό 50 και 500 γενιές, και για μια μείξη MAE(mean absolute error) και MSE οι δύο προαναφερθέντες καλύτεροι συνδυασμοί. Τα αποτελέσματα των γενετικών αλγορίθμων αναλύονται σε επόμενο κεφάλαιο.
8. Τα δύο καλύτερα νευρωνικά δίκτυα με βάση το MSE του training και το MAPE χρησιμοποιήθηκαν, ύστερα από ένα τρέξιμο στη Matlab, για να δημιουργήσουν τις νέες συντεταγμένες εισόδου της προσομοίωσης μοριακής δυναμικής στο Lammps
9. Το τελικό στάδιο είναι η πραγματοποίηση προσομοίωσης μοριακής δυναμικής στο Lammps με τις συντεταγμένες που δημιουργήθηκαν από τα δύο βέλτιστοποιημένα νευρωνικά δίκτυα. Τα αποτελέσματα των δύο τελικών προσομοιώσεων αναλύονται σε επόμενο κεφάλαιο
10. Κλείνοντας, συγκρίνεται ο αρχικός και οι τελικοί χρόνοι των προσομοιώσεων και αναφέρεται η χρησιμότητα της όλης διαδικασίας, που για προσομοιώσεις με μεγάλο υπολογιστικό κόστος έχει ιδιαίτερο νόημα.

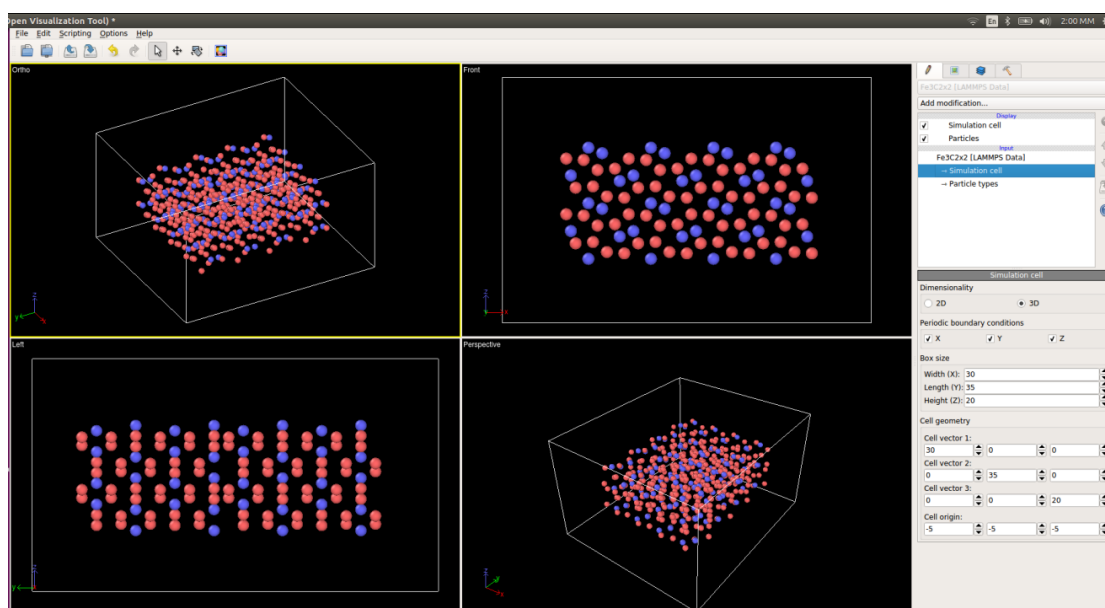
Κεφάλαιο 5

5 Ανάλυση αποτελεσμάτων

5.1 Αποτελέσματα αρχικής προσομοίωσης

Για την προσομοίωση μοριακής δυναμικής στο LAMMPS χρειάστηκε ένας κατάλληλος κώδικας για τη γεωμετρική βελτιστοποίηση της δομής του συστήματος και ένα αρχείο `read_data` για την εισαγωγή των αρχικών θέσεων του συστήματος στο πρόγραμμα.

Η αρχική δομή του σεμεντίτη 512 ατόμων είχε συνολική ενέργεια -1710.18053485 eV, και αρχική μέγιστη δύναμη 38.1487 eV/Å. Η δομή του συστήματος πριν την προσομοίωση φαίνεται στο παρακάτω σχήμα :

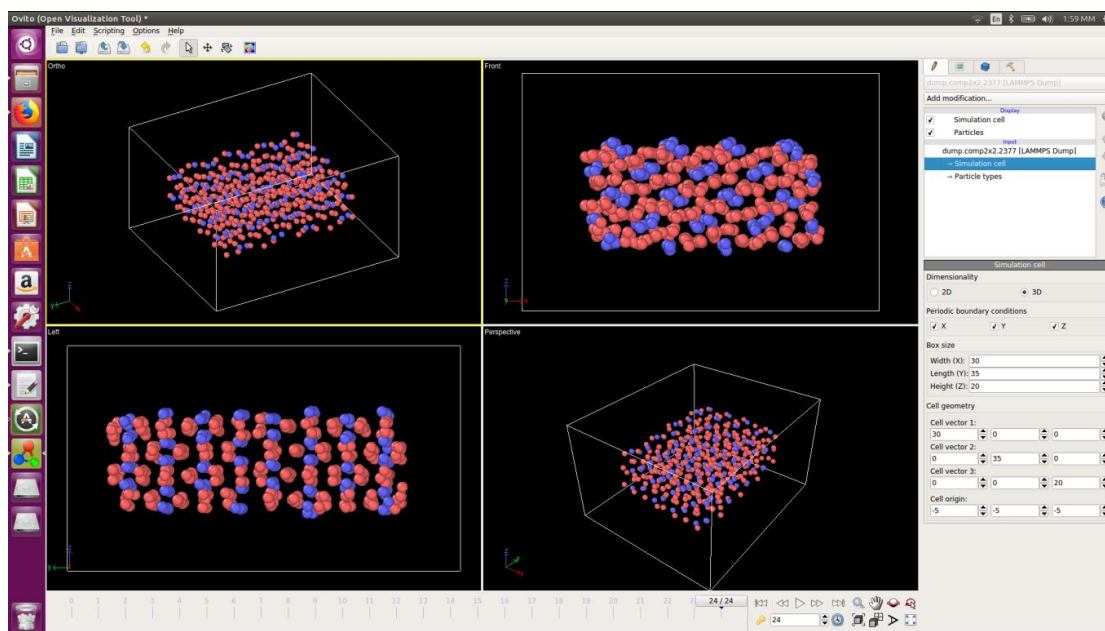


Σχήμα 36 : Αρχική δομή 512 ατόμων σεμεντίτη

Η προσομοίωση διήρκεσε 8.87448 δευτερόλεπτα με 2 πυρήνες 1.8GHz . Χρειάστηκαν 2377 επαναλήψεις και 7520 υπολογισμοί δυνάμεων για τον υπολογισμό της βέλτιστης ενεργειακά δομής. Χρησιμοποιήθηκε δυναμικό EAM για το σύστημα Fe-C και ο αλγόριθμος βελτιστοποίησης `conjugate gradient`. Στόχος ήταν η βελτιστοποίηση της γεωμετρίας του συστήματος, δηλαδή η ελαχιστοποίηση της τελικής ενέργειας, των τελικών πιέσεων στον περιβάλλοντα χώρο και των δυνάμεων μεταξύ των ατόμων.

Η τελική προκύπτουσα δομή είχε ενέργεια -2010.38352235 eV και η τελική μέγιστη δύναμη ήταν $1.91855e-05$ eV/Å.

Η δομή του συστήματος μετά την προσομοίωση φαίνεται στο σχήμα :



Σχήμα 37 : Τελική δομή 512 ατόμων σεμεντίτη

5.2 Αποτελέσματα νευρωνικών δικτύων

Για το συσχετισμό δεδομένων εισόδου (inputs) και εξόδου (targets) στα νευρωνικά δίκτυα χρειάστηκε ένας κατάλληλος κώδικας για εφαρμογή στη MATLAB, όπως φαίνεται παραπάνω. Επίσης χρειάστηκε η επιλογή ορισμένων παραμέτρων των νευρωνικών δικτύων. Αναλυτικά επιλέχθηκαν :

- Ο αλγόριθμος Levenberg-Marquardt για την μάθηση του δικτύου λόγω της προσαρμοστικότητάς του.
- Ποσοστό δεδομένων για training = 75% του συνόλου
- Ποσοστό δεδομένων για validation = 15% του συνόλου
- Ποσοστό δεδομένων για testing = 10% του συνόλου
- Μέθοδος ελέγχου της απόδοσης του δικτύου : Mean Squared Error(MSE) και MAPE(Mean Absolute Percentage Error)
- Μέγιστος αριθμός επαναλήψεων = 1000
- Στόχος MSE = 0, αν και πειραματιστήκαμε με διάφορα performance goals κατά τη διάρκεια των υπολογισμών
- Μέγιστα σφάλματα validation για early-stopping = 6, αν και πειραματιστήκαμε με διάφορες τιμές για τα μέγιστα σφάλματα validation
- Ελάχιστη κλίση της απόδοσης για early-stopping = $1e-07$, αν και έγιναν δοκιμές για διαφορετικές τιμές
- Συνάρτηση ενεργοποίησης για τις κρυφές στρώσεις : υπερβολική εφαπτομένη (tanh) αν και έγιναν δοκιμές με τη λογαριθμική σιγμοειδή.
- Συνάρτηση ενεργοποίησης για τη στρώση εξόδου : γραμμική

Όπως προαναφέρθηκε, έχουν πραγματοποιηθεί υπολογισμοί για διαφορετικούς συνδυασμούς αριθμών νευρώνων στις 2 κρυφές στρώσεις ενός νευρωνικού δικτύου με τις παραπάνω σταθερές παραμέτρους. Τα αποτελέσματα παρουσιάζονται αναλυτικά παρακάτω :

5.2.1 3-5-X-3

	Μέσος όρος 10 δοκιμών			
Δίκτυο	3--5--5--3	3--5--10--3	3--5--15--3	3--5--20--3
best_MSE_val	0,034461727	0,030177525	0,030653567	0,028692516
best_MSE_test	0,033141862	0,030055723	0,032341384	0,029681444
best_MSE_train	0,031756413	0,026648301	0,023735204	0,021018435
regression total	0,99963	0,999684	0,999706	0,999735
time (s)	2,7724	1,4395	2,2402	2,2753
MAPE(%)	12,80782689	11,61972217	11,02478245	10,34421327

Πίνακας 1

	Overall best 10 δοκιμών			
Δίκτυο	3--5--5--3	3--5--10--3	3--5--15--3	3--5--20--3
best_MSE_val	0,028482661	0,022090403	0,023532597	0,024845234
best_MSE_test	0,029475923	0,024625278	0,024866052	0,023876583
best_MSE_train	0,02536262	0,020759839	0,019625556	0,017289762
regression total	0,9997	0,99974	0,99975	0,99977
time	1,10000	0,53400	0,49800	0,62700
MAPE(%)	11,47166702	10,0182249	9,736833137	9,135793686

Πίνακας 2

Αρχικά συγκρίνεται ο μέσος όρος καθεμιάς από τις 10 δοκιμές και η βέλτιστη τιμή που υπολογίστηκε σε κάθε τετράδα αρχιτεκτονικών για να εξαχθούν χρήσιμα συμπεράσματα. Στην περίπτωση που η πρώτη κρυφή στρώση έχει 5 νευρώνες παρατηρείται κυριαρχία των 20 νευρώνων ως μεταβλητής προς βελτιστοποίηση καθώς εκεί παρατηρούνται οι καλύτερες τιμές κατά μέσο όρο αν εξαιρέσουμε το χρόνο, γεγονός λογικό καθώς το υπολογιστικό κόστος ανεβαίνει όσο περισσότεροι νευρώνες προστίθενται, στη γενική περίπτωση, αφού πρέπει να υπολογιστούν παραπάνω βάρη και biases. Όμως αυτή η τάση δεν φαίνεται από τον παραπάνω πίνακα, γεγονός που οφείλεται στις τυχαίες αρχικές εκκινήσεις των βαρών και των biases σε κάθε στρώση.

Η τιμή best_MSE_train είναι το σφάλμα που παρατηρείται κατά την μάθηση του δικτύου και όσο μικρότερο είναι, τόσο μεγαλύτερη ικανότητα εκπαίδευσης έχει το δίκτυο. Το σφάλμα μάθησης είναι σημαντικό στα νευρωνικά δίκτυα καθώς δείχνει το κατά πόσο το δίκτυο μπορεί να συνδέσει τα δεδομένα εισόδου με τα δεδομένα εξόδου, και άρα είναι στοιχείο της αξιοπιστίας του δικτύου.

Στην περίπτωση των 5 νευρώνων στην πρώτη κρυφή στρώση, το μικρότερο σφάλμα κατά μέσο όρο όπως και η βέλτιστη τιμή σφάλματος παρατηρείται με την αρχιτεκτονική 3-5-20-3. Αυτό σημαίνει ότι για τα δοθέντα στοιχεία εισόδου και εξόδου, η αρχιτεκτονική 3-5-20-3 καταφέρνει να λύσει με μεγαλύτερη επιτυχία τη μη-γραμμική εξίσωση σύνδεσης εισόδων και εξόδων.

Η τιμή $best_MSE_test$ είναι το σφάλμα που παρατηρείται κατά τη γενίκευση (generalization) του δικτύου, δηλαδή, μετά την εκμάθηση, το δίκτυο ελέγχεται σε ξεχωριστό dataset (που προέρχεται από το αρχικό σετ δεδομένων) για το κατά πόσο μπορεί να συνδέσει ικανοποιητικά δεδομένα που δεν χρησιμοποιήθηκαν κατά τη διάρκεια της μάθησης. Με λίγα λόγια, το σφάλμα του testing ελέγχει αν το δίκτυο είναι αποτελεσματικό σε δεδομένα που δεν ξέρει (αλλά ανήκουν στην ίδια συνάρτηση πιθανότητας με τα δεδομένα της μάθησης) ή αν έχει δημιουργηθεί overfitting, δηλαδή υπερβολική προσήλωση στα γνωστά δεδομένα, τέτοια που οδηγεί στο να χαθεί η μεγάλη εικόνα.

Στην περίπτωση των 5 νευρώνων στην πρώτη κρυφή στρώση, το μικρότερο σφάλμα testing κατά μέσο όρο και η βέλτιστη τιμή σφάλματος παρατηρείται με την αρχιτεκτονική 3-5-20-3. Αυτό σημαίνει ότι η εν λόγω αρχιτεκτονική, συγκριτικά με τις υπόλοιπες, έχει τη μεγαλύτερη δυνατότητα γενίκευσης.

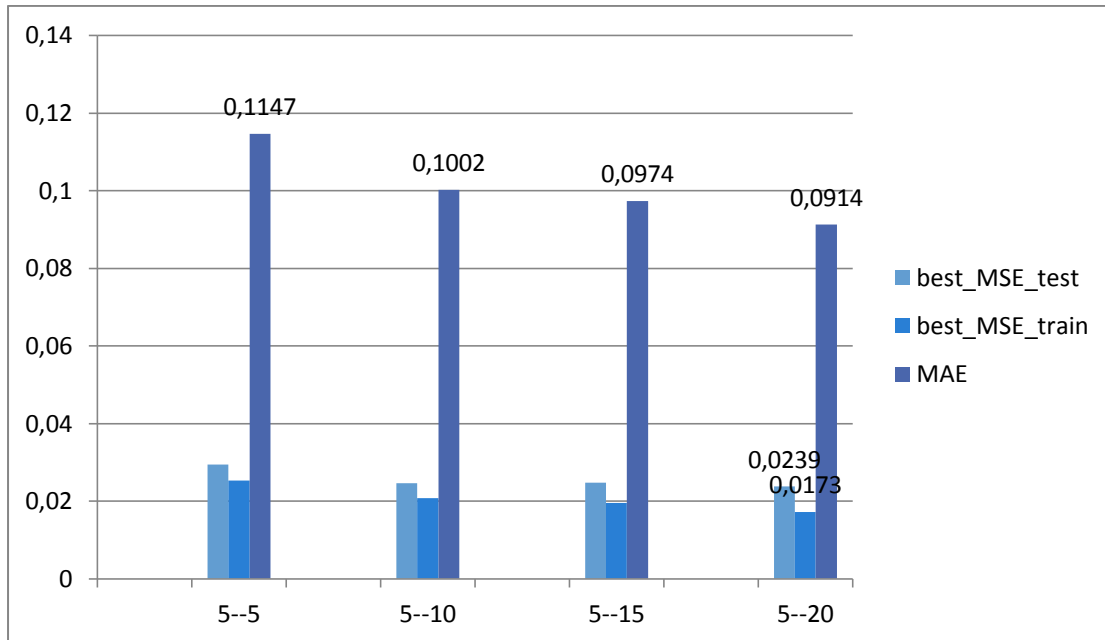
Η τιμή της γραμμικής σύγκλισης/παλινδρόμησης (regression- R^2) όλων των dataset (total) σε όλα τα δίκτυα που δοκιμάστηκαν είναι πάνω από 99,9% γεγονός που ενισχύει την αξιοπιστία των εν λόγω δικτύων.

Το validation dataset χρησιμοποιείται κατά τη διάρκεια της μάθησης, προφυλάσσει με τη μέθοδο early stopping το δίκτυο από overfitting και χρησιμοποιείται για να επιλεγθούν οι κατάλληλες τιμές των βαρών μεταξύ πολλών υποψήφιας τιμών. Το $best_MSE_val$ δεν παρουσιάζει αναλυτικό ενδιαφέρον παρά μόνο σε σύγκριση με τα $best_MSE_test$ και $best_MSE_train$, δηλαδή αν είναι της ίδιας τάξης μεγέθους.

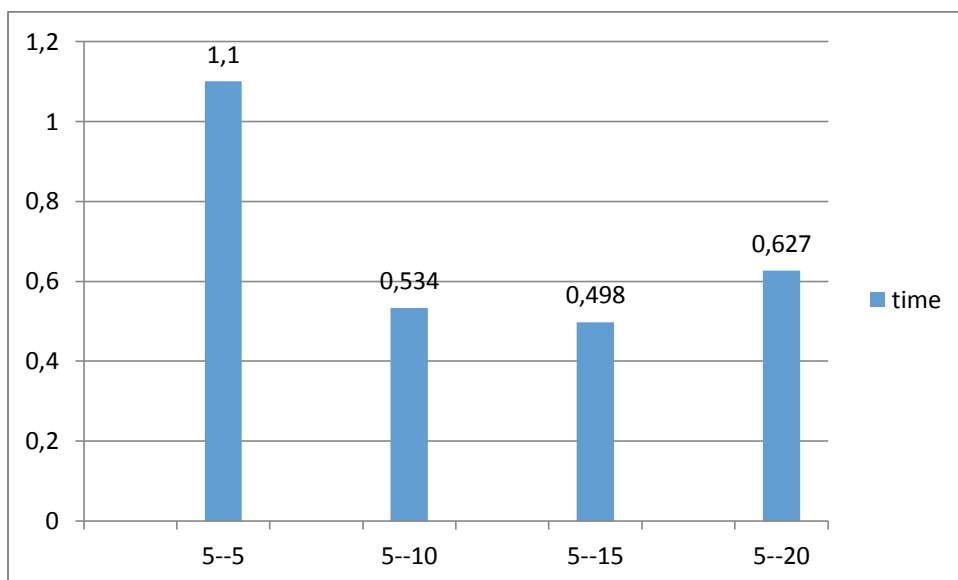
Ο χρόνος (time) δείχνει το υπολογιστικό κόστος της χρήσης των νευρωνικών δικτύων, που είναι μικρό όπως φαίνεται από τους πίνακες, και βέβαια μικρότερο από τον χρόνο τρεξίματος προσομοιώσεων μοριακής δυναμικής. Ακόμα, είναι πολλές τάξεις μεγέθους μικρότερο από προσομοιώσεις DFT. Όπως προαναφέρθηκε, δεν μπορούν να εξαχθούν ασφαλή συμπεράσματα για την επίδραση του αριθμού των νευρώνων στο υπολογιστικό κόστος των υπό μελέτη νευρωνικών δικτύων λόγω της τυχαιότητας που οφείλεται στις τυχαίες αρχικές τιμές των βαρών κάθε νευρώνα και του bias κάθε στρώσης.

Το Mean Absolute Percentage Error (MAPE) είναι το μέσο απόλυτο ποσοστιαίο σφάλμα των νευρωνικών δικτύων και κρίνει το κατά πόσο τα δεδομένα που εξάγονται από το δίκτυο είναι χρησιμοποιήσιμα για άλλους υπολογισμούς. Στην περίπτωσή μας, σκοπός είναι τα δεδομένα να χρησιμοποιηθούν ως αρχικές θέσεις για προσομοιώσεις μοριακής δυναμικής μειωμένου χρόνου. Ένα σύνηθες κριτήριο για το MAPE είναι να έχει τιμή κάτω από 10% για να κριθούν ικανοποιητικά τα δεδομένα εξόδου.

Στην περίπτωση που η πρώτη κρυφή στρώση έχει 5 νευρώνες, παρατηρείται ότι η πολυπλοκότητα του δικτύου δεν καταφέρνει να παρέχει συστηματικά τιμές του MAPE κάτω από 10%. Όμως είναι αρκετά κοντά στο όριο αυτό όπως φαίνεται από τις βέλτιστες τιμές του σφάλματος στις αρχιτεκτονικές 3-5-15-3 και 3-5-20-3. Όπως θα φανεί αργότερα, όσο η πολυπλοκότητα του δικτύου μεγαλώνει με την προσθήκη επιπλέον νευρώνων, το δίκτυο γίνεται πιο αποτελεσματικό και το MAPE μειώνεται περαιτέρω και συστηματικά.



Σχήμα 38 : Σύγκριση των βέλτιστων τιμών των αρχιτεκτονικών 3-5-X-3



Σχήμα 39 : Σύγκριση των βέλτιστων τιμών χρόνου των αρχιτεκτονικών 3-5-X-3

5.2.2 3-10-X-3

	Μέσος όρος 10 δοκιμών			
Δίκτυο	3--10--5--3	3--10--10--3	3--10--15--3	3--10--20--3
best_MSE_val	0,030567806	0,030274803	0,028384164	0,031292435
best_MSE_test	0,03297182	0,033219935	0,031723673	0,02741024
best_MSE_train	0,025657788	0,024310225	0,019593643	0,018135028
regression total	0,999688	0,999701	0,999747	0,99976
time (s)	2,7094	2,0466	1,1928	1,5148
MAPE(%)	11,43707242	11,02572189	10,21445242	9,795740978

Πίνακας 3

	Overall best 10 δοκιμών			
Δίκτυο	3--10--5--3	3--10--10--3	3--10--15--3	3--10--20--3
best_MSE_val	0,02392924	0,024218519	0,024195526	0,025438112
best_MSE_test	0,024711319	0,022608439	0,025814536	0,021639796
best_MSE_train	0,019377126	0,017660146	0,016662507	0,016554032
regression total	0,99975	0,99976	0,99978	0,99978
time	0,458	0,871	0,751	0,86
MAPE(%)	9,841277231	9,367532155	9,209296627	9,133779264

Πίνακας 4

Στην περίπτωση των 10 νευρώνων στην πρώτη κρυφή στρώση, παρατηρείται ότι και πάλι η βέλτιστη αρχιτεκτονική είναι η 3-10-20-3 καθώς έχει τις μικρότερες τιμές σφάλματος training, σφάλματος testing, MAPE, και τις μέγιστες τιμές regression, κατά μέσο όρο και κατά βέλτιστη τιμή σε σύγκριση με τις υπόλοιπες αρχιτεκτονικές.

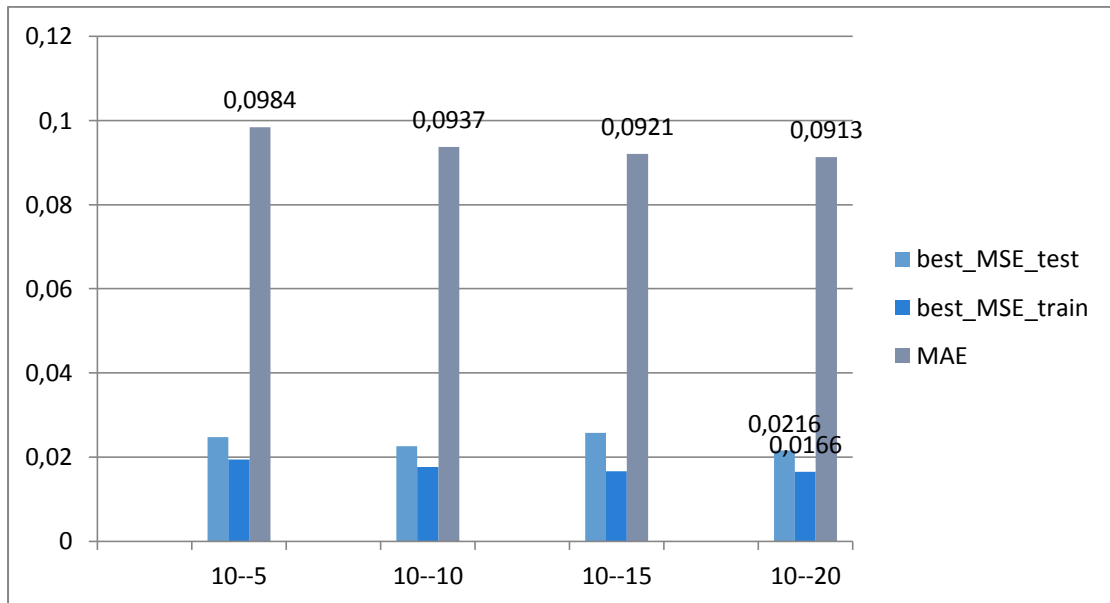
Το σφάλμα training πέφτει κατά μέσο όρο κάτω από 0,02 στην τιμή του 0,0181 και συγκριτικά με το καλύτερο των αρχιτεκτονικών 3-5-X-3 είναι μικρότερο.

Το σφάλμα testing πέφτει επίσης κατά μέσο όρο σε σύγκριση με τις αρχιτεκτονικές 3-5-X-3, από 0,0296 σε 0,0216 και το regression αυξάνεται. Ακόμα, η αρχιτεκτονική 3-10-20-3 παρουσιάζει το καλύτερο MSE στο testing dataset κατά μέσο όρο από όλους τους (συνολικά 16) εξεταζόμενους συνδυασμούς.

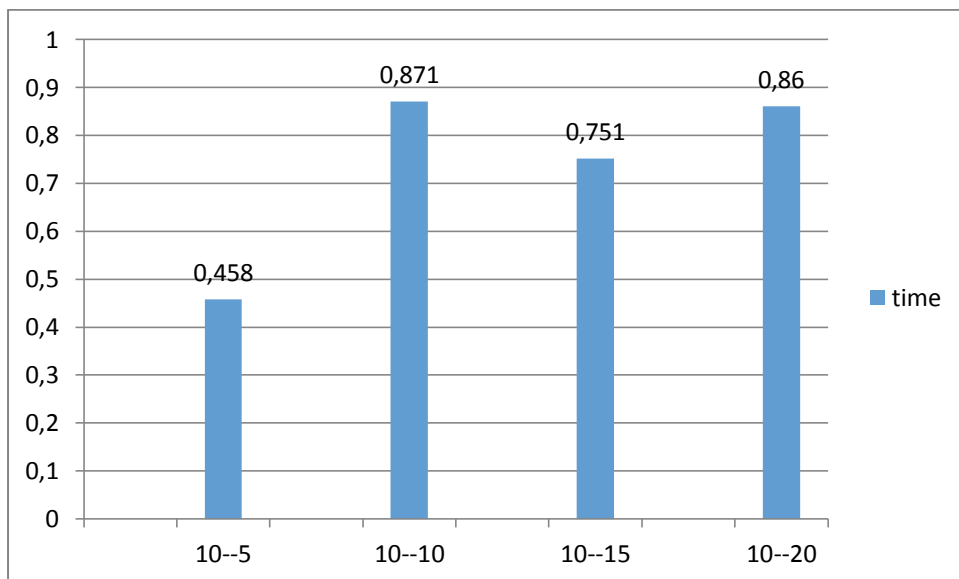
Ο χρόνος παραμένει αρκετά κάτω του ορίου των 8,5 δευτερολέπτων της προσομοίωσης μοριακής δυναμικής που πραγματοποιήθηκε στο lammps.

Το πιο σημαντικό στοιχείο των παραπάνω πινάκων είναι ότι η αρχιτεκτονική 3-10-20-3 καταφέρνει συστηματικά να περάσει το όριο του 10% στο mean absolute percentage error. Οι υπόλοιπες αν και δεν το καταφέρνουν, είναι κοντά καθώς οι βέλτιστες (ακραίες) τιμές τους έχουν κάτω από 10% MAPE.

Τα παραπάνω σημαίνουν ότι η προστιθέμενη πολυπλοκότητα που απέκτησε το δίκτυο συνεισέφερε στα βελτιωμένα αποτελέσματα που παρατηρούνται.



Σχήμα 40 : Σύγκριση των βέλτιστων τιμών των αρχιτεκτονικών 3-10-X-3



Σχήμα 41 : Σύγκριση των βέλτιστων τιμών χρόνου των αρχιτεκτονικών 3-10-X-3

5.2.3 3-15-X-3

	Μέσος όρος 10 δοκιμών			
Δίκτυο	3--15--5--3	3--15--10--3	3--15--15--3	3--15--20--3
best_MSE_val	0,028502539	0,033239724	0,031060794	0,03033365
best_MSE_test	0,02960834	0,030807325	0,032476562	0,034040446
best_MSE_train	0,02171783	0,021268672	0,016672313	0,015312077
regression total	0,999731	0,999724	0,999764	0,999776
time (s)	2,5922	1,3087	1,2829	1,4073
MAPE(%)	10,50709969	10,5809156	9,647607812	9,207114842

Πίνακας 5

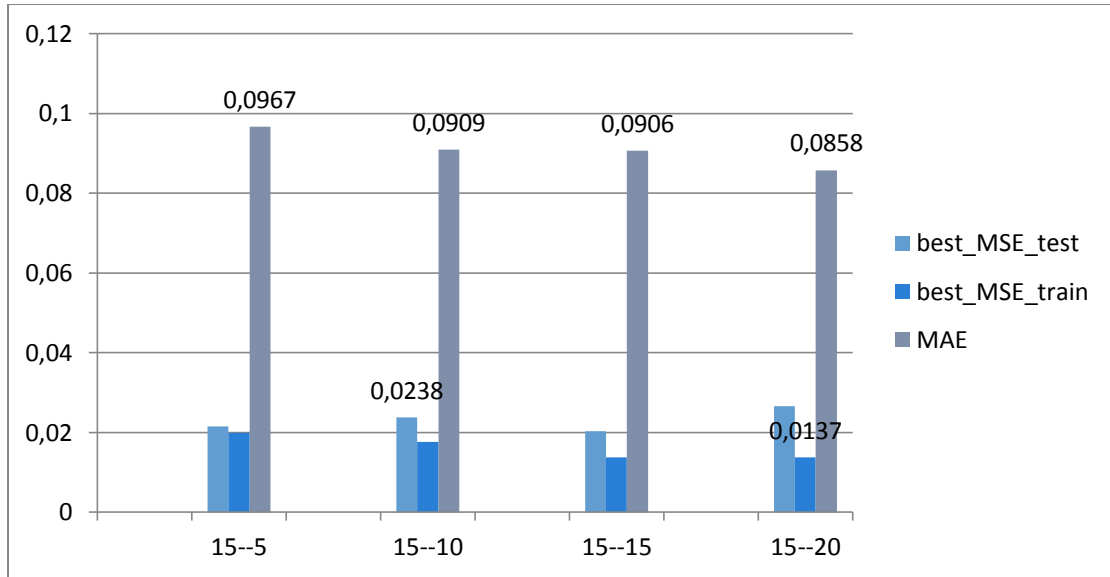
	Overall best 10 δοκιμών			
Δίκτυο	3--15--5--3	3--15--10--3	3--15--15--3	3--15--20--3
best_MSE_val	0,022096803	0,022828096	0,025805207	0,024393958
best_MSE_test	0,021514204	0,023810734	0,020350559	0,026590348
best_MSE_train	0,019874766	0,017635086	0,013740765	0,013711771
regression total	0,99975	0,99977	0,99978	0,99979
time	0,556	0,607	0,796	1,013
MAPE(%)	9,667479954	9,091124812	9,064059368	8,576150855

Πίνακας 6

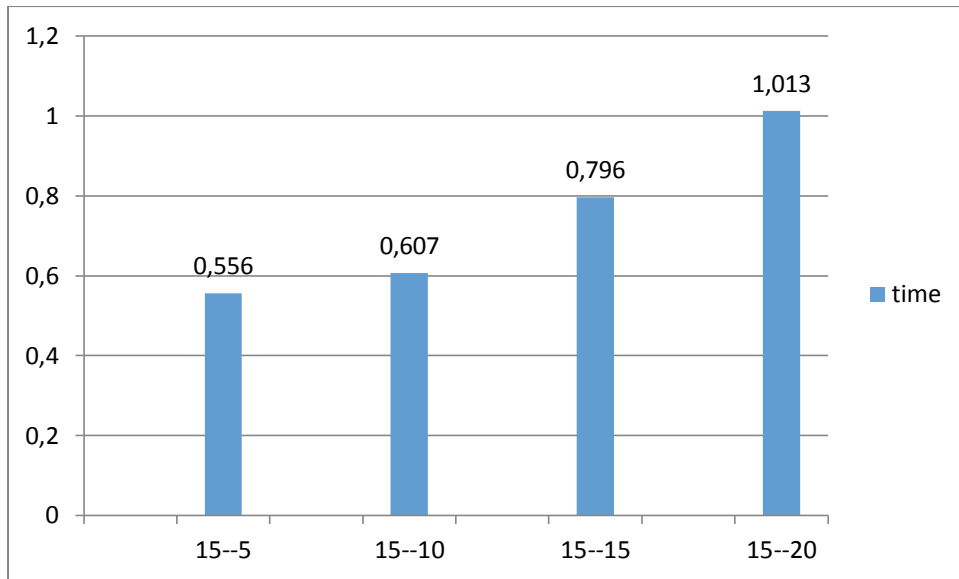
Στην περίπτωση των 15 νευρώνων στην πρώτη κρυφή στρώση, παρατηρείται ότι πλέον δεν υπάρχει συγκεκριμένη αρχιτεκτονική με τις καλύτερες τιμές κατά μέσο όρο σε όλες τις ουσιώδεις κατηγορίες, καθώς η αρχιτεκτονική 3-15-20-3 έχει το μεγαλύτερο regression και τα μικρότερα MAPE και train MSE κατά μέσο όρο, αλλά η αρχιτεκτονική 3-15-5-3 έχει μεγαλύτερη δυνατότητα γενίκευσης καθώς έχει μικρότερο testing MSE.

Ακόμα παρατηρείται ότι το testing MSE στην περίπτωση του 3-15-15-3 και 3-15-20-3 έχει αυξηθεί σε σχέση με τους προηγούμενους υπολογισμούς, ενώ το testing MSE στα 3-15-5-3 και 3-15-10-3 συνεχίζει να μειώνεται. Αυτό σημαίνει ότι σιγά σιγά το δίκτυο μειώνει την ικανότητα γενίκευσής του με την προσθήκη επιπλέον νευρώνων παράλο που το σφάλμα του training μειώνεται και έτσι έχουμε το πρώτο δείγμα overfitting σε πολύ μικρή κλίμακα(καθώς στο πραγματικό overfitting τα σφάλματα του training και του testing διαφέρουν κατά αρκετές τάξεις μεγέθους).

Όσον αφορά το MAPE, παρατηρείται ότι υπάρχουν πλέον δύο αρχιτεκτονικές, οι 3-15-15-3 και η 3-15-20-3, που είναι συστηματικά κάτω από το όριο του 10%, γεγονός που δείχνει ότι η προσθήκη επιπλέον νευρώνων, αν και έβλαψε λίγο την δυνατότητα γενίκευσης του δικτύου, εν τέλει βοηθά στην εξαγωγή καλύτερων αποτελεσμάτων για τον σκοπό που μας ενδιαφέρει. Ακόμα, οι βέλτιστες τιμές όλων των δικτύων είναι κάτω του 10%, κάτι που επιβεβαιώνει ότι, σταθερά, η αύξηση των νευρώνων μειώνει το MAPE και τα δίκτυα αποκτούν σιγά σιγά την απαιτούμενη πολυπλοκότητα για να δώσουν την ζητούμενη ακρίβεια. Ο χρόνος παραμένει σταθερά κάτω από το όριο των 8,5 δευτερολέπτων.



Σχήμα 42 : Σύγκριση των βέλτιστων τιμών των αρχιτεκτονικών 3—15---X—3



Σχήμα 43 : Σύγκριση των βέλτιστων τιμών χρόνου των αρχιτεκτονικών 3—15—X—3

5.2.4 3-20-X-3

Δίκτυο	Μέσος όρος 10 δοκιμών			
	3--20--5--3	3--20--10--3	3--20--15--3	3--20--20--3
best_MSE_val	0,030512922	0,030560964	0,038152939	0,043335868
best_MSE_test	0,032633691	0,031942619	0,033837643	0,041143388
best_MSE_train	0,021265434	0,016796568	0,014568444	0,013560984
regression total	0,999728	0,999769	0,99977	0,999762
time (s)	1,9992	1,787	1,413	1,4471
MAPE(%)	10,48965264	9,352600861	9,301508425	9,178592707

Πίνακας 7

Η αρχιτεκτονική 3–20–20–3 παρουσιάζει το καλύτερο MSE training κατά μέσο όρο συνολικά(από το σύνολο των 16 μελετούμενων περιπτώσεων). Ακόμα, παρουσιάζει το ελάχιστο MAPE από το σύνολο των συνδυασμών.

	Overall best 10 δοκιμών			
Δίκτυο	3--20--5--3	3--20--10--3	3--20--15--3	3--20--20--3
best_MSE_val	0,025939679	0,025328627	0,024066605	0,032064539
best_MSE_test	0,023712525	0,024555743	0,028596327	0,033493278
best_MSE_train	0,018444118	0,014703173	0,012196345	0,010810352
regression total	0,99976	0,99979	0,9998	0,99978
time	0,675	0,849	0,998	1,229
MAPE(%)	9,690621833	8,5190726	7,992544468	8,508880238

Πίνακας 8

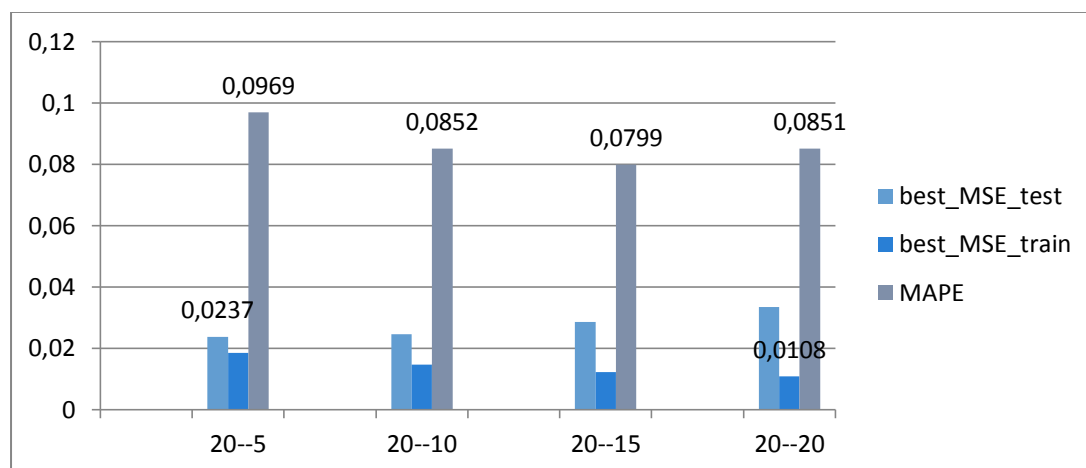
Στην περίπτωση όπου η πρώτη κρυφή στρώση αποτελείται από 20 νευρώνες, δεν υπάρχει πλέον κυρίαρχη αρχιτεκτονική. Το καλύτερο σφάλμα training κατά μέσο όρο και το καλύτερο MAPE μπορεί να παρατηρείται στην αρχιτεκτονική 3-20-20-3, αλλά το testing σφάλμα συνεχίζει να αυξάνεται με τον ίδιο ρυθμό και το regression αρχίζει να μειώνεται λόγω των αποκλίσεων των dataset του testing και του validation.

Το καλύτερο σφάλμα testing κατά μέσο όρο παρατηρείται στην αρχιτεκτονική 3-20-10-3, όμως όλα τα σφάλματα testing έχουν αυξηθεί κατά μέσο όρο.

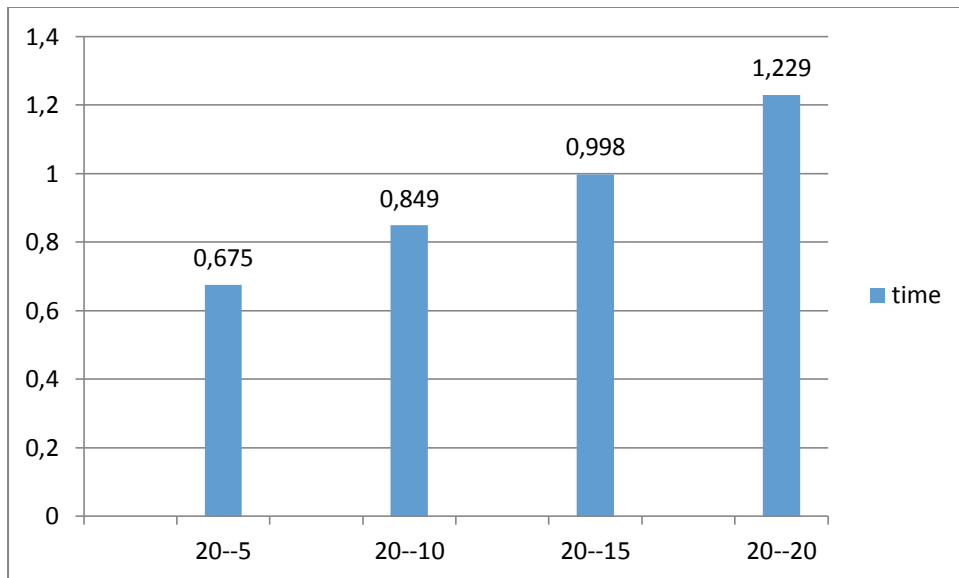
Ο χρόνος παραμένει σταθερά πολύ κάτω από το όριο των 8,5 δευτερολέπτων.

Το καλύτερο regression πλέον παρατηρείται στη δομή 3-20-15-3, στην οποία παρατηρείται και το καλύτερο MAPE κατά βέλτιστη τιμή.

Τέλος, παρατηρείται ότι τρεις αρχιτεκτονικές έχουν πλέον έρθει κάτω από το όριο του 10%, αν και με κόστος στη γενίκευση του δικτύου, και ότι όλες οι βέλτιστες τιμές του MAPE είναι κάτω του 10%, με την τιμή του δικτύου 3-20-15-3 να πέφτει και κάτω από το 8%.



Σχήμα 44 : Σύγκριση των βέλτιστων τιμών των αρχιτεκτονικών 3–20--X–3



Σχήμα 45 : Σύγκριση των βέλτιστων τιμών χρόνου των αρχιτεκτονικών 3—20—X—3

Συμπερασματικά, η αύξηση των νευρώνων γενικά οδηγεί σε καλύτερη μάθηση του δικτύου λόγω της προστιδόμενης πολυπλοκότητας, όμως μετά από κάποιο σημείο η τάση αυτή, αν και εξακολουθεί να υπάρχει, δίνει αρνητικά χαρακτηριστικά στο δίκτυο, όπως μειωμένη ικανότητα γενίκευσης και μειωμένη σύγκλιση.

Η μέθοδος trial and error που παρουσιάστηκε οδηγεί σε ορισμένα χρήσιμα συμπεράσματα αλλά και σε αβεβαιότητα σχετικά με τη βέλτιστη αρχιτεκτονική του δικτύου, επομένως το παραπάνω πρόβλημα πρέπει να μελετηθεί συστηματικά.

Επειδή οι δύο μεταβλητές που μελετήθηκαν (αριθμοί νευρώνων σε κάθε κρυφή στρώση) είναι λίγες για την απόκτηση ολοκληρωμένης εικόνας σχετικά με τα εκάστοτε δίκτυα, καθώς τα ποσοστά διαχωρισμού του dataset σε training, testing, και validation είναι σημαντικά για την απόδοση του δικτύου, χρησιμοποιήθηκαν γενετικοί αλγόριθμοι για την ολοκληρωμένη προσέγγιση του προβλήματος. Έτσι μπορούν να βρεθούν συνδυασμοί νευρώνων που δεν μπορούσαν να μελετηθούν με trial and error, όπως επίσης μπορούν να προστεθούν παραπάνω παράμετροι στο πρόβλημα βελτιστοποίησης που αντιμετωπίζεται.

5.3 Αποτελέσματα γενετικών αλγορίθμων

Η συστηματική βελτιστοποίηση της αρχιτεκτονικής των νευρωνικών δικτύων έγινε με τη χρήση γενετικών αλγορίθμων για την εύρεση των βέλτιστων παραμέτρων του νευρωνικού δικτύου, συμπεριλαμβανομένων αλλά όχι μόνο, των νευρώνων στις κρυφές στρώσεις. Αναλυτικά οι παράμετροι που επιλέχθηκαν προς βελτιστοποίηση είναι οι εξής :

- Νευρώνες στην πρώτη κρυφή στρώση
- Νευρώνες στη δεύτερη κρυφή στρώση
- Ποσοστό του dataset για training
- Ποσοστό του dataset για testing
- Αριθμός μέγιστων validation σφαλμάτων

Σε αυτό το πλαίσιο πραγματοποιήθηκαν δοκιμές για διαφορετικές τιμές του crossover fraction (0,7-0,8-0,9), διαφορετικές τιμές του elite count (2-3-4), και διαφορετικά population sizes (20 και 50). Ακόμη, για την αντικειμενική συνάρτηση χρησιμοποιήθηκε ως κριτήριο το MSE του training σε όλες τις παραπάνω περιπτώσεις, το MAPE στις 2 καλύτερες περιπτώσεις του MSE-training, και συνδυασμός του MSE και του MAPE για την αποδοτικότερη από τις παραπάνω περιπτώσεις.

Οι παράμετροι που κρατήθηκαν σταθερές ήταν :

- migration direction : forward
- migration fraction : 0,2
- generations : 500
- stall generations : 50
- function tolerance : 1e-6
- selection function : stochastic uniform
- mutation function : uniform, 0,01
- fitness scaling : rank

Περιορισμοί :

- Αριθμός νευρώνων μεταξύ 5 και 20
- Ποσοστό training μεταξύ 60% και 80%
- Ποσοστό testing μεταξύ 10% και 20%
- αριθμός μέγιστων σφαλμάτων validation μεταξύ 6 και 20

Τα αποτελέσματα για το MSE του training και population size 20 φαίνονται στους παρακάτω πίνακες :

crossover fraction	0,7		
population size	20		
elite count	2	3	4
generations	60/500	56/500	78/500
layersize x1	19	17	14
layersize x2	18	15	14
training percent	73,732	73,519	60,129
testing percent	10,842	10,968	18,955
validation percent	15,426	15,513	20,916
val. Max fail	9	7	8
obj.function value	0,022654	0,020353	0,021158

Πίνακας 9

Με βάση τους γενετικούς αλγόριθμους, για population size 20 και crossover fraction 0,7, η καλύτερη τιμή του elite count είναι 3, με MSE για το training 0,020353 και αρχιτεκτονική δικτύου 3-17-15-3, 73,5% του dataset για training, 11% του dataset για testing και 7 max validation fails.

crossover fraction	0,8		
population size	20		
elite count	2	3	4
generations	54/500	54/500	74/500
layersize x1	15	13	14
layersize x2	17	17	15
training percent	64,218	63,262	76,674
testing percent	13,304	16,454	13,111
validation percent	22,478	20,284	10,215
val. Max fail	12	12	17
obj.function value	0,019868	0,021638	0,023796

Πίνακας 10

Για population size 20 και crossover fraction 0,8, η καλύτερη τιμή του elite count είναι 2, με MSE για το training 0,019868 και αρχιτεκτονική δικτύου 3-15-17-3, 64,2% του dataset για training, 13,3% του dataset για testing και 12 max validation fails. Προφανώς οι αριθμοί διαχωρισμού του dataset θα πρέπει να στρογγυλοποιηθούν με βάση τον υπάρχον αριθμό (512) των στοιχείων του dataset.

crossover fraction	0,9		
population size	20		
elite count	2	3	4
generations	69/500	56/500	66/500
layersize x1	14	11	9
layersize x2	18	17	18
training percent	63,776	78,548	62,385
testing percent	13,139	14,968	18,669
validation percent	23,085	6,484	18,946
val. Max fail	10	12	18
obj.function value	0,024921	0,019913	0,020458

Πίνακας 11

Για population size 20 και crossover fraction 0,9 , η καλύτερη τιμή του elite count είναι 3, με MSE για το training 0,019913 και αρχιτεκτονική δικτύου 3-11-17-3, 78,5% του dataset για training, 15% του dataset για testing και 12 max validation fails.

Τα αποτελέσματα για το MSE του training και για population size 50 φαίνονται αναλυτικά στους παρακάτω πίνακες :

crossover fraction	0,7		
population size	50		
elite count	2	3	4
generations	52/500	85/500	51/500
layersize x1	16	14	13
layersize x2	13	16	19
training percent	75,854	75,187	72,601
testing percent	18,52	10,109	15,029
validation percent	5,626	14,704	12,37
val. Max fail	8	13	7
obj.function value	0,020261	0,018054	0,022482

Πίνακας 12

Για population size 50 και crossover fraction 0,7 , η καλύτερη τιμή του elite count είναι 3, με MSE για το training 0,018054 και αρχιτεκτονική δικτύου 3-14-16-3, 75,2% του dataset για training, 10,1% του dataset για testing και 13 max validation fails.

Η τιμή 0,018054 είναι και η καλύτερη τιμή που έδωσαν οι γενετικοί αλγόριθμοι για το training σφάλμα, από τις συνολικά 18 που μελετήθηκαν.

crossover fraction	0,8		
population size	50		
elite count	2	3	4
generations	58/500	83/500	62/500
layersize x1	20	14	17
layersize x2	20	16	16
training percent	61,674	71,079	64,696
testing percent	10,439	14,119	13,989
validation percent	27,887	14,802	21,315
val. Max fail	20	9	18
obj.function value	0,023438	0,019444	0,021888

Πίνακας 13

Για population size 50 και crossover fraction 0,8 , η καλύτερη τιμή του elite count είναι 3, με MSE για το training 0,019444 και αρχιτεκτονική δικτύου 3-14-16-3, 71% του dataset για training, 14,1% του dataset για testing και 9 max validation fails.

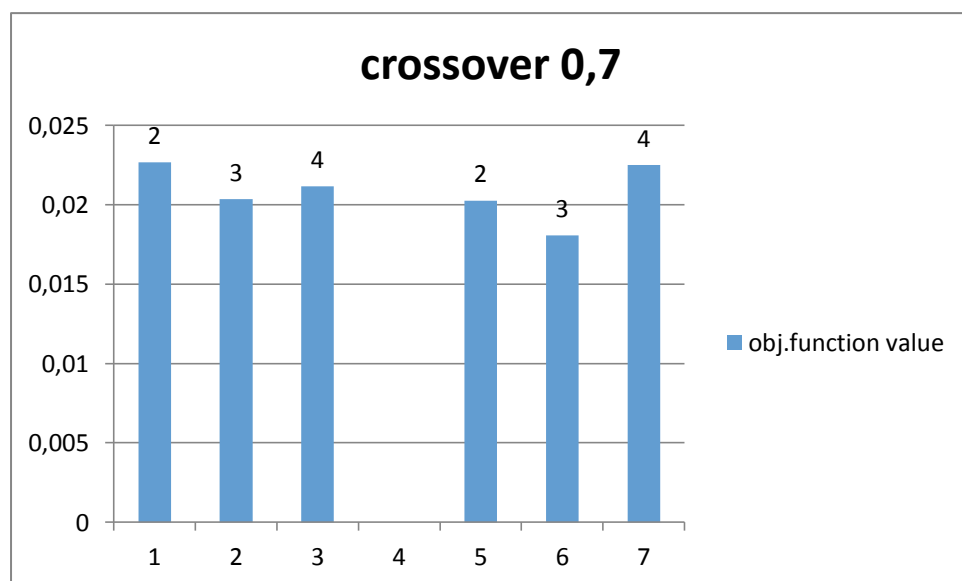
Παρατηρείται ότι για δεύτερη συνεχόμενη φορά η καλύτερη αρχιτεκτονική είναι 3-14-16-3 και τα ποσοστά διαχωρισμού του dataset αρκετά κοντινά.

crossover fraction	0,9		
population size	50		
elite count	2	3	4
generations	54/500	112/500	51/500
layersize x1	18	16	16
layersize x2	17	14	17
training percent	73,077	71,773	71,092
testing percent	13,657	18,392	19,873
validation percent	13,266	9,835	9,035
val. Max fail	10	8	13
obj.function value	0,026409	0,02595	0,018092

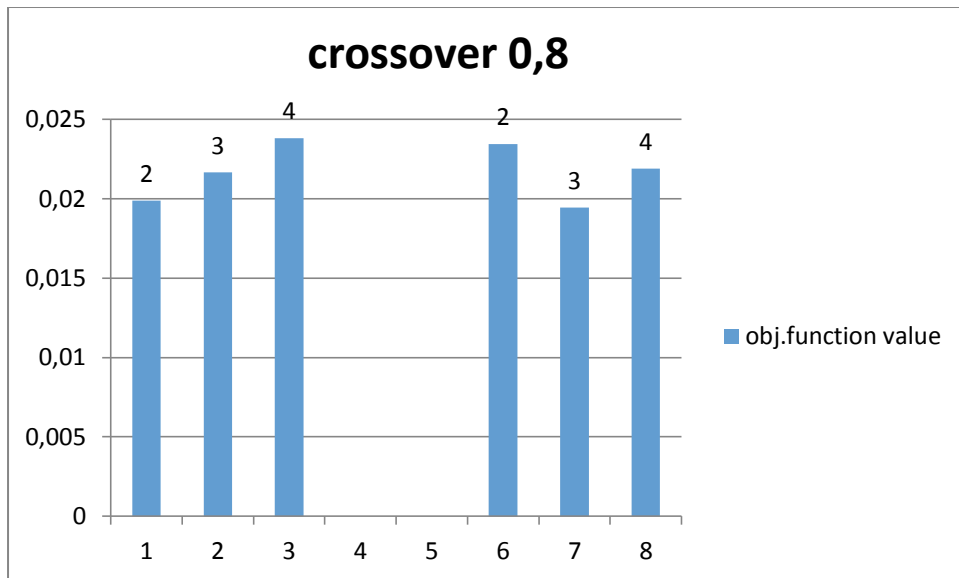
Πίνακας 14

Για population size 50 και crossover fraction 0,9 , η καλύτερη τιμή του elite count είναι 4, με MSE για το training 0,018092 και αρχιτεκτονική δικτύου 3-16-17-3, 71% του dataset για training, 19,8% του dataset για testing και 13 max validation fails.

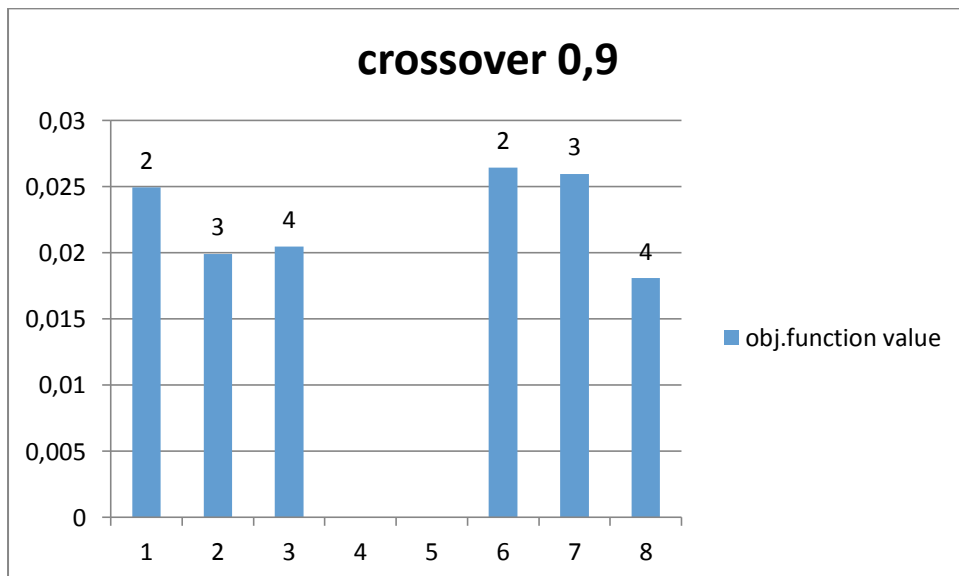
Η τιμή του σφάλματος training 0,018092 είναι η δεύτερη καλύτερη του συνόλου των υπολογισμών, με πολύ μικρή διαφορά από την πρώτη (0,018054).



Σχήμα 46 : Σύγκριση των τιμών της αντικειμενικής συνάρτησης για crossover fraction 0,7 , elite count 2,3 και 4, και population size 20 (αριστερά) και 50(δεξιά)



Σχήμα 47: Σύγκριση των τιμών της αντικειμενικής συνάρτησης για crossover fraction 0,8 , elite count 2,3 και 4, και population size 20 (αριστερά) και 50(δεξιά)



Σχήμα 48 : Σύγκριση των τιμών της αντικειμενικής συνάρτησης για crossover fraction 0,9 , elite count 2,3 και 4, και population size 20 (αριστερά) και 50(δεξιά)

Παρατηρείται ότι υπάρχει αύξηση στο MSE του training σε σχέση με το trial and error, καθώς η βέλτιστη τιμή του training MSE στο trial and error ήταν (ενδεικτικά) 0,010810352. Αυτό όμως δεν σημαίνει ότι το δίκτυο δεν έχει καλύτερη απόδοση. Ο ποσοστιαίος διαχωρισμός του dataset παίζει μεγάλο ρόλο στην απόδοση του δικτύου όπως και ο αριθμός των μέγιστων σφαλμάτων του validation.

Με τον τρόπο αυτό προσπαθήσαμε να βρούμε τις παραμέτρους του δικτύου που βελτιστοποιούν το σφάλμα του training και καταλήξαμε στην τριάδα crossover, elite count, population size : (0,7-3-50). Όμως, επειδή αυτό που μας αφορά σε τελική ανάλυση είναι το MAPE, για τις δύο καλύτερες τριάδες παραμέτρων των γενετικών αλγορίθμων χρησιμοποιήσαμε ως αντικειμενική συνάρτηση το MAPE αντί για το training MSE.

Τα αποτελέσματα για το MAPE για τις τριάδες (0,7-3-50) και (0,9-4-50) φαίνονται παρακάτω :

crossover fraction	0,7	0,9
elite count	3	4
population size	50	50
generations	208/500	500/500
layersize x1	18	15
layersize x2	18	15
training percent	65,671	74,343
testing percent	17,449	14,034
validation percent	16,88	11,623
val. Max fail	15	8
obj.function value	5,312404456	6,041013

Πίνακας 15

Παρατηρείται σημαντική μείωση του MAPE σε σχέση με τη διαδικασία trial and error. Ενδεικτικά, σε ένα παρόμοιο δίκτυο του trial and error, με πολύ παρόμοια αρχιτεκτονική με το δεύτερο αποτέλεσμα του γενετικού αλγορίθμου, το 3-15-15-3 που είχε ποσοστό training 75%, ποσοστό testing 10% και max validation fails 6, το καλύτερο MAPE που επιτεύχθηκε ήταν 9,064059 και ο μέσος όρος ήταν 9,64761. Δηλαδή προσθέτοντας 5% του dataset από το validation dataset στο testing, και αυξάνοντας κατά 2 τα μέγιστα σφάλματα κατά το validation, μειώνεται το MAPE πάνω από 3%. Επίσης αξ σημειωθεί ότι στη δεύτερη περίπτωση 500 γενιές (πολλές ώρες τρεξίματος) δεν ήταν αρκετές για να σταματήσει ο γενετικός αλγόριθμος λόγω σύγκλισης, κάτι που σημαίνει ότι είναι πιθανό να υπάρχει και καλύτερη λύση.

Ακόμα παρατηρείται ότι οι βέλτιστοι νευρώνες δεν είναι απαραίτητα και οι περισσότεροι.

Συγκριτικά με το trial and error βλέπουμε ότι με την αλλαγή ορισμένων παραμέτρων αυτοματοποιημένα μπορούμε να βελτιώσουμε κατά πολύ την ποιότητα νευρωνικών δικτύων που με πρώτη ματιά είναι στο όριο μεταξύ επιτυχούς και αποτυχούς πρόβλεψης. Με μια ολοκληρωμένη διαδικασία όπως οι γενετικοί αλγόριθμοι που ψάχνει ολικά και όχι τοπικά βέλτιστα είναι δυνατή η εύρεση βέλτιστων αρχιτεκτονικών νευρωνικών δικτύων με ευρείς περιορισμούς παραμέτρων, παραμέτρων οι οποίες δεν μπορούν να προσδιοριστούν ικανοποιητικά με την εμπειρία.

Παρακάτω θα συγκριθεί η καλύτερη περίπτωση σφάλματος training του trial and error με την καλύτερη περίπτωση σφάλματος training των γενετικών αλγορίθμων. Για να γίνει αυτό τρέχουμε με τη Matlab μία φορά ένα νευρωνικό δίκτυο με τις προτεινόμενες από τους γενετικούς αλγόριθμους παραμέτρους.

Δίκτυο	3--14--16--3	3--20--20--3
best_MSE_val	0,026135813	0,03206
best_MSE_test	0,030935009	0,03869
best_MSE_train	0,015687103	0,01377
regression total	0,99978	0,99978
time (s)	2,827	1,45100
MAPE(%)	9,120865	8,50888

Πίνακας 16

Παρατηρείται ότι το δίκτυο 3-20-20-3 υπερέρχει στα βασικά χαρακτηριστικά όπως το MAPE, ο χρόνος, και το σφάλμα του training, κάτι που είναι λογικό καθώς έχει περισσότερες νευρώνες.

Αντίθετα, το δίκτυο 3-14-16-3 υπερέρχει στο σφάλμα του testing και ισοβαθεί στο regression. Για να κατανοήσουμε τη βελτίωση του δικτύου με τους γενετικούς αλγόριθμους θα συγκριθεί το βελτιωμένο δίκτυο με το μέσο όρο της κατηγορίας 3-15-15-3 του trial and error.

Δίκτυο	3--14--16--3	3--15--15--3 (μ.ο)
best_MSE_val	0,026135813	0,031060794
best_MSE_test	0,030935009	0,032476562
best_MSE_train	0,015687103	0,016672313
regression total	0,99978	0,999764
time (s)	2,827	1,2829
MAPE(%)	9,120865	9,647607812

Πίνακας 17

Εδώ φαίνεται καθαρά η υπεροχή του βελτιωμένου δικτύου σε σχέση με το trial and error. Το βελτιωμένο δίκτυο έχει μικρότερα όλα τα σφάλματα συμπεριλαμβανομένου του MAPE, και μεγαλύτερο regression. Το μόνο μειονέκτημά του είναι ότι υστερεί σε χρόνο, παραμένοντας όμως κάτω από το όριο των 7 δευτερολέπτων.

Αντίστοιχα θα συγκριθεί η καλύτερη περίπτωση MAPE του trial and error με την καλύτερη περίπτωση MAPE των γενετικών αλγορίθμων. Για να γίνει αυτό τρέχουμε με τη Matlab μία φορά ένα νευρωνικό δίκτυο με τις προτεινόμενες από τους γενετικούς αλγόριθμους παραμέτρους.

Δίκτυο	3-18-18-3	3-20-20-3
best_MSE_val	0,0400227	0,03206
best_MSE_test	0,037687588	0,03869
best_MSE_train	0,0118625	0,01377
regression total	0,99976	0,99978
time (s)	1,809	1,451
MAPE(%)	8,919031	8,508880

Πίνακας 18

Παρατηρείται ότι το καλύτερο συνολικά δίκτυο του trial and error είναι καλύτερο στο regression, το χρόνο, και το MAPE, ενώ το βελτιωμένο δίκτυο είναι καλύτερο στο σφάλμα του training και του testing. Αν όμως συγκρίνουμε το τυχαίο δείγμα του βελτιωμένου νευρωνικού δικτύου με το μέσο όρο του trial and error έχουμε :

Δίκτυο	3-18-18-3	3-20-20-3 (μ.ο.)
best_MSE_val	0,0400227	0,043335868
best_MSE_test	0,037687588	0,041143388
best_MSE_train	0,0118625	0,013560984
regression total	0,99976	0,999762
time (s)	1,809	1,4471
MAPE	8,919031	9,178592707

Πίνακας 19

Εδώ φαίνεται ότι το βελτιωμένο δίκτυο υπερτερεί του μέσου όρου της αρχιτεκτονικής 3-20-20-3 του trial and error σε όλα τα σφάλματα, και υστερεί στο χρόνο και το regression, παρόλο που έχει λιγότερους νευρώνες.

Δηλαδή φαίνεται ότι δεν υπάρχει ακριβώς γραμμική σχέση μεταξύ του αριθμού των νευρώνων και του σφάλματος του training, παρά μόνο αν όλες οι υπόλοιπες παράμετροι διατηρούνται σταθερές.

Σαν μια τελική μελέτη δοκιμάστηκε η παρεμβολή στην αντικειμενική συνάρτηση που βελτιστοποιούν οι γενετικοί αλγόριθμοι του MAE(mean absolute error) και του MSE του training, με συντελεστή 0,5 για το κάθε σφάλμα, εφόσον το MAE και το MSE είναι της ίδιας τάξης μεγέθους. Επιλέχθηκαν πάλι οι δύο καλύτεροι συνδυασμοί crossover fraction, elite count και population size, δηλαδή (0,7-3-50) και (0,9-4-50).

Ο σκοπός της μελέτης αυτής είναι ο έλεγχος της μεταβολής των βελτιστοποιημένων παραμέτρων των νευρωνικών δικτύων σε σχέση με τις αντίστοιχες του βέλτιστου MSE και MAPE.

Τα αποτελέσματα για MAE και MSE με population size 50 φαίνονται παρακάτω :

crossover fraction	0,7	0,9
elite count	3	4
population size	50	50
generations	52/500	73/500
layersize x1	19	18
layersize x2	16	18
training percent	67,051	60,884
testing percent	17,58	17,652
validation percent	15,369	21,464
val. Max fail	14	17
obj.function	0,039759193	0,043986131

value		
MAPE(%)	5,839236611	6,479577
MSE	0,021126019	0,023176

Πίνακας 20

Παρατηρούμε ότι στη δεύτερη περίπτωση οι αριθμοί των νευρώνων έμειναν ίδιοι με την αντίστοιχη περίπτωση βελτιστοποίησης του MAPE ενώ άλλαξαν τα ποσοστά, ενώ στην πρώτη περίπτωση έμειναν ίδιοι οι νευρώνες στη δεύτερη κρυφή στρώση. Τα ποσοστά διαχωρισμού του dataset άλλαξαν κατά πολύ ενώ τα μέγιστα σφάλματα validation από 13 έγιναν 14 στην πρώτη περίπτωση, και από 15 έγιναν 17 στη δεύτερη περίπτωση.

Με σκοπό την επιτάχυνση της προσομοίωσης μοριακής δυναμικής εξήχθησαν οι συντεταγμένες εξόδου των δύο βέλτιστων νευρωνικών δικτύων, δηλαδή των :

crossover fraction	0,7	0,7
elite count	3	3
population size	50	50
layersize x1	18	14
layersize x2	18	16
training percent	65,671	75,187
testing percent	17,449	10,109
validation percent	16,88	14,704
val. max fails	15	13
obj. function value(MAPE-MSE)	5,3124045	0,0180540

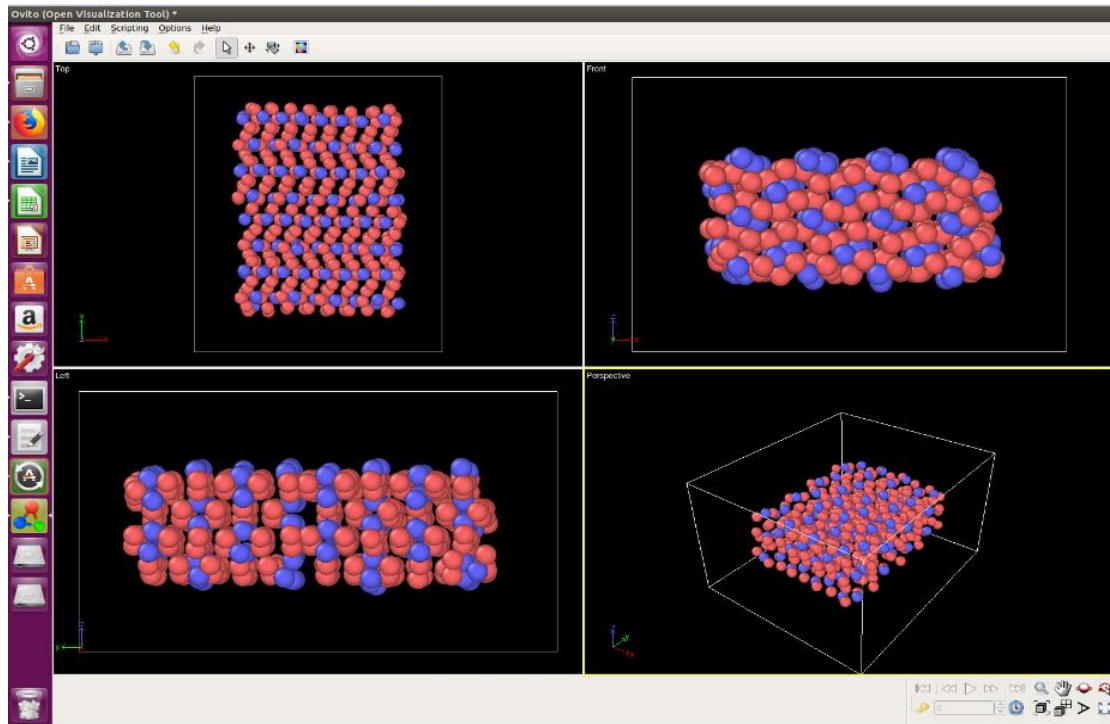
Πίνακας 21

5.4 Αποτελέσματα τελικών προσομοιώσεων

Παρουσιάζονται οι προσομοιώσεις που είχαν συντεταγμένες εισόδου προς γεωμετρική βελτιστοποίηση τις συντεταγμένες εξόδου από τα νευρωνικά δίκτυα, και χωρίζονται σε προσομοιώσεις με βάση το MSE και με βάση το MAPE, όπως εξηγήθηκε στην προηγούμενη ενότητα.

5.4.1.1 Δομή βασισμένη σε νευρωνικά δίκτυα με βάση το MSE

Η αρχική δομή της προσομοίωσης είχε ως δεδομένα εισόδου τις συντεταγμένες που προέβλεψε το νευρωνικό δίκτυο, αφού βελτιστοποιήθηκε ως προς το MSE του training. Η αρχική ενέργεια της δομής ήταν -1728.27576027 eV, η αρχική μέγιστη δύναμη 45.3134 eV/A, και η δομή του συστήματος πριν την προσομοίωση φαίνεται στο παρακάτω σχήμα :



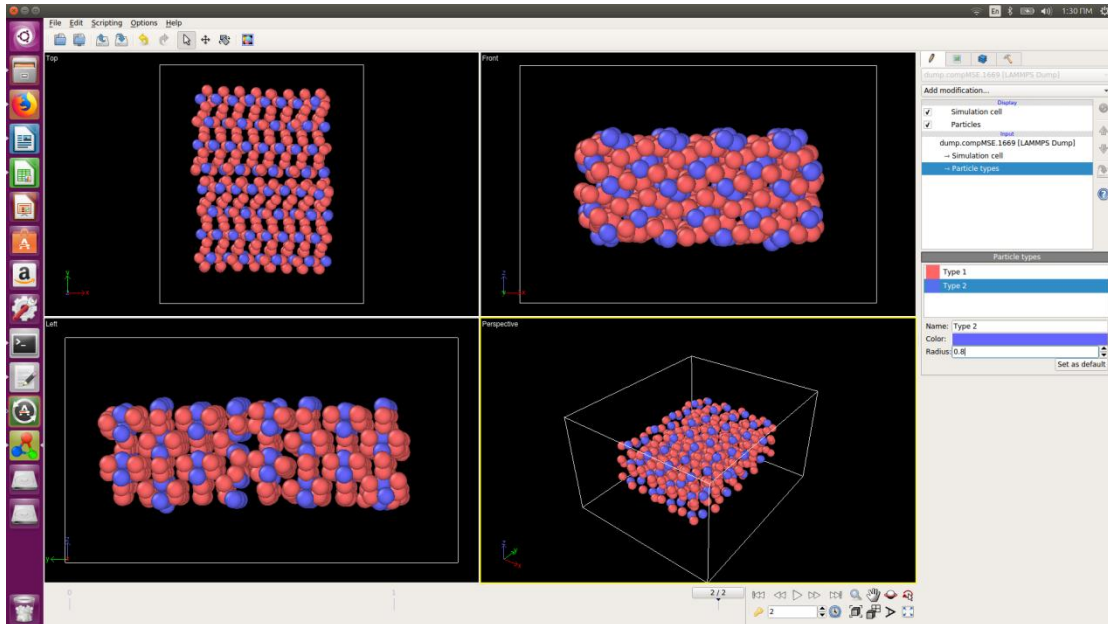
Σχήμα 49 : Αρχική θέση προσομοίωσης 512 ατόμων

Η προσομοίωση διήρκεσε 5.89543 δευτερόλεπτα με 2 πυρήνες 1.8GHz. Χρειάστηκαν 1669 επαναλήψεις και 4964 υπολογισμοί δυνάμεων για τον υπολογισμό της βέλτιστης ενεργειακά δομής.

Χρησιμοποιήθηκε δυναμικό EAM για το σύστημα Fe-C και ο αλγόριθμος βελτιστοποίησης conjugate gradient.

Η τελική προκύπτουσα δομή είχε ενέργεια -2018.52326142 eV και η τελική μέγιστη δύναμη ήταν 1.11235×10^{-4} eV/Å.

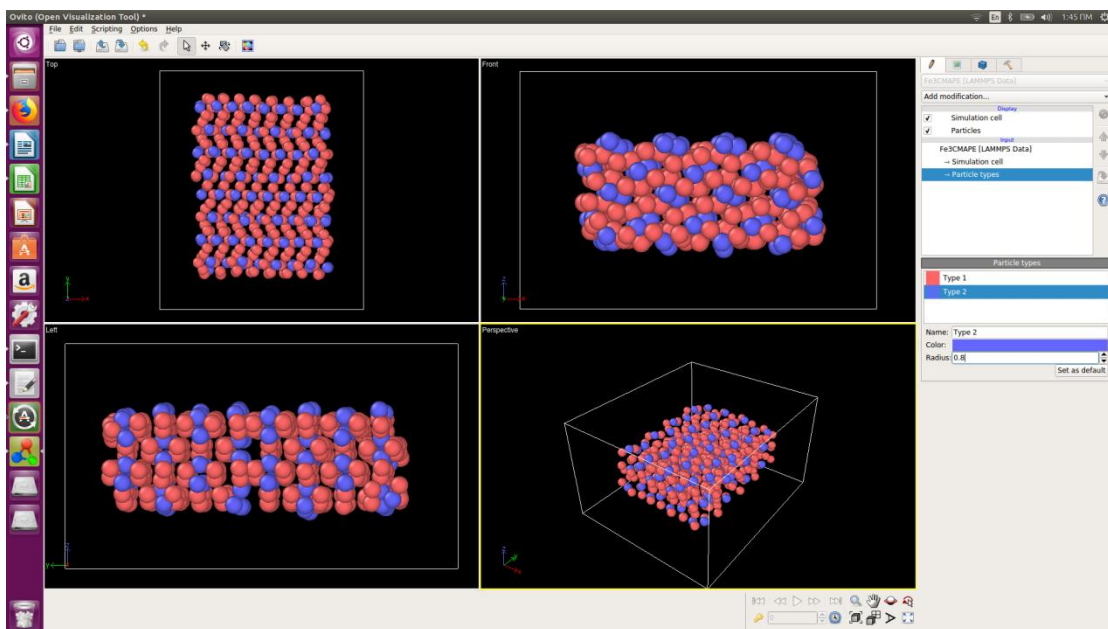
Η δομή του συστήματος μετά την προσομοίωση φαίνεται στο παρακάτω σχήμα :



Σχήμα 50 : Τελική δομή 512 ατόμων μετά την προσομοίωση

5.4.1.2 Δομή βασισμένη σε νευρωνικά δίκτυα με βάση το MAPE

Η αρχική δομή της προσομοίωσης είχε ως δεδομένα εισόδου τις συντεταγμένες που πρόβλεψε το νευρωνικό δίκτυο, αφού βελτιστοποιήθηκε ως προς το MAPE. Η αρχική ενέργεια της δομής ήταν -1688.25413794 eV, η αρχική μέγιστη δύναμη 61.3266 eV/Å, και η δομή του συστήματος πριν την προσομοίωση φαίνεται στο παρακάτω σχήμα :



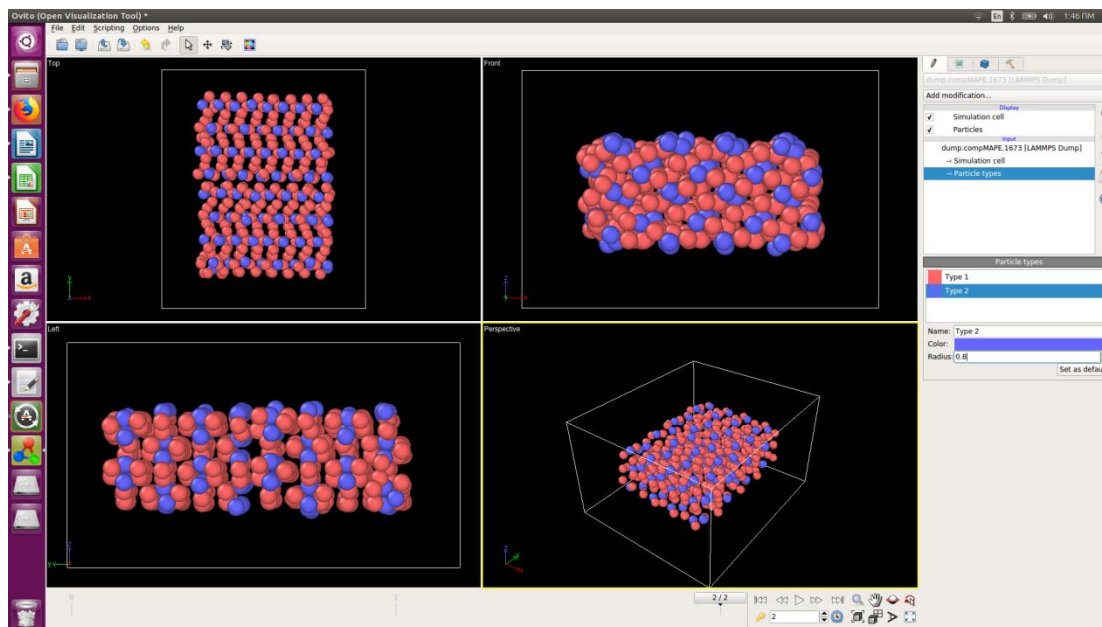
Σχήμα 51 : Αρχικές θέσεις προσομοίωσης

Η προσομοίωση διήρκεσε 5.57229 δευτερόλεπτα με 2 πυρήνες 1.8GHz. Χρειάστηκαν 1673 επαναλήψεις και 4921 υπολογισμοί δυνάμεων για τον υπολογισμό της βέλτιστης ενεργειακά δομής.

Χρησιμοποιήθηκε δυναμικό EAM για το σύστημα Fe-C και ο αλγόριθμος βελτιστοποίησης conjugate gradient.

Η τελική προκύπτουσα δομή είχε ενέργεια -2015.89588147 eV και η τελική μέγιστη δύναμη ήταν $1.64354e-04$ eV/Å.

Η δομή του συστήματος μετά την προσομοίωση φαίνεται στο παρακάτω σχήμα :



Σχήμα 52 : Τελική δομή μετά την προσομοίωση

5.5 Σύγκριση χρόνων

Παρατηρείται μείωση του χρόνου προσομοίωσης κατά 33,5687274 % (8.8 sec → 5.89 sec) στην περίπτωση του MSE και κατά 37,2099548 % (8.8 sec → 5.57 sec) στην περίπτωση του MAPE.

Προσθέτοντας και τους χρόνους τρεξίματος των νευρωνικών δικτύων (2,827 sec και 1,809 sec αντίστοιχα), το ποσοστιαίο κέρδος του συνολικού χρόνου υπολογισμών γίνεται 1,7133398 % για το MSE και 16,8256619 % για το MAPE.

Οι ιδιαίτερα χρονοβόρα βελτιστοποίηση με γενετικούς αλγόριθμους στην παρούσα μελέτη έγινε καθαρά για λόγους παρουσίασης της διαδικασίας επιτάχυνσης του χρόνου προσομοιώσεων στην ολότητά της. Με άλλα λόγια, συμφέρει μόνο σε περιπτώσεις όπου οι προσομοιώσεις των συστημάτων είναι πολύ χρονοβόρες, και στην παρούσα εργασία παρουσιάζεται για λόγους πληρότητας και όχι ελαχιστοποίησης υπολογιστικού κόστους.

Είναι προφανές ότι σε τόσο μικρούς χρόνους όπως τα 9 περίπου δευτερόλεπτα που διαρκεί η προσομοίωση μοριακής δυναμικής 512 ατόμων είναι δύσκολο να αποκομιστεί ουσιαστικό κέρδος υπολογιστικού κόστους. Όμως, με βάση την πρόσφατη βιβλιογραφία, υπάρχουν παραδείγματα όπου προσομοιώσεις που διαρκούν ώρες ή και μέρες, με βάση την διαδικασία υπολογισμού συντεταγμένων με νευρωνικά δίκτυα, καταλήγουν να διαρκούν πολλές τάξεις μεγέθους λιγότερο.

Η μάθηση των νευρωνικών δικτύων με εισόδους και εξόδους συντεταγμένες διαρκεί πάντα δευτερόλεπτα και μάλιστα πολύ λίγα, επομένως η υπό μελέτη μεθοδολογία μπορεί να εφαρμοστεί σε συνθετότερα προβλήματα, με πιο μεγάλα υπολογιστικά κόστη. Εκεί μπορούν να θυσιάστούν ορισμένες ώρες για καλύτερο MAPE μέσω των γενετικών αλγορίθμων, κάποια δευτερόλεπτα για τη μάθηση των νευρωνικών δικτύων, και να αποκομιστούν ισχυρά οφέλη σε υπολογιστικό κόστος της προσομοίωσης.

Ο καθαρός χρόνος προσομοίωσης μειώνεται κατά 1/3 στη χειρότερη περίπτωση και το υπολογιστικό αυτό κέρδος είναι το κύριο ζητούμενο της παρούσας διπλωματικής εργασίας.

Κεφάλαιο 6

6 Συμπεράσματα

Στην παρούσα διπλωματική εργασία μελετάται η ελαχιστοποίηση του χρόνου προσομοιώσεων μοριακής δυναμικής με τη χρήση νευρωνικών δικτύων, η αρχιτεκτονική των οποίων έχει προηγουμένως βελτιστοποιηθεί είτε με τη χρήση δοκιμής και πλάνης (trial and error) είτε με τη χρήση γενετικών αλγορίθμων. Στόχος της εργασίας είναι να αποδειχθεί ότι ο καθαρός χρόνος προσομοιώσεων μοριακής δυναμικής μειώνεται κατά πολύ με τη χρήση αρχικών συντεταγμένων που προέρχονται από τεχνητά νευρωνικά δίκτυα. Ένας ακόμα στόχος της εργασίας είναι ο καθορισμός μίας διαδικασίας βελτιστοποίησης της αρχιτεκτονικής των νευρωνικών δικτύων με χρήση γενετικών αλγορίθμων, ώστε να αποκτήσουν τα επιθυμητά χαρακτηριστικά που οδηγούν σε αύξηση της αποδοτικότητάς τους, ανάλογα με την περίπτωση που μελετάται.

Τα νευρωνικά δίκτυα χρησιμοποιούνται συχνά σε υπολογιστικές προσομοιώσεις για προσδιορισμό της ενέργειας ενός συστήματος, γεωμετρική βελτιστοποίησή του ή δημιουργία συναρτήσεων δυναμικού. Η παρούσα εργασία ασχολείται με τη γεωμετρική βελτιστοποίηση ενός συστήματος 512 ατόμων σεμεντίτη Fe_3C με τη μέθοδο μοριακής δυναμικής.

Το κύριο πρόβλημα στην παραπάνω διαδικασία είναι η εύρεση του/των κατάλληλων αρχιτεκτονικών των νευρωνικών δικτύων. Για την επίλυση του προβλήματος αυτού αρχικά δοκιμάστηκαν διάφοροι συνδυασμοί νευρώνων σε δύο κρυφές στρώσεις με trial and error. Λόγω του ότι οι μεταβλητές που παίζουν ρόλο στα νευρωνικά δίκτυα είναι πολλές, δεν είναι δυνατό να μελετηθεί ικανοποιητικά το παραπάνω πρόβλημα με τη μέθοδο trial and error, επομένως πρέπει να χρησιμοποιηθεί κάποια μέθοδος βελτιστοποίησης. Για το λόγο αυτό χρησιμοποιήθηκαν γενετικοί αλγόριθμοι για τη συστηματική εύρεση των βέλτιστων αρχιτεκτονικών των νευρωνικών δικτύων.

Η κύρια τάση που παρατηρήθηκε κατά τη διάρκεια του trial and error ήταν η μείωση του training MSE και του MAPE όσο αυξάνονται οι νευρώνες σε κάθε κρυφή στρώση. Όμως, από κάποιο αριθμό νευρώνων και μετά, που στο πρόβλημά μας ανήκει στο διάστημα [10,15], το δίκτυο αρχίζει να μειώνει την ικανότητα γενίκευσής του καθώς το MSE του testing αρχίζει να αυξάνεται, παρόλο που το training MSE συνεχίζει να μειώνεται. Εάν το ζητούμενο είναι το ελάχιστο training MSE και το MAPE, η αρχιτεκτονική 3-20-20-3 δίνει τα καλύτερα αποτελέσματα στο διάστημα αριθμού νευρώνων ανά κρυφή στρώση [5,20]. Εάν το ζητούμενο είναι η ελαχιστοποίηση του σφάλματος γενίκευσης, διατηρώντας παράλληλα σε

ικανοποιητικό επίπεδο το σφάλμα μάθησης, η αρχιτεκτονική 3-10-20-3 δίνει τα καλύτερα αποτελέσματα.

Με τους γενετικούς αλγόριθμους μελετήθηκαν πέντε μεταβλητές των νευρωνικών δικτύων, δηλαδή οι αριθμοί των νευρώνων σε κάθε στρώση στο διάστημα [5,20], τα ποσοστά training στο διάστημα [60,80] και testing στο διάστημα [10,20] και ο μέγιστος αριθμός σφαλμάτων του validation(max fail) στο διάστημα [6,20]. Πριν προσδιοριστούν οι βέλτιστες τιμές αυτών των πέντε μεταβλητών επιχειρήθηκε ο προσδιορισμός των βέλτιστων παραμέτρων του γενετικού αλγορίθμου. Πιο συγκεκριμένα, εξετάστηκαν τρεις τιμές του crossover fraction στο διάστημα [0,7 , 0,9] και τρεις τιμές του elite count στο διάστημα [2,4]. Η παραπάνω διαδικασία έγινε για population size 20 και 50. Με αντικειμενική συνάρτηση το MSE το καλύτερο δίκτυο είχε αρχιτεκτονική 3-14-16-3, ποσοστό training 75,2%, ποσοστό testing 10,1% και αριθμό σφαλμάτων validation 13. Με αντικειμενική συνάρτηση το MAPE το καλύτερο δίκτυο είχε αρχιτεκτονική 3-18-18-3, ποσοστό training 65,7%, ποσοστό testing 17,4% και αριθμό σφαλμάτων validation 15. Τέλος, χρησιμοποιώντας αντικειμενική συνάρτηση που συνδυάζει τα δύο προηγούμενα είδη σφαλμάτων, το καλύτερο δίκτυο είχε αρχιτεκτονική 3-19-16-3, ποσοστό training 67,051%, ποσοστό testing 17,58% και max fails 14. Τα δύο καλύτερα δίκτυα χρησιμοποιήθηκαν για πρόβλεψη της νέας γεωμετρίας εισόδου της προσομοίωσης μοριακής δυναμικής, και η μείωση του χρόνου που απαιτήθηκε για τον προσδιορισμό της βέλτιστης δομής υπερέβη το 33% και στις δύο περιπτώσεις.

Στην παραπάνω διαδικασία μπορούν να διαφοροποιηθούν μελλοντικά πολλές μεταβλητές, όπως ο αριθμός των ατόμων υπό μελέτη(μεγαλύτερο σύστημα),ο αλγόριθμος εκμάθησης του νευρωνικού δικτύου(BFGS,CG) το είδος νευρωνικού δικτύου που χρησιμοποιείται(ANFIS,RBF), η μέθοδος βελτιστοποίησης που υιοθετήθηκε(particle swarm optimization,bee colony optimization), ακόμα και η μέθοδος υπολογιστικής προσομοίωσης DFT.

7 Βιβλιογραφία

1. Ceperley, D.M., *Path integrals in the theory of condensed helium*. Reviews of Modern Physics, 1995. **67**(2): p. 279.
2. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln*. Annalen der Physik, 1927. **389**(20): p. 457-484.
3. P. Allen, M. and D. J. Tildesley, *Computer simulation of liquids* M.P. Allen, D.J. Tildesley. 1988.
4. Goldstein, H., *Classical mechanics*. 2011: Pearson Education India.
5. Mie, G., *Zur kinetischen Theorie der einatomigen Körper*. Annalen der Physik, 1903. **316**(8): p. 657-697.
6. Morse, P.M., *Diatomic molecules according to the wave mechanics. II. Vibrational levels*. Physical Review, 1929. **34**(1): p. 57.
7. Alder, B.J. and T.E. Wainwright, *Studies in molecular dynamics. I. General method*. The Journal of Chemical Physics, 1959. **31**(2): p. 459-466.
8. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas*. Physical Review, 1964. **136**(3B): p. B864-B871.
9. Sommerfeld, A., *Zur elektronentheorie der metalle auf grund der fermischen statistik*. Zeitschrift für Physik, 1928. **47**(1-2): p. 1-32.
10. Car, R. and M. Parrinello, *Unified approach for molecular dynamics and density-functional theory*. Physical review letters, 1985. **55**(22): p. 2471.
11. Rapaport, D.C. and D.C.R. Rapaport, *The art of molecular dynamics simulation*. 2004: Cambridge university press.
12. Καλτερεμίδου, Κ.-Α.Λ. and Κ.-Α.Λ. Kalteremidou, *Προσομοίωση της κοπής σε νανοκλίμακα με τη μέθοδο της Μοριακής Δυναμικής*. 2012.
13. Hernández, E.R. *Molecular Dynamics: from basic techniques to applications (A Molecular Dynamics Primer)*. in *AIP Conference Proceedings*. 2008. AIP.
14. Andersen, H.C., *Molecular dynamics simulations at constant pressure and/or temperature*. The Journal of chemical physics, 1980. **72**(4): p. 2384-2393.
15. Parrinello, M. and A. Rahman, *Crystal structure and pair potentials: A molecular-dynamics study*. Physical Review Letters, 1980. **45**(14): p. 1196.
16. Nosé, S., *A unified formulation of the constant temperature molecular dynamics methods*. The Journal of chemical physics, 1984. **81**(1): p. 511-519.
17. Hoover, W.G., *Canonical dynamics: equilibrium phase-space distributions*. Physical review A, 1985. **31**(3): p. 1695.
18. Jain, A.K., J. Mao, and K.M. Mohiuddin, *Artificial neural networks: A tutorial*. Computer, 1996. **29**(3): p. 31-44.
19. Zisis, D., E.K. Xidias, and D. Lekkas, *A cloud based architecture capable of perceiving and predicting multiple vessel behaviour*. Applied Soft Computing, 2015. **35**: p. 652-661.
20. Sengupta, N., M. Sahidullah, and G. Saha, *Lung sound classification using cepstral-based statistical features*. Computers in biology and medicine, 2016. **75**: p. 118-129.
21. Ganesan, N., et al., *Application of neural networks in diagnosing cancer disease using demographic data*. International Journal of Computer Applications, 2010. **1**(26): p. 76-85.
22. Alizadeh, E., et al., *Measuring systematic changes in invasive cancer cell shape using Zernike moments*. Integrative Biology, 2016. **8**(11): p. 1183-1193.

23. Lyons, S.M., et al., *Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas*. *Biology open*, 2016. **5**(3): p. 289-299.
24. Nabian, M.A. and H. Meidani, *Deep Learning for Accelerated Seismic Reliability Analysis of Transportation Networks*. *Computer-Aided Civil and Infrastructure Engineering*, 2018. **33**(6): p. 443-458.
25. Nabian, M.A. and H. Meidani, *Accelerating Stochastic Assessment of Post-Earthquake Transportation Network Connectivity via Machine-Learning-Based Surrogates*. 2018.
26. Hydrology, A.T.C.o.A.o.A.N.N.i., *Artificial neural networks in hydrology. I: Preliminary concepts*. *Journal of Hydrologic Engineering*, 2000. **5**(2): p. 115-123.
27. Peres, D., et al., *Significant wave height record extension by neural networks and reanalysis wind data*. *Ocean Modelling*, 2015. **94**: p. 128-140.
28. Dwarakish, G., S. Rakshith, and U. Natesan, *Review on applications of neural network in coastal engineering*. *Artificial Intelligent Systems and Machine Learning*, 2013. **5**(7): p. 324-331.
29. Ermini, L., F. Catani, and N. Casagli, *Artificial neural networks applied to landslide susceptibility assessment*. *Geomorphology*, 2005. **66**(1-4): p. 327-343.
30. Haykin, S., *Neural networks: a comprehensive foundation*. 1994: Prentice Hall PTR.
31. Artrith, N. and A. Urban, *An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO₂*. *Computational Materials Science*, 2016. **114**: p. 135-150.
32. McCulloch, W.S. and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 1943. **5**(4): p. 115-133.
33. Kleene, S.C., *Representation of events in nerve nets and finite automata*. 1951, RAND PROJECT AIR FORCE SANTA MONICA CA.
34. Hebb, D.O., *The first stage of perception: growth of the assembly*, in *The Organization of Behavior*. 2005, Psychology Press. p. 102-120.
35. Turing, A., *Intelligent machinery (1948)*. B. Jack Copeland, 2004: p. 395.
36. Farley, B. and W. Clark, *Simulation of self-organizing systems by digital computer*. *Transactions of the IRE Professional Group on Information Theory*, 1954. **4**(4): p. 76-84.
37. Rochester, N., et al., *Tests on a cell assembly theory of the action of the brain, using a large digital computer*. *IRE Transactions on information Theory*, 1956. **2**(3): p. 80-93.
38. Rosenblatt, F., *The perceptron: a probabilistic model for information storage and organization in the brain*. *Psychological review*, 1958. **65**(6): p. 386.
39. Hubel, D.H. and T.N. Wiesel, *Receptive fields of single neurones in the cat's striate cortex*. *The Journal of physiology*, 1959. **148**(3): p. 574-591.
40. Ivakhnenko, A.G. and V.G.e. Lapa, *Cybernetics and forecasting techniques*. 1967.
41. Minsky, M. and S.A. Papert, *Perceptrons: An introduction to computational geometry*. 2017: MIT press.
42. Werbos, P., *Beyond Regression: "New Tools for Prediction and Analysis in the Behavioral Sciences"*. Ph. D. dissertation, Harvard University, 1974.
43. Qian, N. and T.J. Sejnowski, *Predicting the secondary structure of globular proteins using neural network models*. *Journal of molecular biology*, 1988. **202**(4): p. 865-884.

44. Rost, B. and C. Sander, *Prediction of protein secondary structure at better than 70% accuracy*. Journal of molecular biology, 1993. **232**(2): p. 584-599.
45. Schmidhuber, J., *Learning complex, extended sequences using the principle of history compression*. Neural Computation, 1992. **4**(2): p. 234-242.
46. Hochreiter, S., *Untersuchungen zu dynamischen neuronalen Netzen*. Diploma, Technische Universität München, 1991. **91**(1).
47. Behnke, S., *Hierarchical neural networks for image interpretation*. Vol. 2766. 2003: Springer.
48. Riedmiller, M. and H. Braun. *RPROP-A fast adaptive learning algorithm*. in *Proc. of ISICIS VII*, Universität. 1992. Citeseer.
49. Hinton, G.E., S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets*. Neural computation, 2006. **18**(7): p. 1527-1554.
50. Borsboom, D., G.J. Mellenbergh, and J. Van Heerden, *The theoretical status of latent variables*. Psychological review, 2003. **110**(2): p. 203.
51. Smolensky, P., *Information processing in dynamical systems: Foundations of harmony theory*. 1986, COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE.
52. Le, Q.V. *Building high-level features using large scale unsupervised learning*. in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. 2013. IEEE.
53. Villars, P. and L. Calvert, *Pearson's Handbook of Crystallographic Data for Intermediate Phases*. American Society of Metals, Cleveland, OH, 1985.
54. Villars, P., L. Calvert, and W. Pearson, *Pearson's Handbook of Crystallographic Data for Intermetallic Phases. Volumes 1, 2, 3*. American Society for Metals, 1985, 1985: p. 3258.
55. Hahn, T., U. Shmueli, and A.J.C. Wilson, *International tables for crystallography*. Vol. 1. 1983: Reidel Dordrecht.
56. Smith, W.F. and J. Hashemi, *Foundations of materials science and engineering*. 2011: McGraw-Hill.
57. Durand-Charre, M., *Microstructure of steels and cast irons*. 2013: Springer Science & Business Media.
58. Ashrafzadeh, M., et al., *Cementite Formation from Hematite-Graphite Mixture by Simultaneous Thermal-Mechanical Activation*. Metallurgical and Materials Transactions B, 2015. **46**(2): p. 813-823.
59. Mitchell, M., *An introduction to genetic algorithms*. 1998: MIT press.
60. Fogel, L.J., A.J. Owens, and M.J. Walsh, *Artificial intelligence through simulated evolution*. 1966.
61. Rechenberg, I., *Evolutionsstrategie--Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. 1973.
62. Holland, J.H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. 1992: MIT press.
63. Mitchell, M., *Genetic algorithms: An overview*. Complexity, 1995. **1**(1): p. 31-39.
64. Goldberg, D.E., *Genetic Algorithms in Search*. Optimization & Machine Learning, 1989.
65. Bodenhofer, U., *Genetic algorithms: theory and applications*. 2003, Lecture notes, Fuzzy Logic Laboratorium Linz-Hagenberg, Winter.
66. Hobday, S., R. Smith, and J. Belbruno, *Applications of neural networks to fitting interatomic potential functions*. Modelling and Simulation in Materials Science and Engineering, 1999. **7**(3): p. 397.
67. Brenner, D.W., et al., *A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons*. Journal of Physics: Condensed Matter, 2002. **14**(4): p. 783.

68. Lorenz, S., A. Groß, and M. Scheffler, *Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks*. Chemical Physics Letters, 2004. **395**(4-6): p. 210-215.
69. Raff, L., et al., *Ab initio potential-energy surfaces for complex, multichannel systems using modified novelty sampling and feedforward neural networks*. The Journal of chemical physics, 2005. **122**(8): p. 084104.
70. Khaliullin, R.Z., et al., *Graphite-diamond phase coexistence study employing a neural-network mapping of the ab initio potential energy surface*. Physical Review B, 2010. **81**(10): p. 100103.
71. Eshet, H., et al., *Ab initio quality neural-network potential for sodium*. Physical Review B, 2010. **81**(18): p. 184107.
72. Li, J. and H. Guo, *Communication: An accurate full 15 dimensional permutationally invariant potential energy surface for the OH+ CH4→ H2O+ CH3 reaction*. 2015, AIP Publishing.
73. Morawietz, T., V. Sharma, and J. Behler, *A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges*. The Journal of chemical physics, 2012. **136**(6): p. 064103.
74. Morawietz, T. and J.r. Behler, *A density-functional theory-based neural network potential for water clusters including van der waals corrections*. The Journal of Physical Chemistry A, 2013. **117**(32): p. 7356-7366.
75. Sosso, G.C., et al., *Neural network interatomic potential for the phase change material GeTe*. Physical Review B, 2012. **85**(17): p. 174103.
76. Acı, M. and M. Avcı, *Artificial neural network approach for atomic coordinate prediction of carbon nanotubes*. Applied Physics A, 2016. **122**(7): p. 631.
77. AVCI, M., *An ANFIS Model for Geometry Optimization of Single Wall Carbon Nanotubes*.
78. Jang, J.-S., *ANFIS: adaptive-network-based fuzzy inference system*. IEEE transactions on systems, man, and cybernetics, 1993. **23**(3): p. 665-685.
79. Acı, M., Ç.İ. Acı, and M. Avcı, *PERFORMANCE COMPARISON OF ANFIS, ANN, SVR, CART AND MLR TECHNIQUES FOR GEOMETRY OPTIMIZATION OF CARBON NANOTUBES USING CASTEP*. Turkish Journal of Engineering, 2018. **2**(3): p. 119-124.
80. Artrith, N. and J. Behler, *High-dimensional neural network potentials for metal surfaces: A prototype study for copper*. Physical Review B, 2012. **85**(4): p. 045439.
81. Restrepo, S.E., S.T. Giraldo, and B.J. Thijsse, *Using artificial neural networks to predict grain boundary energies*. Computational Materials Science, 2014. **86**: p. 170-173.
82. Olmsted, D.L., S.M. Foiles, and E.A. Holm, *Survey of computed grain boundary properties in face-centered cubic metals: I. Grain boundary energy*. Acta Materialia, 2009. **57**(13): p. 3694-3703.
83. Dhole, V.S. and N.N. Patil, *A Review of Cascade Correlation Neural Network for Software Cost Estimation*.
84. Castin, N., J. Fernández, and R. Pasianot, *Predicting vacancy migration energies in lattice-free environments using artificial neural networks*. Computational Materials Science, 2014. **84**: p. 217-225.
85. Artrith, N. and A.M. Kolpak, *Grand canonical molecular dynamics simulations of Cu-Au nanoalloys in thermal equilibrium using reactive ANN potentials*. Computational Materials Science, 2015. **110**: p. 20-28.
86. Metropolis, N. and S. Ulam, *The monte carlo method*. Journal of the American statistical association, 1949. **44**(247): p. 335-341.

87. Khorshidi, A. and A.A. Peterson, *Amp: A modular approach to machine learning in atomistic simulations*. Computer Physics Communications, 2016. **207**: p. 310-324.
88. Behler, J. and M. Parrinello, *Generalized neural-network representation of high-dimensional potential-energy surfaces*. Physical review letters, 2007. **98**(14): p. 146401.
89. Behler, J., et al., *Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential*. Physical review letters, 2008. **100**(18): p. 185501.
90. Artrith, N., T. Morawietz, and J. Behler, *High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide*. Physical Review B, 2011. **83**(15): p. 153101.
91. Plimpton, S., *Fast parallel algorithms for short-range molecular dynamics*. Journal of computational physics, 1995. **117**(1): p. 1-19.