



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών
Εργαστήριο Μικροϋπολογιστών και Ψηφιακών Συστημάτων VLSI

Πολυεπίπεδη Ταχεία Προτυποποίηση και Σύνθεση Εξειδικευμένων και Επαναδιαμορφούμενων Συστημάτων Πολλαπλών Επιταχυντών Υλικού

Διδακτορική Διατριβή
Διονύσιος Διαμαντόπουλος

Αθήνα, 2015



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών
Εργαστήριο Μικροϋπολογιστών και Ψηφιακών Συστημάτων VLSI

Πολυεπίπεδη Ταχεία Προτυποποίηση και Σύνθεση Εξειδικευμένων και Επαναδιαμορφούμενων Συστημάτων Πολλαπλών Επιταχυντών Υλικού

Διδακτορική Διατριβή

του

Διονύσιου Διαμαντόπουλου

Διδάκτορος Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών
Διπλωματούχου Μηχανικού Ηλεκτρονικών Υπολογιστών και Πληροφορικής

Υποβληθείσα στο Τμήμα
Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
του Εθνικού Μετσόβιου Πολυτεχνείου

Αθήνα, 2015



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής & Υπολογιστών
Εργαστήριο Μικροϋπολογιστών και Ψηφιακών Συστημάτων VLSI

Πολυεπίπεδη Ταχεία Προτυποποίηση και Σύνθεση Εξειδικευμένων και Επαναδιαμορφούμενων Συστημάτων Πολλαπλών Επιταχυντών Υλικού

Τριμελής Συμβουλευτική Επιτροπή

.....
Δημήτριος Σούντρης
Αν. Καθηγητής Ε.Μ.Π.

.....
Κιαμάλ Πεχμεστζή
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Οικονομάκος
Επ. Καθηγητής Ε.Μ.Π.

Επταμελής Εξεταστική Επιτροπή

.....
Δημήτριος Σούντρης
Αν. Καθηγητής Ε.Μ.Π.

.....
Κιαμάλ Πεχμεστζή
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Οικονομάκος
Επ. Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Θεοδωρίδης
Επ. Καθηγητής Π. Πατρών

.....
Διονύσιος Ρεΐσης
Αν. Καθηγητής Ε.Κ.Π.Α.

.....
Michael Hübner
Καθηγητής RUB

.....
Διονύσιος Πνευματικάτος
Καθηγητής Π. Κρήτης



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
 Σχολή Ηλεκτρολόγων Μηχ. και Μηχ. Υπολογιστών
 Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
 Εργ. Μικροϋπολογιστών και Ψηφιακών Συστημάτων

**Πολυεπίπεδη Ταχεία Προτυποποίηση και Σύνθεση
 Εξειδικευμένων και Επαναδιαμορφούμενων
 Συστημάτων Πολλαπλών Επιταχυντών Υλικού**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ
ΤΟΥ
ΔΙΟΝΥΣΙΟΥ ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ

Συμβουλευτική Επιτροπή: ΔΗΜΗΤΡΙΟΣ ΣΟΥΝΤΡΗΣ
 ΚΙΑΜΑΛ ΠΕΚΜΕΣΤΖΗ
 ΓΕΩΡΓΙΟΣ ΟΙΚΟΝΟΜΑΚΟΣ

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 17η Ιουλίου 2015

ΔΗΜΗΤΡΙΟΣ ΣΟΥΝΤΡΗΣ
 ΑΝΑΠΛ. ΚΑΘΗΓΗΤΗΣ

ΚΙΑΜΑΛ ΠΕΚΜΕΣΤΖΗ
 ΚΑΘΗΓΗΤΗΣ

ΓΕΩΡΓΙΟΣ ΟΙΚΟΝΟΜΑΚΟΣ
 ΕΠΙΚ. ΚΑΘΗΓΗΤΗΣ

ΔΙΟΝΥΣΙΟΣ ΡΕΪΣΗΣ
 ΑΝΑΠΛ. ΚΑΘΗΓΗΤΗΣ

ΓΕΩΡΓΙΟΣ ΘΕΟΔΩΡΙΔΗΣ
 ΕΠΙΚ. ΚΑΘΗΓΗΤΗΣ

ΔΙΟΝΥΣΙΟΣ
 ΠΝΕΥΜΑΤΙΚΑΤΟΣ
 ΚΑΘΗΓΗΤΗΣ

MICHAEL HÜBNER
 PROFESSOR

Αθήνα, Ιούλιος 2105

.....

Διονύσιος Διαμαντόπουλος

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Εθνικό Μετσόβιο Πολυτεχνείο
Διπλωματούχος Μηχανικός Ηλεκτρονικών Υπολογιστών και Πληροφορικής
Πολυτεχνική Σχολή, Πανεπιστήμιο Πατρών

Πνευματική Ιδιοκτησία © 2015 Διονύσιος Διαμαντόπουλος
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Η παρούσα Διδακτορική Διατριβή συγχρηματοδοτήθηκε από τα ερευνητικά προγράμματα της Ευρωπαϊκής Υπηρεσίας Διαστήματος (ESA), “SPARing Robotics Technologies For Autonomous Navigation (SPARTAN)” (ESA / ESTEC ITT Reference AO / 1-6512 / 10 / NL / EK), “Spartan EXTension Activity” (SEXTANT) (ESA / ESTEC ITT Reference 4000103357 / 11 / NL / EK) και “Code Optimisation Modication Partitioning” (COMPASS) (ESA / ESTEC ITT Reference), καθώς και από τα ευρωπαϊκά ερευνητικά προγράμματα FP7-248716 2PARMA και FP7-215244 MOSART. Επίσης, μέρος της παρούσας Διδακτορικής Διατριβής συγχρηματοδοτήθηκε από εθνικά κεφάλαια και από το Εθνικό Στρατηγικό Πλαίσιο Αναφοράς (ΕΣΠΑ) 2007-2013 για τα έργα “Next Generation Millimeter Wave Backhaul Radio” (NexGenMiliwave - ΜΙΚΡΟ2-ΣΕ-B/E-II), καθώς και από τα συνεργατικά έργα έρευνας και ανάπτυξης MENELAOS και CIDCIP του βιομηχανικού οργανισμού si-Cluster στον τομέα των διαστημικών τεχνολογιών και εφαρμογών.

Ηλεκτρονική έκδοση της παρούσας διδακτορικής διατριβής είναι διαθέσιμη στον ιστότοπο: <http://nestor.microlab.ntua.gr/>

Περίληψη

Οι τεχνολογικές εξελίξεις των τελευταίων ετών έθεσαν τα θεμέλια εδραίωσης της πληροφοριοποίησης της κοινωνίας, επιδρώντας σε οικονομικές, πολιτικές, πολιτιστικές και κοινωνικές διαστάσεις. Στο απόγειο αυτής της πραγματώσης, σήμερα, ολοένα και περισσότερες καθημερινές συσκευές συνδέονται στο παγκόσμιο ιστό, αποδίδοντας τον όρο «Ίντερνετ των πραγμάτων». Το μέλλον επιφυλάσσει την πλήρη σύνδεση και αλληλεπίδραση των συστημάτων πληροφορικής και επικοινωνιών με τον φυσικό κόσμο, οριοθετώντας τη μετάβαση στα συστήματα φυσικού κυβερνοχώρου και προσφέροντας μετα-υπηρεσίες στον φυσικό κόσμο όπως προσωποποιημένη ιατρική περίθαλψη, αυτόνομες μετακινήσεις, έξυπνες ενεργειακά πόλεις κ.α. . Σκιαγραφώντας τις ανάγκες αυτής της δυναμικά εξελισσόμενης αγοράς, οι μηχανικοί υπολογιστών καλούνται να υλοποιήσουν υπολογιστικές πλατφόρμες που αφενός ενσωματώνουν αυξημένη συστημική πολυπλοκότητα και αφετέρου καλύπτουν ένα ευρύ φάσμα μετα-χαρακτηριστικών, όπως λ.χ. το κόστος σχεδιασμού, ο χρόνος σχεδιασμού, η αξιοπιστία και η επαναχρησιμοποίηση, τα οποία προδιαγράφονται από ένα αντικρουόμενο σύνολο λειτουργικών, τεχνολογικών και κατασκευαστικών περιορισμών. Η παρούσα διατριβή στοχεύει στην αντιμετώπιση των παραπάνω σχεδιαστικών προκλήσεων, μέσω της ανάπτυξης μεθοδολογιών και εργαλείων συν-σχεδίασης υλικού/λογισμικού που επιτρέπουν την ταχεία υλοποίηση καθώς και την αποδοτική σύνθεση αρχιτεκτονικών λύσεων, οι οποίες προδιαγράφουν τα μετα-χαρακτηριστικά λειτουργίας που απαιτεί η σύγχρονη αγορά. Συγκεκριμένα, στα πλαίσια αυτής της διατριβής, παρουσιάζονται α) μεθοδολογίες επιτάχυνσης της ροής σχεδιασμού τόσο για επαναδιαμορφούμενες όσο και για εξειδικευμένες αρχιτεκτονικές, β) ετερογενή αδρομερή αρχιτεκτονικά πρότυπα επιτάχυνσης επεξεργασίας και επικοινωνίας και γ) αποδοτικές τεχνικές πολυκριτηριακής σύνθεσης τόσο σε υψηλό αφαιρετικό επίπεδο προγραμματισμού, όσο και σε φυσικό επίπεδο πυριτίου.

Αναφορικά προς την επιτάχυνση της ροής σχεδιασμού, προτείνεται μια μεθοδολογία που χρησιμοποιεί εικονικές πλατφόρμες, οι οποίες αφαιρώντας τις αρχιτεκτονικές λεπτομέρειες καταφέρνουν να μειώσουν σημαντικά το χρόνο εξομοίωσης. Παράλληλα, εισηγείται η συστημική συν-εξομοίωση με τη χρήση επαναδιαμορφούμενων πλατφορμών, ως μέσων επιτάχυνσης. Με αυτόν τον τρόπο, ο κύκλος ανάπτυξης ενός προϊόντος υλικού, μετατεθειμένος από την κάθετη σειριακή ροή σε έναν κυκλικό αλληλεπιδραστικό βρόγχο, καθίσταται ταχύτερος, ενώ οι δυνατότητες προσομοίωσης εμπλουτίζονται με αποδοτικότερες μεθόδους εντοπισμού και διόρθωσης σχεδιαστικών σφαλμάτων, καθώς και μεθόδους ελέγχου των μετρικών απόδοσης του συστήματος σε σχέση με τις επιθυμητές προδιαγραφές, σε όλες τις φάσεις ανάπτυξης του συστήματος. Σε ορθογώνια συνάφεια με το προαναφερθέν μεθοδολογικό πλαίσιο, προτείνονται νέα αρχιτεκτονικά πρότυπα που στοχεύουν στη γεφύρωση του χάσματος μεταξύ της σχεδιαστικής πολυπλοκότητας και της τεχνολογικής παραγωγικότητας, με τη χρήση συστημάτων εξειδικευμένων επιταχυντών υλικού σε ετερογενή συστήματα-σε-ψηφίδα καθώς και δίκτυα-σε-ψηφίδα. Παρουσιάζεται κατάλληλη μεθοδολογία συν-σχεδίασης των επιταχυντών υλικού και του λογισμικού προκειμένου να αποφασισθεί η κα-

τανομή των εργασιών στους διαθέσιμους πόρους του συστήματος/δικτύου. Το μεθοδολογικό πλαίσιο προβλέπει την υλοποίηση των επιταχυντών είτε με συμβατικές μεθόδους προγραμματισμού σε γλώσσα περιγραφής υλικού είτε με αφαιρετικό προγραμματιστικό μοντέλο με τη χρήση τεχνικών υψηλού επιπέδου σύνθεσης. Σε κάθε περίπτωση, δίδεται η δυνατότητα στο σχεδιαστή για βελτιστοποίηση συστημικών μετρικών, όπως η ταχύτητα επεξεργασίας, η ρυθμαπόδοση, η αξιοπιστία, η κατανάλωση ενέργειας και η επιφάνεια πυριτίου του σχεδιασμού. Τέλος, προκειμένου να αντιμετωπισθεί η αυξημένη πολυπλοκότητα στα σχεδιαστικά εργαλεία επαναδιαμορφούμενων συστημάτων, προτείνονται νέοι εξελικτικοί αλγόριθμοι πολυκριτηριακής βελτιστοποίησης, οι οποίοι εκμεταλλεύονται τους σύγχρονους πολυπύρηνους επεξεργαστές και την αδρομερή φύση των πολυνηματικών περιβαλλόντων προγραμματισμού (π.χ. OpenMP), μειώνουν το χρόνο επίλυσης του προβλήματος της τοποθέτησης των λογικών πόρων σε φυσικούς, ενώ ταυτόχρονα, ομαδοποιώντας τις εφαρμογές βάση των εγγενών χαρακτηριστικών τους, διερευνούν αποτελεσματικότερα το χώρο σχεδίασης.

Η αποδοτικότητά των προτεινόμενων αρχιτεκτονικών προτύπων και μεθοδολογιών επαληθεύτηκε σε σχέση με τις υφιστάμενες λύσεις αιχμής τόσο σε αυτοτελής εφαρμογές, όπως η ψηφιακή επεξεργασία σήματος, τα πολυμέσα και τα προβλήματα αριθμητικής πολυπλοκότητας, καθώς και σε συστημικά ετερογενή περιβάλλοντα, όπως ένα σύστημα όρασης υπολογιστών για αυτόνομα διαστημικά ρομποτικά οχήματα και ένα σύστημα πολλαπλών επιταχυντών υλικού για σταθμούς εργασίας και κέντρα δεδομένων, στοχεύοντας εφαρμογές υψηλής υπολογιστικής απόδοσης (HPC). Τα αποτελέσματα ενισχύουν την πεποίθηση του γράφοντα, ότι η παρούσα διατριβή παρέχει ανταγωνιστική τεχνογνωσία για την αντιμετώπιση των πολύπλοκων σύγχρονων - και προβλεπόμενα μελλοντικών - σχεδιαστικών προκλήσεων.

Ευχαριστίες

“There is no one-man-show”, μία φράση που πρωτάκουσα στο εργαστήριο κατά την έναρξη της ερευνητικής μου ζωής και σταδιακά, έζησα, κατανόησα και αφομοίωσα. Σε αυτό το σημείο, λοιπόν, θα ήθελα να ευχαριστήσω όλα τα άτομα που συνέβαλαν στη επιτυχή ολοκλήρωση της παρούσας διδακτορικής διατριβής.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή μου κ. Δημήτριο Σούντρη για την εμπιστοσύνη του να με δεχθεί ως συνεργάτη του. Η συνεχής υποστήριξή του σε ερευνητικό επίπεδο, η αισιόδοξη ματιά του σε κάθε ερευνητική μου επιλογή και οι αναλυτικές διορθώσεις του συντέλεσαν στην επίτευξη υψηλής εγκυρότητας. Ωστόσο δε θα ήθελα να σταθώ μόνο σε αυτά τα γνωρίσματα. Ο μεθοδολογικός τρόπος σκέψης του και η επιχειρηματική κριτική του με βοήθησαν τόσο σε επιστημονικό, όσο και σε προσωπικό επίπεδο και τον ευχαριστώ θερμά για αυτό. Οι πολύωρες συζητήσεις μας, συνθέτουν ένα πολύτιμο προσωπικό βιβλίο με επιστημονικά, ιστορικά, γεωγραφικά και φιλοσοφικά κεφάλαια.

Ευχαριστώ επίσης τον Διευθυντή του εργαστηρίου Μικροϋπολογιστών και Ψηφιακών Συστημάτων, Καθηγητή κ. Κιαμάλ Πεκμεστζή για τη πολύτιμη υποστήριξη που μου παρείχε όλα αυτά τα χρόνια της έρευνάς μου στο εργαστήριο. Η επιστημονική του κατάρτιση, οι συμβουλές του, η θετική του προσέγγιση και η άρτια οργανωμένη δομή του εργαστηρίου που ο ίδιος εξασφαλίζει, αποτέλεσαν τις ιδανικότερες συνθήκες για την εκπόνηση της διατριβής μου.

Επίσης θέλω να ευχαριστήσω τον Επίκ. Καθηγητή κ. Γεώργιο Οικονομάκο. Είναι βέβαιο ότι δίχως την υποστήριξη του σε επίπεδο οργάνωσης εργαλείων λογισμικού/υλικού, αδειών χρήσης και βιβλιοθηκών σχεδίασης, καμία μέτρηση και αξιολόγηση σε αυτή τη διατριβή δεν θα μπορούσε να συλλεχθεί. Παράλληλα ο κ. Οικονομάκος διαθέτει ένα μοναδικό άοκνο ερευνητικό προφίλ, το οποίο είχα την ευκαιρία να θαυμάσω από τη συνεργασία μας στις ερευνητικές εργασίες της παρούσας διατριβής, σχετικές με το πεδίο υψηλού επιπέδου σύνθεσης (HLS).

Σε αυτό το σημείο θέλω να ευχαριστήσω δύο ανθρώπους που κατεύθυναν τα ερευνητικά θεμέλια αυτής της διατριβής. Πρόκειται για τους μεταδιδακτορικούς ερευνητές Δρ. Κωσταντίνο Σιώζιο και Δρ. Σωτήριο Ξύδη. Ο μεν, πρώτος, διασφάλισε ότι αυτή η διατριβή θα ξεκινήσει και ο δε, δεύτερος, ότι θα ολοκληρωθεί. Ο κ. Σιώζιος έχει μοναδικό τρόπο να κινητοποιεί και να οργανώνει τις ερευνητικές εργασίες, ειδικά στα πρώτα χρόνια της διδακτορικής διατριβής, όπου τα ερευνητικά μονοπάτια περιπλέκονται και φαίνονται όλα σύνθετα. Αντίστοιχα, ο κ. Ξύδης διαθέτει ένα ιδιαίτερο χάρισμα στο να εντοπίζει τα τρέχοντα ερευνητικά προβλήματα και να προτείνει καινοτόμες προσεγγίσεις. Και οι δύο διακρίνονται για την υψηλού επιπέδου επιστημονική κατάρτισή τους και είμαι σίγουρα πολύ τυχερός για τη συνεργασία μας τα τελευταία πέντε χρόνια, αλλά και κυρίως για τη φιλία μας.

Επιπλέον θέλω να ευχαριστήσω τον Αν. Καθηγητή κ. Διονύσιο Ρεΐση για την υποστήριξη που μου παρείχε κατά τη συνεργασία μας στα πλαίσια του ερευνητικού έργου NexGenMiliWave. Η συνεργασία αυτή μου έδωσε τη δυνατότητα να ασχοληθώ με εφαρμογές επεξεργασίας σήματος και να εισπράξω πολύτιμη γνώση στο πεδίο αυτό αλλά και στο τρόπο σχεδιασμού που είχε αναπτύξει η ερευνητική του ομάδα στο εργαστήριο Ψηφιακών Συστημάτων του ΕΚΠΑ, το οποίο και διευθύνει. Από την ερευνητική ομάδα του κ. Ρεΐση θέλω να ευχαριστήσω τον κ. Αθανάσιο Καραχάλιο και τον κ. Παναγιώτη Γαλιατσάτο για την υποστήριξή τους. Στα πλαίσια αυτού του έργου, ευχαριστώ την ερευνήτρια Δρ. Ροδούλα Μακρή για τη συνεχή βοήθεια της σε διαδικαστικά και τεχνικά θέματα.

Εν συνεχεία θα ήθελα να ευχαριστήσω τον ερευνητή Δρ. Γεώργιο Λεντάρη και τον Ύ.Δ. κ. Ιωάννη Σταμούλια για τη συνεργασία μας στα ερευνητικά έργα του οργανισμού ESA. Η άφθογη οργανωτικότητα του πρώτου και η συστηματικότητα του δεύτερου συνετέλεσαν σε επιτυχή και ενθουσιώδη αποτελέσματα.

Θέλω επίσης να ευχαριστήσω τα υπόλοιπα μέλη της ομάδας του εργαστηρίου Μικροϋπολογιστών και Ψηφιακών Συστημάτων για το υπέροχο περιβάλλον που έχουν αναπτύξει. Αρχικά τη γραμματέα του εργαστηρίου κα. Ζέφη Σκίνη που φροντίζει για όλες τις διαδικαστικές υποθέσεις μας. Ευχαριστώ τους ερευνητές -παρελθόντες και παρόντες- Επ.Καθ. Νίκο Μοσχόπουλο, Επ.Καθ. Ηρακλή Αναγνωστόπουλο, Δρ. Νικόλαο Ζομπάκη, Δρ. Νικόλαο Αξελό, Δρ. Αντώνη Παπανικολάου, Δρ. Αλέξανδρο Μπάρτζα, Δρ. Χρήστο Μπαλούκα και Δρ. Χριστόφορο Κάχρη, κ. Δημήτριο Μπικιάρη, κ. Ιωάννη Κούτρα, κ. Κωνσταντίνο Τσουμάνη, κ. Ευστάθιο Σωτηρίου-Ξανθόπουλο, κ. Χάρη Σιδηρόπουλο, κα. Μάρω Μπάκα, κ. Δημήτριο Ροδόπουλο, κ. Γεώργιο Ζερβάκη, κ. Νικόλαο Ευταξιοπούλο, κ. Δημήτριο Αναγνωστό, κ. Βασίλη Τσοούτσουρα, κ. Κωνσταντίνο Μαραγκό, κ. Λάζαρο Παπαδόπουλο, και κ. Γιώργο Χατζηκωσταντή. Ιδιαίτερες ευχαριστίες θέλω να αποδώσω στον κ. Μπικιάρη για την πολυτιμότερη τεχνογνωσία του στις σχεδιαστικές ροές υλικού, κατά τα πρώτα χρόνια ένταξής μου στο εργαστήριο. Επίσης ευχαριστώ τον Δρ. Κάχρη για την ευκαιρία που μου έδωσε να ασχοληθώ με το ερευνητικό πεδίο των microservers. Τέλος, ευχαριστώ τον Καθηγητή κ. Βασίλη Παυλίδη για τη πολύτιμη υποστήριξη που μου παρείχε κατά την επίσκεψή μου, ως ερευνητής, στο ακαδημαϊκό ίδρυμα EPFL, στη Λωζάνη της Ελβετίας.

Είναι αυτονόητο ότι χωρίς καλούς φίλους/φίλες να γεμίζουν μοναδικά τον ελεύθερο χρόνο, καμία ισορροπία δεν μπορεί να διατηρηθεί. Τους ευχαριστώ όλους από καρδιάς, όπως και τους συναθλητές ψαροντουφεκάδες, ιστιοπλόους, ποδηλάτες και δρομείς.

Καθώς οι αφιερώσεις τοποθετούνται στο τέλος, επέλεξα, επί τούτου, αυτό το σημείο για να ευχαριστήσω τους ανθρώπους στους οποίους και αφιερώνω τη παρούσα διδακτορική διατριβή. Πρόκειται για τους γονείς μου, Νικόλαο και Ανεζίνα και την αδερφή μου Βασιλική. Θεωρώ ότι δεν υπάρχει πιο ισχυρός ανθρώπινος σύνδεσμος από αυτόν της οικογένειας, πόσο μάλλον στη δική μου περίπτωση, που είχα τη τύχη να μεγαλώσω σε ένα περιβάλλον απεριόριστης αγάπης. Η αμέριστη συμπαράσταση και η υποστήριξη που μου παρείχαν όλα αυτά τα χρόνια είναι μόλις δύο λέξεις για τις οποίες τους ευχαριστώ. Σίγουρα τα συναισθήματα υπερισχύουν των λέξεων.

Περιεχόμενα

Περίληψη	xi
Ευχαριστίες	xiii
Κατάλογος σχημάτων	xix
Κατάλογος πινάκων	xxvii
Λίστα Συμβόλων	xxxii
1 Εισαγωγή	1
1.1 Αρχή Ήμισυ Παντός	1
1.2 Περιγραφή Χώρου Έρευνας	4
1.3 Επισκόπηση Διδακτορικής Διατριβής	9
1.3.1 Ερευνητικά Προβλήματα και Συνεισφορά	14
1.3.2 Οργάνωση Κεφαλαίων	15
Βιβλιογραφία	17
2 Ταχεία Προτυποποίηση με Υβριδικά Εικονικά Συστήματα-σε-Ψηφίδα	21
2.1 Εισαγωγή - Κίνητρο έρευνας για ταχεία εικονική προτυποποίηση	21
2.2 Υφιστάμενες Προσεγγίσεις Προτυποποίησης Συστημάτων με Εικονικές Πλατφόρμες	24
2.2.1 Επικοινωνία Μεταξύ Συστήματος Υποδοχής και Εικονικής πλατφόρμας (Host2VP)	25
2.2.2 Επικοινωνία Μεταξύ Εικονικής πλατφόρμας και Υλικού	29
2.3 Μεθοδολογία Προτυποποίησης	30
2.3.1 HotTalk API: Διεπαφή Επικοινωνίας Host2VP και VP2HW.	34
2.3.1.1 Host2VP	34
2.3.1.2 VP2HW	35
2.3.1.3 Υλοποίηση του διαύλου επικοινωνίας HotTalk.	36
2.3.2 Αξιολόγηση της τρισδιάστατης τεχνολογίας ολοκλήρωσης για ενσωματωμένα συστήματα	39
2.3.2.1 Προ-επεξεργασία	41
2.3.2.2 Δημιουργία 3-D στοίβας	42
2.3.2.3 Συστημική προτυποποίηση 3-D	43
2.4 Πειραματικά Αποτελέσματα	45
2.4.1 Αξιολόγηση του Διαύλου Επικοινωνίας	46
2.4.2 Αξιολόγηση του επεξεργαστή OpenRISC σε τρισδιάστατη τεχνολογία ολοκλήρωσης 3-D	51
2.4.3 Εφαρμογή του πλαισίου Plug&Chip σε ένα ετερογενές ενσωματωμένο σύστημα - Το έργο SPARTAN	52

2.5	Γενικά Συμπεράσματα	59
	Βιβλιογραφία	61
3	Πολυεπίπεδη Σύνθεση Ετερογενών Αρχιτεκτονικών Ειδικού Σκοπού	65
3.1	Εισαγωγή στις Αρχιτεκτονικές Ειδικού Σκοπού	65
3.2	Αρχιτεκτονικές SDR με βελτιστοποιημένο προφίλ θερμότητας και αξιοπιστίας	67
3.2.1	Εισαγωγή	67
3.2.2	Περιγραφή αρχιτεκτονικού προτύπου SDR.	70
3.2.3	Κίνητρο έρευνας για θερμικά-ασφαλείς αρχιτεκτονικές SDR	72
3.2.4	Επέμβαση σε επίπεδο μικρο-αρχιτεκτονικής	74
3.2.5	Προτεινόμενη μεθοδολογία.	78
	3.2.5.1 Σύνθεση αρχιτεκτονικής	80
	3.2.5.2 Επιλογή βάση επιφάνειας πυριτίου	80
	3.2.5.3 Επιλογή βάση χρονικής καθυστέρησης	81
	3.2.5.4 Επιλογή βάση θερμικού προφίλ	82
3.2.6	Πειραματικά αποτελέσματα	84
	3.2.6.1 Περιγραφή πειραματικού περιβάλλοντος	84
	3.2.6.2 Επιπτώσεις της επιλεκτικής αντιγραφής στη θερμοκρασία 85	
	3.2.6.3 Μελέτη επίδρασης της θερμοκρασίας στην αξιοπιστία	90
	3.2.6.4 Σχεδιασμός αρχιτεκτονικών SDR πολλαπλών επεξεργαστών ενιαίας ψηφίδας	94
3.2.7	Συμπεράσματα.	96
3.3	Ετερογενείς Αρχιτεκτονικές Δικτύων-σε-Ψηφίδα για Εφαρμογές Πολυμέ- σων	98
3.3.1	Εισαγωγή	98
3.3.2	Η αρχιτεκτονική της προτεινόμενης προσέγγισης διασύνδεσης	99
	3.3.2.1 Σχεδιασμός 2-D και 3-D δρομολογητών	101
3.3.3	Προτεινόμενη μεθοδολογία σχεδιασμού 3-D NoC	105
3.3.4	Πειραματικά αποτελέσματα	106
3.3.5	Συμπεράσματα.	115
3.4	Συστήματα Διαστημικών Εφαρμογών	116
3.4.1	Εισαγωγή	116
3.4.2	Η αρχιτεκτονική του συστήματος SPARTAN.	118
3.4.3	Μεθοδολογία συν-σχεδιασμού Υλικού/Λογισμικού.	120
	3.4.3.1 Φάση 1: Αλγοριθμική ανάλυση	120
	3.4.3.2 Φάση II: Βελτιστοποιήσεις και τροποποιήσεις ανεξαρτήτως της πλατφόρμας υλοποίησης	122
	3.4.3.3 Φάση III: Χαρακτηρισμός εφαρμογής μέσω λογισμικού	123
	3.4.3.4 Φάση IV: Συν-σχεδιασμός Υλικού/Λογισμικού	126
	3.4.3.5 Φάση V: Βελτιστοποίηση εξαρτώμενη από την πλατφόρμα υλοποίησης	127

3.4.4	Πειραματικά Αποτελέσματα	128
3.4.4.1	Γενικευμένα αποτελέσματα χαρακτηρισμού των αλγορίθμων CV.	128
3.4.4.2	Χρονική διάρκεια ζωής δεδομένων	129
3.4.4.3	Εκτιμήσεις σχετικά με την κατάτμηση υλικού/λογισμικού 130	
3.4.4.4	Ανάλυση ροής δεδομένων	134
3.4.5	Συμπεράσματα	134
3.5	3-D SoC Αρχιτεκτονικές Ψηφιακής Επεξεργασίας Σήματος	135
3.5.1	Εισαγωγή	135
3.5.2	Προτεινόμενο μεθοδολογικό πλαίσιο για 3-D SoC	135
3.5.3	Πειραματικά αποτελέσματα	141
3.5.4	Συμπεράσματα	145
	Βιβλιογραφία	147
4	Εργαλεία Σχεδιασμού για Επαναδιαμορφούμενες Πλατφόρμες	153
4.1	Εισαγωγή	153
4.2	Συναφής βιβλιογραφία.	157
4.3	Το προτεινόμενο πλαίσιο σχεδίασης.	159
4.3.1	Ο εξελικτικός αλγόριθμος δρομολόγησης GENESIS	161
4.3.1.1	Επιλογή βάση ελιτισμού	169
4.3.1.2	Σύζευξη βάση αναμιξιμότητας	169
4.3.1.3	Διασταύρωση βάση ποιότητας μονοπατιών	170
4.3.1.4	Αυτο-προσαρμοστική Μετάλλαξη	172
4.3.2	Η αδρομερής μηχανή παραλληλισμού του αλγορίθμου GENESIS	174
4.4	Προσαρμογή του εργαλείου GENESIS στα χαρακτηριστικά της εφαρμογής 177	
4.4.1	Προσαρμοσμένη Ομαδοποίηση σε Επίπεδο Εφαρμογής	179
4.4.2	Εξερεύνηση Χώρου Ομαδοποίησης για τη Βελτιστοποίηση του Χαρακτηρισμού Αξιολόγησης.	180
4.5	Πειραματικά Αποτελέσματα	185
4.6	Γενικά Συμπεράσματα	194
	Βιβλιογραφία	195
5	Σύνθεση Επαναδιαμορφούμενων Συστημάτων Πολλαπλών Επιταχυντών Υλικού	199
5.1	Αρχιτεκτονικό πρότυπο και προγραμματιστική διεπαφή για συστήματα Π.Ε. 200	
5.1.1	Εισαγωγή - Κίνητρο έρευνας σε συστήματα Π.Ε. για FPGAs	200
5.1.2	Το πλαίσιο DMM-HLS για συστήματα πολλαπλών επιταχυντών σε FPGAs	205
5.1.3	Αξιολόγηση του πλαισίου DMM-HLS	212
5.2	Κλιμάκωση συστημάτων πολλαπλών επιταχυντών σε σταθμούς εργασίας 218	
5.2.1	Εισαγωγή	218

5.2.2	Η αρχιτεκτονική HLSMapReduceFlow	220
5.2.2.1	Το πλαίσιο Phoenix MapReduce	220
5.2.2.2	Επιτάχυνση σε επίπεδο ροής δεδομένων	221
5.2.2.3	Η μεθοδολογία του πλαισίου HLSMapReduceFlow με το εργαλείο Vivado-HLS	225
5.2.3	Πειραματικά αποτελέσματα	226
5.3	Συμπεράσματα	230
	Βιβλιογραφία	231
6	Συμπεράσματα και Μελλοντικές Προεκτάσεις	235
6.1	Επισκόπηση διδακτορικής διατριβής	235
6.2	Μελλοντικές Προεκτάσεις	239
	Βιβλιογραφία	241
	Δημοσιεύσεις	243
	Βιογραφικό	249

Κατάλογος σχημάτων

1	Σύννεφο λέξεων διδακτορικής διατριβής, όπως αναλύθηκε λεκτικά από τα πακέτα λογισμικού <i>Detex v2.6</i> και <i>IBM Word Cloud build 32</i> .	x
1.1	Το αυτόματο θέατρο τού Ήρωνος. Υπολογίζεται ότι σχεδιάστηκε κατά το διάστημα 1ο π.Χ. αι - 1ο μ.Χ. αι. Αποτελείται από πλήθος επαναπρογραμματιζόμενων και εξειδικευμένων κατασκευών, ώστε να επαναδιαμορφώνεται σε διαφορετικές παραστάσεις.	3
1.2	A) Ιδιοφυείς προγραμματιζόμενοι κύλινδροι με αριστερόστροφες, δεξιόστροφες και ελεύθερες περιελίξεις νήματος σε πύρους, B) Εξειδικευμένος μηχανισμός κίνησης: Τροχοκιβώτιο με μηχανισμό ανεξάρτητης κίνησης των τροχών.	3
1.3	Παγκόσμιες πωλήσεις ημιαγωγών 1988-2014 (σε δισεκατομμύρια δολάρια). Πηγή: World Semiconductor Trade Statistics, [7].	5
1.4	Κλιμάκωση στατικής και δυναμικής κατανάλωσης ενέργειας σε ολοκληρωμένα κυκλώματα αναφορικά προς τον τεχνολογικό κόμβο. Πηγή: Mentor Graphics 2014.	6
1.5	Προβλέψεις του οργανισμού ITRS για τη κλιμάκωση της μέγιστης συχνότητας λειτουργίας ανά τακτές χρονολογικές μελέτες.	7
1.6	Τάσεις της “Σχοτεινής Σιλικόνης” για διαφορετικούς κόμβους τεχνολογίας ολοκλήρωσης [14].	7
1.7	Κλιμάκωση κατασκευαστικού κόστους ανά πύλη για διαφορετικούς κόμβους τεχνολογίας ολοκλήρωσης. Πηγή: IBS 2012	8
1.8	Σύγκριση σχεδιασμών ολοκληρωμένων κυκλωμάτων διαφορετικής αρχιτεκτονικής προσέγγισης, αναφορικά προς την ενεργειακή τους απόδοση. Πηγή: Bob Broderon, Berkeley Wireless group, ISSCC proceedings, Microsoft, 2011.	9
1.9	Ποσοστό αύξησης απόδοσης της συμβατικής τεχνολογικής κλιμάκωσης και της καινοτομίας αρχιτεκτονικής και υλικών, για κάθε τεχνολογικό κόμβο. Πηγή: IBM Microelectronics, Intel IC Insights, 2012.	10
1.10	Επισκόπηση διδακτορικής διατριβής.	11
2.1	Τάσεις κόστους ανάπτυξης υλικού και λογισμικού, αναφορικά προς τη κλιμάκωση της τεχνολογίας ολοκλήρωσης.	23
2.2	Το προτεινόμενο μεθοδολογικό πλαίσιο Plug&Chip.	31
2.3	Μηχανισμοί επικοινωνία μεταξύ συστήματος υποδοχής και εικονικής πλατφόρμας VP.	35
2.4	Μηχανισμοί επικοινωνία μεταξύ εικονικής πλατφόρμας VP και φυσικής πλατφόρμας υλοποίησης.	36

2.5	Η ροή επικοινωνίας στον πυρήνα συναλλαγής HotTalk.	37
2.6	Η αρχιτεκτονική του προτεινόμενου πυρήνα συναλλαγής.	38
2.7	Διεργασίες κατά το στάδιο προ-επεξεργασίας.	41
2.8	Διεργασίες κατά το στάδιο της δημιουργίας 3-D στοιβάς.	42
2.9	Διεργασίες κατά το στάδιο της συστημικής προτυποποίησης 3-D.	44
2.10	Παράδειγμα αξιολόγησης ενός 3-D σχεδιασμού με τέσσερα επίπεδα πυριτίου: (α) σχεδιασμός με τέσσερα εικονικά επίπεδα πυριτίου και δίκτυα TSV και (β) σχεδιασμός μετά από λεπτομερή δρομολόγηση.	45
2.11	Κέρδη επιτάχυνσης σε απόλυτο χρόνο ρολογιού για: (α) την προτεινόμενη βιβλιοθήκη Host2VP και (β) την προτεινόμενη βιβλιοθήκη VP2HW, συγκριτικά με την προσέγγιση semi-hosting[4].	47
2.12	Αξιολόγηση της επιβάρυνσης επικοινωνίας μεταξύ: (α) συστήματος υποδοχής και εικονικής πλατφόρμας VP, (β) εικονικής πλατφόρμας VP και υλικού, για μέγεθος πακέτου 4 bytes.	48
2.13	Αξιολόγηση της απόδοσης της προτεινόμενης συν-προσομοίωσης για επιλεγμένους αλγόριθμους της βιβλιοθήκης εφαρμογών MiBench.	50
2.14	Κατάτμηση του επεξεργαστή OpenRISC για εναλλακτικές απαιτήσεις σχεδίασης: (i) κατάτμηση με αλγόριθμο min-cut and (b) κατάτμηση προσανατολισμένη στη τεχνολογία υλοποίησης.	52
2.15	Φυσική υλοποίηση του επεξεργαστή OpenRISC σε τεχνολογία 3-D TSMC 45nm, με τη χρήση του εργαλείου Cadence SoC Encounter, για τη κατάτμηση των Σχημάτων 2.14(α) και 2.14(β).	53
2.16	Αρχιτεκτονική σε επίπεδο συστήματος του έργου SPARTAN.	54
2.17	Αξιολόγηση του συστήματος SPARTAN χρησιμοποιώντας ανάλυση σχεδιαστικών συμβιβασμών βασισμένη σε σενάρια λειτουργίας.	56
2.18	Αξιολόγηση του συστήματος SPARTAN αναφορικά προς την αποτελεσματικότητα της συν-προσομοίωσης.	57
2.19	Στιγμιότυπο αξιολόγησης του συστήματος SPARTAN στο περιβάλλον ROS, με εφαρμογή του πλαισίου Plug&Chip.	58
2.20	Στιγμιότυπο αξιολόγησης ενός συστήματος ελέγχου κινητήρα αυτοκινήτου (ECU) σε περιβάλλον Linux, με εφαρμογή του πλαισίου Plug&Chip.	60
3.1	Το προτεινόμενο αρχιτεκτονικό πρότυπο πλατφόρμας SDR.	71
3.2	Διάγραμμα κατανομής για (a) κατανάλωση ενέργειας και (b) πυκνότητα ενέργειας του επεξεργαστή LEON3.	73
3.3	Θερμικό προφίλ του επεξεργαστή LEON3: (a) δίχως τη χρήση μονάδων αντιγραφής, (b) με τη χρήση μονάδων αντιγραφής 2×local data/instruction memories, 2×L1 data/instruction caches, 2×register file, και (c) με τη χρήση μονάδων αντιγραφής 2×instruction unit, 2×cache controller, 2×AHB controller.	74
3.4	Προτεινόμενες μικρο-αρχιτεκτονικές τροποποιήσεις.	76
3.5	Ο ελεγκτής θερμοκρασίας πραγματικού χρόνου για κάθε αντίγραφο μονάδας υλικού.	78

3.6	Η προτεινόμενη μεθοδολογία για διαχείριση της θερμοκρασίας μέσω αντιγράφων μονάδων υλικού.	79
3.7	Διακυμάνσεις θερμοκρασίας για διαφορετικές αρχιτεκτονικές υλοποιήσεις της πλατφόρμας LEON3 SDR.	86
3.8	Η πυκνότητα κατανάλωσης ενέργειας σε σχέση με τη μέγιστη θερμοκρασία.	88
3.9	Η επιφάνεια πυριτίου σε σχέση με τη μέγιστη θερμοκρασία.	90
3.10	Αξιολόγηση της παραμέτρου επιτάχυνσης γήρανσης A_f για αρχιτεκτονικές με διαφορετική μέση τιμή θερμοκρασίας.	91
3.11	Αξιολόγηση διαφορετικών αρχιτεκτονικών αναφορικά προς το πρόβλημα TDDb.	93
3.12	Θερμική ανάλυση για ένα σύστημα CMP που αποτελείται από 2x2 επεξεργαστές LEON3.	94
3.13	Κανονικοποιημένη πυκνότητα κατανάλωσης ενέργειας σε σχέση με την επιφάνεια πυριτίου, για συστήματα πολλαπλών επεξεργαστών LEON3.	95
3.14	Κανονικοποιημένη μέγιστη θερμοκρασία σε σχέση με την επιφάνεια πυριτίου, για συστήματα πολλαπλών επεξεργαστών LEON3.	96
3.15	Παράδειγμα γράφου επικοινωνίας εφαρμογής.	100
3.16	Εναλλακτικές αρχιτεκτονικές 3-D NoCs για την εφαρμογή που απεικονίζεται στο Σχήμα 3.15.	102
3.17	Προτεινόμενο αρχιτεκτονικό πρότυπο 3-D δρομολογητή.	103
3.18	Δομή του πακέτου της προτεινόμενης αρχιτεκτονικής NoC.	104
3.19	Προτεινόμενη μεθοδολογία σχεδιασμού 3-D NoCs.	105
3.20	Απεικόνιση της εφαρμογής VOPD σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.	107
3.21	Απεικόνιση της εφαρμογής MWD σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.	108
3.22	Απεικόνιση της εφαρμογής MPEG-4 σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.	108
3.23	Απεικόνιση της εφαρμογής MMS σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.	109
3.24	Αξιολόγηση δισδιάστατων και τρισδιάστατων δρομολογητών αναφορικά προς τον αριθμό των δεσμευμένων λογικών πυλών, σε σχέση με τον αριθμό των θηρών E/E.	110
3.25	Αξιολόγηση δισδιάστατων και τρισδιάστατων δρομολογητών αναφορικά προς τη καθυστέρηση, σε σχέση με τον αριθμό των θηρών E/E.	111
3.26	Αξιολόγηση δισδιάστατων και τρισδιάστατων δρομολογητών αναφορικά προς τη κατανάλωση ενέργειας, σε σχέση με τον αριθμό των θηρών E/E.	112
3.27	Φυσική απεικόνιση της προτεινόμενης ετερογενούς αρχιτεκτονικής 3-D NoC για την εφαρμογή MPEG-4.	113

3.28	Αριθμός packet hops για διαφορετικές αρχιτεκτονικές προσεγγίσεις: (i) ομογενές 2-D NoC, (ii) ομογενές 3-D NoC και (iii) προτεινόμενο ετερογενές 3-D NoC.	114
3.29	Μέγιστη συχνότητα λειτουργίας για διαφορετικές αρχιτεκτονικές προσεγγίσεις NoC.	115
3.30	Power consumption for different instantiations of NoC.	115
3.31	Σχηματική απεικόνιση των λειτουργιών χαρτογράφησης και εντοπισμού του συστήματος SPARTAN.	119
3.32	Ροή διεργασιών της προτεινόμενης μεθοδολογίας συν-σχεδιασμού Υλικού/Λογισμικού.	120
3.33	Διεργασίες που εκτελούνται κατά τη φάση της αλγοριθμικής ανάλυσης.	121
3.34	Προτεινόμενη ροή εργασιών για βελτιστοποιήσεις και τροποποιήσεις ανεξαρτήτως της πλατφόρμας υλοποίησης.	122
3.35	Προτεινόμενα εργαλεία χαρακτηρισμού εφαρμογών.	124
3.36	Προτεινόμενη μεθοδολογία για τον χαρακτηρισμό αλγορίθμων CV μέσω λογισμικού.	126
3.37	Προτεινόμενη ροή εργασιών για συν-σχεδιασμό υλικού/λογισμικού.	127
3.38	Προτεινόμενη ροή εργασιών προς βελτιστοποίηση του συστήματος SPARTAN, λαμβάνοντας υπόψιν την πλατφόρμα υλοποίησης.	127
3.39	Αποτελέσματα χαρακτηρισμού των αλγορίθμων CV για τις λειτουργίες (α) χαρακτηρισμού και (β) εντοπισμού.	129
3.40	Χρονική διάρκεια ζωής δεδομένων για τις λειτουργίες (α) της χαρτογράφησης και (β) του εντοπισμού.	130
3.41	Αδρομερής κατάτμηση υλικού/λογισμικού για τη λειτουργία της χαρτογράφησης του ρομποτικού οχήματος του έργου SPARTAN.	132
3.42	Αδρομερής κατάτμηση υλικού/λογισμικού για τη λειτουργία του εντοπισμού του ρομποτικού οχήματος του έργου SPARTAN.	133
3.43	Προτεινόμενο μεθοδολογικό πλαίσιο για τη υποστήριξη ταχείας αξιολόγησης 3-D SoCs συστημάτων.	136
3.44	Αναλυτική ροή εκτέλεσης εργασιών του προτεινόμενου 3-D πλαισίου.	136
3.45	Δημιουργία της 3-D στοίβας με το εργαλείο 3DPart, χρησιμοποιώντας τον αλγόριθμο βελτιστοποίησης <i>Tabu Search</i> [66].	138
3.46	Συστημική προτυποποίηση 3-D.	139
3.47	Αξιολόγηση μετρικών απόδοσης σχεδιασμών 3-D με υφιστάμενα εμπορικά και νέα ακαδημαϊκά εργαλεία CAD.	141
3.48	Ενσωματωμένο σύστημα βασισμένο στον επεξεργαστή Leon3 [18].	142
3.49	Παράδειγμα 3-D σχεδιασμού για ένα ενσωματωμένο σύστημα βασισμένο στον επεξεργαστή LEON3 με το συμβατικό εργαλείο 2-D σχεδίασης Cadence SoC Encounter [24].	143
4.1	Σύγκριση της απόδοσης των επεξεργαστών γενικού σκοπού με τα FPGAs, σε σχέση με τη κλιμάκωση των λογικών πόρων.[3].	154

4.2	Προτεινόμενο μεθοδολογικό πλαίσιο για παράλληλη τοποθέτηση σε FPGAs.	160
4.3	Η προτεινόμενη αναπαράσταση λύσεων τοποθέτησης με τη χρήση χρωμοσωμάτων και ένα παράδειγμα κωδικοποίησης netlist με αλφαριθμητικά αλληλόμορφα.	163
4.4	ο χώρος λύσεων της εφαρμογής <i>alu4</i> με έξι αλληλόμορφα ανά αλληλότοπο. Τα αλληλόμορφα $CLB_{1:6}$ στον αλληλότοπο-CLB και τα $IO_{1:6}$ στον αλληλότοπο-IO συνιστούν 36 γονότυπους CLB_jIO_k με μετρική καταλληλότητας $f_{j,k}$, που αντιπροσωπεύεται από το ύψος των ράβδων.	165
4.5	(a) Δομικό πρότυπο αρχιτεκτονικής FPGA. (b) Σύνθεση HDL-σε-CLBs. (c) Αναπαράσταση ενός σχεδιασμού με γράφο εκτέλεσης. (d) Φυσική τοποθέτηση σε FPGA.	168
4.6	Παράδειγμα επίδρασης του ποσοστού ελιτισμού για το κύκλωμα s38417 του πακέτου MCNC με 6406 CLBs, 29 εισόδους και 106 εξόδους, $n_p=100$, $g_p=500$, $mutation_rate=25$	170
4.7	Η πολυ-νηματική δομή του υποσυστήματος μνήμης και το αντίστοιχο παράλληλο μοντέλο επεξεργασίας fork-join.	175
4.8	Κατανομή των ποσοστών του χρόνου εκτέλεσης για τις συναρτήσεις πηγαίου κώδικα του αλγορίθμου GENESIS.	178
4.9	Ομαδοποίηση στο χώρο χαρακτηριστικών F	182
4.10	Ο χώρος Pareto για την χρονική καθυστέρηση του κυκλώματος σε σχέση με το χρόνο εκτέλεσης του εργαλείου για κάθε ομάδα F_i . Τα δεδομένα είναι κανονικοποιημένα αναφορικά προς την λιγότερο αποτελεσματική λύση ανά ομάδα.	185
4.11	Αξιολόγηση για διαφορετικούς συνδυασμούς των παραμέτρων <i>population size</i> και <i>population age</i> για την εφαρμογή <i>bigkey</i>	187
4.12	Υποψήφιες λύσεις που αξιολογούνται κατά τα στάδια εκτέλεσης των εργαλείων GENESIS και VPR.	191
4.13	Κλιμάκωση της επιτάχυνσης εκτέλεσης του εργαλείου GENESIS, προς τον αριθμό των φυσικών πυρήνων και των νημάτων επεξεργασίας.	192
4.14	Δέσμευση φυσικής μνήμης για (a) ελάχιστο μέγεθος FPGA και (b) διπλάσιο μέγεθος FPGA, αναφορικά προς τις απαιτήσεις της εφαρμογής.	193
4.15	Τάση κλιμάκωσης της δέσμευσης μνήμης για FPGA με μέγεθος (α) ελάχιστο αναφορικά με το μέγεθος της εφαρμογής και (β) διπλάσιο αναφορικά με το μέγεθος της εφαρμογής.	193
5.1	Ανάλυση κλιμάκωσης επιταχυντών του αλγορίθμου ομαδοποίησης K_{means} : A_i -Επιταχυντές= [1 : 128], N_p -Δεδομένα= 2×10^4 , P_k -Ομάδες=3	202

5.2	Παραδειγματικό σενάριο εκτέλεσης τεσσάρων επιταχυντών υλικού. <i>Acc-ι Στατική Χρήση Μνήμης%, Καθυστέρηση]. Όλοι οι επιταχυντές αναμένεται να εκκινήθουν τη χρονική στιγμή 0. Ο σχεδιασμός κρίνεται μη-συνθέσιμος με στατική δέσμευση μνήμης. Το κατώτερο τμήμα του σχήματος απεικονίζει τον τελικό χρονοπρογραμματισμό με χρήση τεχνικών ΔΔΜ. Το ανώτερο τμήμα του σχήματος απεικονίζει την αντίστοιχη δέσμευση της συνολικής μνήμης. Εξαιτίας της κλιμάκωσης του χρόνου, στο συγκεκριμένο σχήμα δεν απεικονίζεται το αποτύπωμα μνήμης κατά τις στιγμιαίες μεταβάσεις νέων δεσμεύσεων/αποδεσμεύσεων μνήμης.</i>	204
5.3	Προτεινόμενο λογικό αρχιτεκτονικό πρότυπο για συστήματα πολλαπλών επιταχυντών υλικού σε FPGAs.	205
5.4	Προτεινόμενο αρχιτεκτονικό πρότυπο για συστήματα πολλαπλών επιταχυντών υλικού σε FPGAs, σε επίπεδο υλικού.	206
5.5	Κέρδος απόδοσης λόγω παραλληλίας και επικάλυψης διεργασιών εκτέλεσης, μέσω του πλαισίου DMM-HLS. Χρονοπρογραμματισμός των επιταχυντών και αποτύπωμα μνήμης για α) συμβατικά εργαλεία HLS με στατική δέσμευση μνήμης, β) πλαίσιο DMM-HLS με ένα μοναδικό σωρό, γ) πλαίσιο DMM-HLS με δύο σωρούς. Συσκευή FPGA: 90 BRAMs. Συνολική ζήτηση μνήμης: 113 BRAMs.	207
5.6	Αρχιτεκτονική διάταξη των ελεγκτών μνήμης του πλαισίου DMM-HLS για ΔΔΜ σε FPGAs.	209
5.7	Προτεινόμενη επέκταση του εργαλείου Vivado HLS για την υποστήριξη ΔΔΜ σε συστήματα πολλαπλών επιταχυντών σε FPGAs.	211
5.8	Κανονικοποιημένη καθυστέρηση ανά επιταχυντή, λόγω μηχανισμών ΔΔΜ, αναφορικά προς τον αριθμό των χρησιμοποιούμενων σωρών.	213
5.9	Σύγκριση επεξεργαστικής ισχύος μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔMM DMM-HLS.	214
5.10	Σύγκριση ρυθμαπόδοσης μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔMM DMM-HLS.	214
5.11	Σύγκριση επιβάρυνσης πόρων σε σχέση με τη ρυθμαπόδοση και την ενέργεια, μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔMM DMM-HLS.	215
5.12	Σύγκριση επιβάρυνσης πόρων σε σχέση με τη ρυθμαπόδοση και την ενέργεια, μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔMM DMM-HLS, για όλες τις εφαρμογές αξιολόγησης.	217
5.13	Το πλαίσιο MapReduce.	221
5.14	Τοπολογία αρχιτεκτονικής για α) υφιστάμενο τυπικό πλαίσιο MapReduce και β) προτεινόμενο πλαίσιο HLSMapReduceFlow.	222
5.15	Η αρχιτεκτονική HLSMapReduceFlow: Κάθε επεξεργαστικός dataflow-κόμβος εργάζεται επί αποκλειστικής μνήμης. Η συνολική μνήμη του συστήματος διαιρείται σε μνήμες <i>k-port</i> και <i>j-port</i> για έναν αριθμό νημάτων εκτέλεσης <i>k-map</i> και <i>j-reduce</i>	223
5.16	Εξερεύνηση ροής δεδομένων από τη ροή έλεγχου της εφαρμογής εισόδου, με το εργαλείο Vivado HLS.	224

- 5.17 Παράδειγμα χρήσης ντιρεκτίβας *DATAFLOW* (a) Ακολουθιακή εκτέλεση (b) Παράλληλη επικαλυπτόμενη επεξεργαστική ροή. . . . 225
- 5.18 Προτεινόμενη επέκταση του εργαλείου Vivado HLS για υποστήριξη του πλαισίου MapReduce σε πλατφόρμες FPGA. 226
- 5.19 Ανάλυση επιδόσεων κλιμάκωσης με τη χρήση του πλαισίου HLSMapReduceFlow. 228

Κατάλογος πινάκων

2.1	Ποιοτική σύγκριση για μηχανισμούς επικοινωνίας Host-to-VP. . . .	27
2.2	Ποιοτική σύγκριση για μηχανισμούς επικοινωνίας VP-to-Hardware.	28
2.3	Χαρακτηριστικά της χρησιμοποιούμενης τεχνολογίας TSV [23]. . .	51
2.4	Αποτίμηση υλοποίησης σε φυσικό επίπεδο μάσκας πυριτίου του επεξεργαστή OpenRISC, με τεχνολογία 3-D.	52
3.1	Παράμετροι του μοντέλου θερμικής εξομοίωσης της αρχιτεκτονι- κής LEON3 για εφαρμογές SDR.	84
3.2	Χαρακτηριστικά της επιλεγμένης τεχνολογίας TSV.	112
3.3	Μετρικές υλοποίησης των εφαρμογών αξιολόγησης.	114
3.4	Μετρικές σχετικά με τη φυσική υλοποίηση σε επίπεδο μάσκας πυριτίου.	144
3.5	Χρόνος εξομοίωσης και κατανάλωση ενέργειας για διαφορετικούς αλγορίθμους επεξεργασίας σήματος.	145
4.1	Υφιστάμενες λύσεις παράλληλων αλγορίθμων τοποθέτησης. . . .	158
4.2	Οι παράμετροι του προτεινόμενου γενετικού αλγορίθμου.	169
4.3	Χαρακτηρισμός μετρικών για τις επιλεγμένες εφαρμογές αξιολό- γησης.	181
4.4	Αξιολόγηση εναλλακτικών ρυθμίσεων του αλγορίθμου GENESIS. .	183
4.5	Σύγκριση του χρόνου εκτέλεσης, της μέγιστη επιτευχθείσας συ- χνότητα λειτουργίας και της κατανάλωσης ενέργειας για τους αλ- γορίθμους τοποθέτησης GENESIS και VPR.	188
5.1	Παραδειγματικό σενάριο μετατροπής στατικών δεσμεύσεων μνή- μης σε δυναμικές, χρησιμοποιώντας κλείσεις του DMM-HLS API. .	212
5.2	Χαρακτηρισμός εφαρμογών αξιολόγησης.	212
5.3	Σύγκριση του πλαισίου HLSMapReduceFlow-FPGA και επεξεργα- στή γενικού σκοπού, σε αντιπροσωπευτικές εφαρμογές σταθμού εργασίας.	229

Λίστα Συμβόλων

<i>3SLOG</i>	three (3) Step LOGarithmic search
<i>ADPCM</i>	Adaptive Differential Pulse-Code Modulation
<i>AHB</i>	Advanced High-performance Bus
<i>AMBA</i>	Advanced Microcontroller Bus Architecture
<i>APB</i>	Advanced Peripheral Bus
<i>API</i>	Application Programming Interface
<i>ASIC</i>	Application Specific Integrated Circuit
<i>BLE</i>	Basic Logic Element
<i>CAD</i>	Computer Aided Design
<i>CLB</i>	Configurable Logic Block
<i>CPS</i>	Cyber Physical Systems
<i>CRC</i>	Cyclic Redundancy Check
<i>CV</i>	Computer Vision
<i>DMA</i>	Direct Memory Access
<i>DMM</i>	Dynamic Memory Management
<i>DSP</i>	Digital Signal Processor
<i>DSU</i>	Debug Support Unit
<i>DUT</i>	Design Under Test
<i>DVFS</i>	Dynamic Voltage and Frequency Scaling
<i>ECO</i>	Engineering change orders
<i>ECU</i>	Engine Control Unit
<i>EDA</i>	Electronic Design Automation
<i>ESL</i>	Electronic System Level
<i>ESoC</i>	Ecosystem-on-a-chip

<i>FFT</i>	Fast Fourier Transform
<i>FIFO</i>	First-In, First-Out
<i>FS</i>	Full Search
<i>FSB</i>	Front-Side Bus
<i>FSM</i>	Finite-State Machine
<i>HDL</i>	Hardware Description Language
<i>HPC</i>	High Performance Computing
<i>HPWL</i>	Half-perimeter Wirelength
<i>HS</i>	Hierarchical Search
<i>HT</i>	HyperTransport
<i>IC</i>	Integrated Circuit
<i>IDW</i>	Inverse Distance Weighting
<i>IP</i>	Intellectual Property
<i>IPC</i>	Inter-Process Communication
<i>ISA</i>	Instruction Set Architecture
<i>ISS</i>	Instruction Set Simulator
<i>ITRS</i>	International Technology Roadmap for Semiconductors
<i>LAB</i>	Logic Array Block
<i>LUT</i>	Look-up Table
<i>MMS</i>	MultiMedia System
<i>MPSoCs</i>	Multi-Processor System-on-Chips
<i>MTTF</i>	Mean Time To Failure
<i>MTU</i>	Maximum Transmission Unit
<i>NASA</i>	National Aeronautics and Space Administration
<i>NIC</i>	Network Interface Controller
<i>NUCA</i>	Non-Uniform Cache Architecture
<i>PHODS</i>	Parallel Hierarchical One-Dimensional Search
<i>POSIX</i>	Portable Operating System Interface for Unix

<i>QoR</i>	Quality of Results
<i>QPI</i>	Intel QuickPath Interconnect
<i>ROS</i>	Robotic Operating System
<i>RTL</i>	Register-Transfer Level
<i>RTS</i>	Run-Time Situation
<i>SDR</i>	Software Defined Radio
<i>SiP</i>	System-in-package
<i>SLAM</i>	Simultaneous Localization And Mapping
<i>SNR</i>	Signal-to-Noise Ratio
<i>SoC</i>	System-on-chip
<i>SPEF</i>	Standard Parasitic Exchange Format
<i>SS</i>	Spiral Search
<i>SURF</i>	Speeded Up Robust Features
<i>T – VPack</i>	Timing-Driven Versatile Packing
<i>TDDb</i>	Time-Depended Dielectric Breakdown
<i>TDP</i>	Thermal Design Power
<i>TSV</i>	Through Silicon Via
<i>VCD</i>	Value Change Dump
<i>VPR</i>	Versatile Placement and Routing

1

Εισαγωγή

1.1. Αρχή Ήμισυ Παντός

Η τεχνολογία, εκ της σύνθεσης των λέξεων “τέχνη” και “λόγος”, αποτελεί σκηνοθετικό παράγοντα της αλληλεπίδρασης του ανθρώπου με το περιβάλλον του, κατά τη μετάβαση από το παρελθόν, στο παρόν και μέλλον. Η ιστορία έχει δείξει ότι δημιουργήματα που θεωρούνται σήμερα δεδομένα, για την εποχή που ανακαλύφθηκαν ήταν σημαντικά τεχνολογικά επιτεύγματα. Υπό άλλη οπτική γωνία, ο χαρακτηρισμός “τέλειο” αναφερόμενος στη χρονική στιγμή του σήμερα, αντιστοιχεί στο παρωχημένο του μέλλοντος. Η επιστήμη σχετίζεται με την αέναη αναζήτηση της αλήθειας και άρα της τελειότητας. Τα επιστημονικά και τεχνολογικά επιτεύγματα κάθε εποχής εμπεριέχουν αδιαμφισβήτητη την ύψιστη γνώση, δημιουργικότητα, τέχνη και λόγο που χαρακτηρίζουν κάθε πολιτισμό.

Η δημιουργική σύμπραξη νέας ιδέας, εξειδίκευσης και επαναχρησιμοποίησης αποτελεί διαχρονική διαδικασία μετάβασης από μία ώριμη τεχνολογία σε μία αναδύομενη και ρηζικέλευθη. Αυτή η διαπίστωση μπορεί να επιβεβαιωθεί ανατρέχοντας στην ιστορία μεγάλων πολιτισμών και παρατηρώντας τη σταδιακή εξέλιξη των τεχνολογικών επιτευγμάτων. Χαρακτηριστικό παράδειγμα είναι το θέατρο, τόσο ως τέχνη, όσο και ως οικοδόμημα. Ως παράδειγμα πολιτισμού αναφοράς, επιλέγεται ο αρχαιότερος πολιτισμός που ασχολήθηκε με το θέατρο, ο Αρχαίος Ελληνικός πολιτισμός.

Το αρχαίο ελληνικό θέατρο, θεσμός της αρχαιοελληνικής πόλης-κράτους αναφέρεται στη διδασκαλία και τέλεση θεατρικών παραστάσεων, επ’ ευκαιρία των εορτασμών του Διονύσου. Αναπτύχθηκε στα τέλη της αρχαϊκής περιόδου (750 π.Χ. - 479 π.Χ) και διαμορφώθηκε πλήρως κατά την κλασική περίοδο (499 π.Χ. - 323 π.Χ.), κυρίως στην Αθήνα. Φέρει έναν έντονο θρησκευτικό και μυστηριακό χαρακτήρα κατά τη διαδικασία της γέννησής του, αλλά και έναν εξίσου έντονο κοινωνικό και πολιτικό χαρακτήρα κατά την περίοδο της ανάπτυξής του [1].

Όλες οι δραματικές παραστάσεις διδάσκονταν κατά τη διάρκεια των γιορτασμών προς τιμήν του Διόνυσου, του επαναστάτη θεού που προτίμησε τη θηλυκή φρενίτιδα για τη λατρεία του, μια μανιακή φρενίτιδα και αμφισβήτηση της πατριαρχικής εξουσίας εκφρασμένη στις πράξεις των ηρωίδων του αττικού δράματος. Ο Διόνυσος, ως θεός της γονιμότητας και της βλάστησης συνδέεται με το πάθος, την έκσταση των συμμετεχόντων στη λατρεία του, τη χαρά, τον τρόμο και κάλλιστα μπορεί να ενωθεί τόσο με την τραγωδία όσο και με την κωμωδία [2-4].

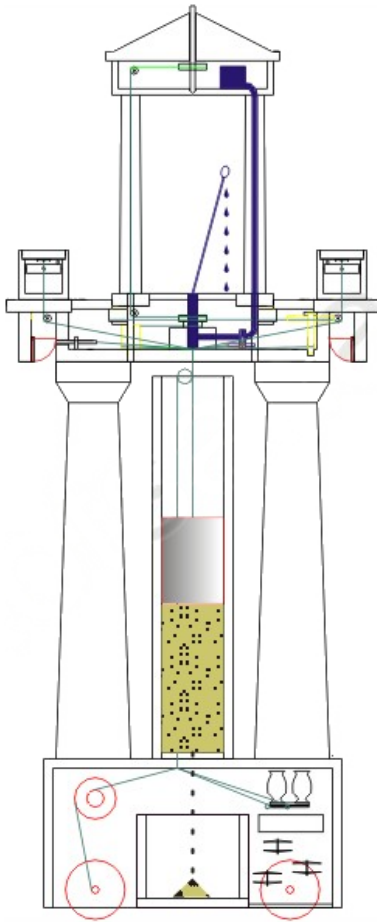
Το σπουδαιότερο αρχαίο ελληνικό θέατρο ήταν το θέατρο του Διονύσου¹, το οποίο θεμελιώθηκε πιθανώς τον 6ο π.Χ. αιώνα και έχει χωρητικότητα 15.000 θεατών. Η απόσταση της πρώτης σειράς (προεδρία) από την τελευταία εκτιμάται γύρω στα 80 μέτρα, ενώ από τους ηθοποιούς γύρω στα 12 μέτρα. Η ψηλότερη σειρά θέσεων του θεάτρου υψωνόταν περίπου 35 μέτρα επάνω από το χαμηλότερο μέρος του περιβόλου.

Το αρχαίο ελληνικό θέατρο υπήρξε μια εκπληκτική καινοτομία σε παγκόσμια κλίμακα και ενσωμάτωνε κάθε φορά την τεχνολογία αιχμής της εποχής. Η ίδια η εξέλιξη του θεάτρου ως οικοδόμημα υπήρξε εντυπωσιακή από την απομονωμένη “ορχήστρα” με την προσθήκη του ιδιοφυούς “κοίλου” και τη χρήση απλών ξύλινων σκηνικών έως τα εκθαμβωτικά λίθινα συγκροτήματα με την εκπληκτική “σκηνή”, με τις υπέροχες κιονοστοιχίες του “προσκηνίου” και των δύο “παρασκηνίων”, την επιβλητική εξέδρα του “λογείου” και τις απόλυτα εναρμονισμένες στο χώρο “παρόδους” [5].

Από τα πρώιμα χρόνια της εμφάνισης του θεάτρου (6ος αιώνας π.Χ.) πλήθος τεχνολογικών καινοτομιών βρήκαν εφαρμογή ή επινοήθηκαν για να βελτιστοποιήσουν τη λειτουργία του και να εξυπηρετήσουν το θέαμα. Η “Χαρώνειος κλίμακα” μια κρυφή υπόγεια διάβαση για την έλευση του Χάροντα, τα “αναπίεσματα»” σύστημα καταπακτών για την ταχεία εμφάνιση και εξαφάνιση προσώπων και αντικειμένων, ο “από μηχανής θεός” μια ανυψωτική μηχανή για την εναέρια μεταφορά προσώπων (πρόδρομος των σημερινών ανυψωτικών θεατρικών μηχανισμών), το “εκκύκλημα” και η “εξώστρα” για την απλή μεταφορά νεκρών ή αντικειμένων στη σκηνή έως την ταχεία εμφάνιση ή και αλλαγή του εσωτερικού σκηνικού όμοια με τις σύγχρονες περιστροφικές σκηνές πάνω σε βαγονέτο [5].

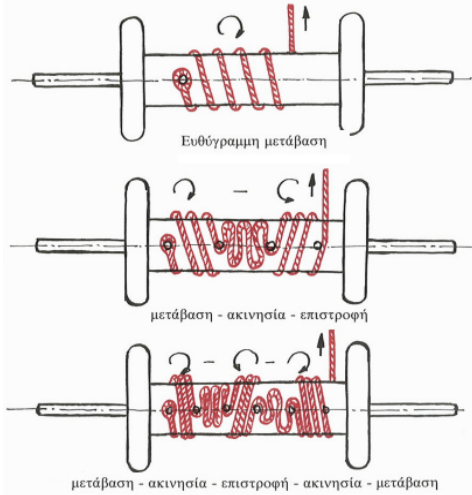
Ωστόσο, το εντυπωσιακότερο, ίσως, τεχνολογικό κατασκεύασμα, σχετικό με το αρχαίο θέατρο, είναι το σταθερό αυτόματο θέατρο του Ήρωνος του Αλεξανδρινού, το οποίο είναι ακριβής ανακατασκευή του σταθερού αυτόματου θεάτρου του Φίλωνος του Βυζαντίου (3ος αι. π.Χ.), που βελτίωσε και περιέγραψε με λεπτομέρειες ο Ήρων ο Αλεξανδρεύς στο βιβλίο του “Αυτοματοποιητική” [6] και απεικονίζεται στο Σχήμα 1.1. Το κατασκεύασμα αυτό παρουσίαζε αυτόματα το μύθο του Ναύπλιου, πατέρα του Ομηρικού ήρωα Παλαμήδη. Ο Ναύπλιος επιθυμώντας να εκδικηθεί τους Αχαιούς που είχαν σκοτώσει τον Παλαμήδη δια

¹Παρόλο που το θέατρο του Διονύσου ήταν το σπουδαιότερο, ωστόσο, το αρχαιότερο θέατρο στην Ελλάδα είναι το θέατρο του Θορικού, το οποίο κατασκευάστηκε τον 6 αιώνα π.χ. και η ιδιαιτερότητα του βρίσκεται στο ιδιόμορφο ελλειψοειδές σχήμα του σε αντίθεση με την ημικυκλική κατασκευή των νεότερων σε χρονολογία θεάτρων. Η χωρητικότητα του ανέρχεται στους 4.000 θεατές.

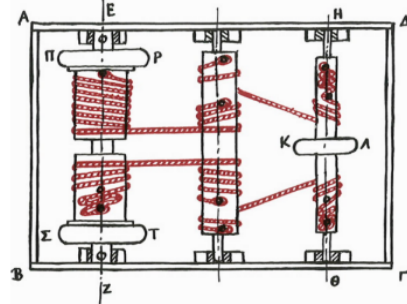


Σχήμα 1.1 Το αυτόματο θέατρο του Ήρωνος. Υπολογίζεται ότι σχεδιάστηκε κατά το διάστημα 1ο π.Χ. αι - 1ο μ.Χ. αι. Αποτελείται από πλήθος επαναπρογραμματιζόμενων και εξειδικευμένων κατασκευών, ώστε να επαναδιαμορφώνεται σε διαφορετικές παραστάσεις.

(A) Επαναδιαμορφούμενος κύλινδρος



(B) Εξειδικευμένο σύστημα κυλίνδρων



Σχήμα 1.2 Α) Ιδιοφυείς προγραμματιζόμενοι κύλινδροι με αριστερόστροφες, δεξιόστροφες και ελεύθερες περιελίξεις νήματος σε πύρους. Β) Εξειδικευμένος μηχανισμός κίνησης: Τροχοκιβώτιο με μηχανισμό ανεξάρτητης κίνησης των τροχών.

λιθοβολισμού (μοναδική βάρβαρη πράξη σε όλη την ελληνική ιστορία), παραπλή-
νησε τα πλοία τους με φωτεινά σήματα και τα έστρεψε προς τις ακτές, όπου
βούλιαξαν άτακτα. Το συγκεκριμένο κατασκευάσμα φέρει πλήθος τεχνολογι-
κών αυτοματισμών που επιτρέπουν την αυτοματοποιημένη ροή εκτέλεσης των
θεατρικών πράξεων. Ο Ήρων στο βιβλίο του περιγράφει με μεγάλη λεπτομέρεια
τους πολύπλοκους μηχανισμούς που συνθέτουν το θέατρο, όπως πόρτες που
ανοιγοκλείνουν αυτόματα, κρήνες διακοσμημένες με πουλιά που κινούνται και
τραγουδούν καθώς και παραγωγή διάφορων ήχων, από τύμπανα και κύμβαλα
έως κεραυνούς. Οι κατασκευές αυτές θεωρούνται ιδιοφυείς διότι ήταν προγραμ-

ματιζόμενες δηλαδή μπορούσαν με ανάλογες ρυθμίσεις να εκτελούν διαφορετικές λειτουργίες. Στο σύνολό του, το θέατρο μπορούσε να επαναδιαμορφωθεί για άλλη θεατρική παράσταση, εφόσον διέθετε για κάθε λειτουργία του επαναδιαμορφούμενους κυλίνδρους και εξειδικευμένα συστήματα κυλίνδρων, όπως αυτά που απεικονίζονται στο Σχήμα 1.2.

Η δημιουργική συνύπαρξη νέας ιδέας, εξειδίκευσης και επαναχρησιμοποίησης θα πρέπει να αποτελεί βασικό πυλώνα για την υποστήριξη μίας διδακτορικής διατριβής στο πλαίσιο έρευνας των πλέον σύγχρονων τεχνολογικών προβλημάτων της επιστήμη πληροφορικής και επικοινωνιών, την επιστήμη χαρακτηρισμού του 21ου αιώνα. Η παρούσα διατριβή στοχεύει σε αυτή τη κατεύθυνση.

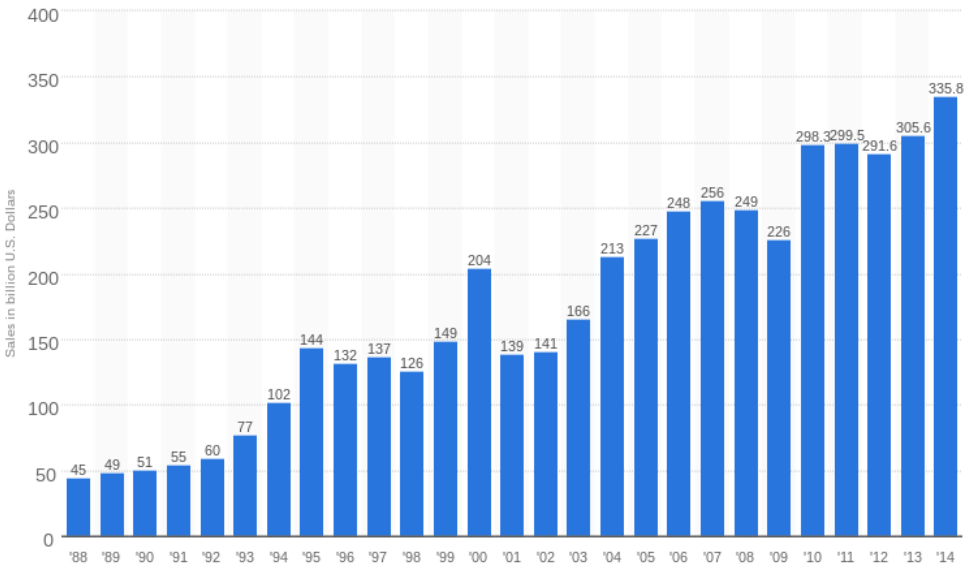
1.2. Περιγραφή Χώρου Έρευνας

Τα ολοκληρωμένα κυκλώματα (IC) αποτελούν το θεμέλιο λίθο των εξελίξεων τόσο στη δια-επιστημονική κοινότητα όσο και στη καθημερινή ζωή του σύγχρονου ανθρώπου. Ο αντίκτυπος αυτών των εξελίξεων είναι τόσο ισχυρός, ώστε πλέον θεωρούνται δεδομένες οι τεχνολογικές μεταβάσεις ανά χρονική περίοδο. Οι καταναλωτές έχουν μάθει να περιμένουν όλο και πιο εξελιγμένα ηλεκτρονικά προϊόντα, ενώ ταυτόχρονα ο επιχειρηματικός κόσμος αναμένει αύξηση της παραγωγικότητας μέσω της βελτίωσης της τεχνολογίας των πληροφοριακών συστημάτων. Παράλληλα, προσεγγίζοντας την μακροοικονομική κλίμακα της τάσης αυτής, ως φαίνεται, η ωριμότητα της βιομηχανίας ημιαγωγών αποτελεί τον καταλύτη της παραγωγικότητας και την ανάπτυξης σε όλους σχεδόν τους τομείς της οικονομικής δραστηριότητας, σημειώνοντας ήδη μία αγορά της τάξεως των 3 τρισεκατομμυρίων δολαρίων [7], όπως φαίνεται και στο Σχήμα 1.3.

Αυτή η εντυπωσιακή αγορά, σύμφωνα με προβλέψεις [8], υπόσχεται να συνδέσει 2.67 δισεκατομμύρια συσκευές στο διαδίκτυο έως το τέλος του 2017 (από 1.11 σήμερα), αποκλειστικά στα πλαίσια των επερχόμενων "έξυπνων πόλεων", και 25 δισεκατομμύρια συσκευές συνολικά στα πλαίσια κοινωνικών, πολιτιστικών και οικονομικών εξελίξεων έως το τέλος του 2020 (από 4.8 σήμερα). Είναι σαφές ότι αυτά τα θορυβώδη νούμερα τροφοδοτούνται από την διασφάλιση κλιμάκωσης της τεχνολογίας ημιαγωγών με τις ανάγκες τις αγοράς.

Τα τελευταία 50 χρόνια της -γραμμικής σχεδόν- αύξησης της υπολογιστικής ισχύος και της αντίστοιχης μείωσης της κατανάλωσης ενέργειας, αποτέλεσαν το υπομόχλιο αυτής της πεποίθησης. Ωστόσο αυτή η γραμμικότητα τροφοδοτούταν από τη συσχέτιση δύο τεχνολογικών αρχών κλιμάκωσης, το νόμο του Moore [9] και το νόμο του Dennard [10]. Ο πρώτος, ο οποίος ισχύει έως σήμερα², υπόσχεται τον διπλασιασμό του αριθμού των στοιχειωδών υλικών (π.χ. τρανζίστορ) που συνιστούν ένα ολοκληρωμένο σύστημα κάθε 18 περίπου μήνες. Πρακτικά αυτός ο νόμος επέτρεψε την αύξηση της υπολογιστικής ισχύος ανά γενεά τεχνολογικής ολοκλήρωσης. Ο δεύτερος νόμος, ο οποίος έπαψε να ισχύει περίπου το 2007, περιέγραφε τις διαστάσεις και το ηλεκτρικά χαρακτηριστικά ενός τρανζί-

²Καθώς η ερμηνεία του νόμου το Moore δεν είναι μοναδική, υπάρχουν απόψεις ότι διανύουμε τη χρονική περίοδο όπου ο νόμος αυτός θα πάψει να ισχύει άμεσα ή έχει ήδη πάψει [11, 12].

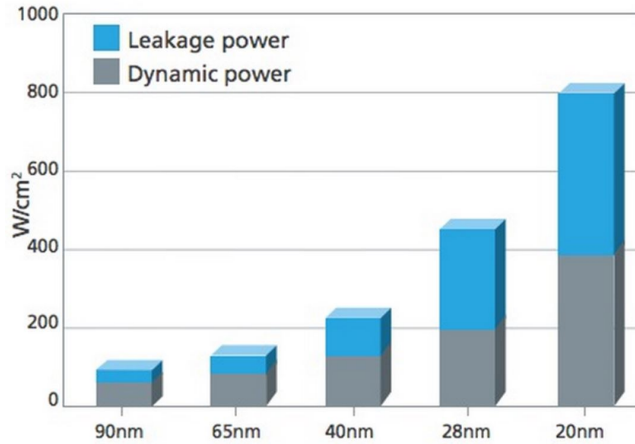


Σχήμα 1.3 Παγκόσμιες πωλήσεις ημιαγωγών 1988-2014 (σε δισεκατομμύρια δολάρια). Πηγή: World Semiconductor Trade Statistics, [7].

στον ώστε να καταστεί δυνατή η διαδοχική συρρίκνωση αυτών, προκειμένου να βελτιώνεται ταυτόχρονα η πυκνότητα της επιφάνειας πυριτίου, η ταχύτητα και η ενεργειακή απόδοση. Πρακτικά ο νόμος αυτός επέτρεψε τη σταθερή διατήρηση της κατανάλωσης ενέργειας, σε κάθε μετάβαση νέας γενιάς ολοκλήρωσης, κατά την οποία ο αριθμός των τρανζίστορ διπλασιαζόταν. Επίσης λόγω της κλιμάκωσης της τεχνολογίας, ο νόμος αυτός προέβλεπε αύξηση στη συχνότητα επεξεργασίας έως 40%. [13].

Ο συνδυασμός αυτών των δύο νόμων, επέτρεψε για τέσσερις σχεδόν δεκαετίες την κλιμάκωση της τάσης τροφοδοσίας και τάσης κατωφλίου κατά περίπου τον ίδιο παράγοντα με τη κλιμάκωση ολοκλήρωσης του μεγέθους των τρανζίστορ. Με αυτό το τρόπο οι σχεδιαστές ήταν σε θέση να λάβουν μια ανάλογη μείωση της ενέργειας τροφοδοσίας ανά τρανζίστορ, έτσι ώστε η πυκνότητα ισχύος (κατανάλωση ενέργειας ανά επιφάνεια πυριτίου) παρέμενε περίπου σταθερή, μεταβαίνοντας από τον έναν τεχνολογικό κόμβο στον επόμενο.

Ωστόσο, μεταβαίνοντας σταδιακά σε τεχνολογικούς κόμβους μερικών νανομέτρων (deep sub-micron technology nodes), η μείωση της τάσης κατωφλίου οδήγησε σε εκθετική αύξηση της ενέργειας διαρροής ρεύματος. Το Σχήμα 1.4 απεικονίζει την κλιμάκωση της στατικής και της δυναμικής κατανάλωσης ενέργειας σε ολοκληρωμένα κυκλώματα, αναφορικά προς τον τεχνολογικό κόμβο. Όπως φαίνεται, για ένα σύνολο τυπικών σχεδιασμών ολοκληρωμένων κυκλωμάτων, η δυναμική ενέργεια κατανάλωσης στον τεχνολογικό κόμβο των 20 nm έχει σχεδόν ξεπεράσει την αντίστοιχη στατική. Αυτή η παρατήρηση οφείλεται στο γεγονός ότι η τάση κατωφλίου δεν είναι πλέον κλιμακούμενη σε μικρότερες

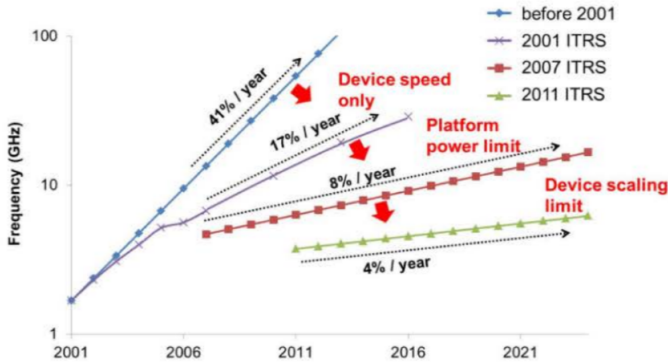


Σχήμα 1.4 Κλιμάκωση στατικής και δυναμικής κατανάλωσης ενέργειας σε ολοκληρωμένα κυκλώματα αναφορικά προς τον τεχνολογικό κόμβο. Πηγή: Mentor Graphics 2014.

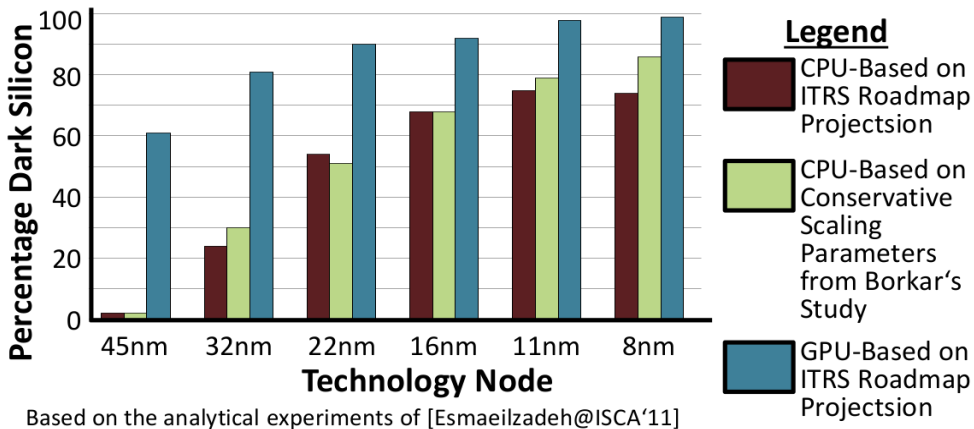
τεχνολογίες ολοκλήρωσης και κατά συνέπεια η τάση τροφοδοσίας δεν μπορεί να κλιμακωθεί περαιτέρω, χωρίς να επηρεάζεται η απόδοση. Αυτή η μείωση της απόδοσης απεικονίζεται στο Σχήμα 1.5. Συγκεκριμένα αυτό το σχήμα απεικονίζει τις προβλέψεις του οργανισμού ITRS για τη κλιμάκωση της μέγιστης συχνότητας λειτουργίας, ανά ταχτές χρονολογικές μελέτες. Ως φαίνεται, οι προβλέψεις κάθε νέας μελέτης είναι απαισιόδοξες όσον αφορά τη κλιμάκωση της συχνότητας επεξεργασίας, αναφέροντας μια αναμενόμενη αύξηση αυτής κατά 41%/χρόνο το 2001, η οποία μειώθηκε στο ποσοστό 4% το 2011.

Συνεπώς, παρόλο που η σύγχρονη τεχνολογία κατασκευής ημιαγωγών εξακολουθεί να υπόσχεται τη σχεδίαση ολοκληρωμένων κυκλωμάτων με περισσότερα τρανζίστορ ανά επιφάνεια πυριτίου σε κάθε τεχνολογικό κόμβο, η ενέργεια τροφοδοσίας ανά τρανζίστορ δεν κλιμακώνεται με ανάλογο τρόπο, και ως εκ τούτου η πυκνότητα ισχύος αυξάνεται. Το φαινόμενο αυτό γίνεται ακόμα πιο έντονο σε συνδυασμό με τα φυσικά όρια που επιβάλλονται από τα υλικά συσκευασίας και την τεχνολογία ψύξης για τη μέγιστη ισχύ και μέγιστη πυκνότητα ισχύος. Το σύνολο της περιγραφής αυτής αποτελεί την εισαγωγή στην εποχή της “Σκοτεινής Σιλικόνης” (Dark Silicon) [13–16].

Ο νέος περιορισμός που επιβάλλεται από τη “Σκοτεινή Σιλικόνη” είναι η έλλειψη ικανότητας ταυτόχρονης ενεργοποίησης όλων των διαθέσιμων τρανζίστορ για δεδομένη θερμική ισχύ σχεδιασμού (TDP). Η τιμή TDP είναι το μέγιστο ποσό της ενέργειας που μπορεί να παρέχεται στο ολοκληρωμένο κύκλωμα προκειμένου αυτό να λειτουργεί σε ασφαλή κατάσταση, δηλαδή, σε θερμικά ασφαλή θερμοκρασία. Σε περίπτωση που η τιμή TDP παραβιάζεται, το ολοκληρωμένο κύκλωμα παράγει θερμότητα με ταχύτερο ρυθμό από αυτόν που μπορεί να διαχέεται στο σύστημα ψύξης. Όπως τονίζεται στη εργασία [14], βάση των προβλέψεων του οργανισμού ITRS και της εταιρείας Intel, στον τεχνολογικό κόμβο



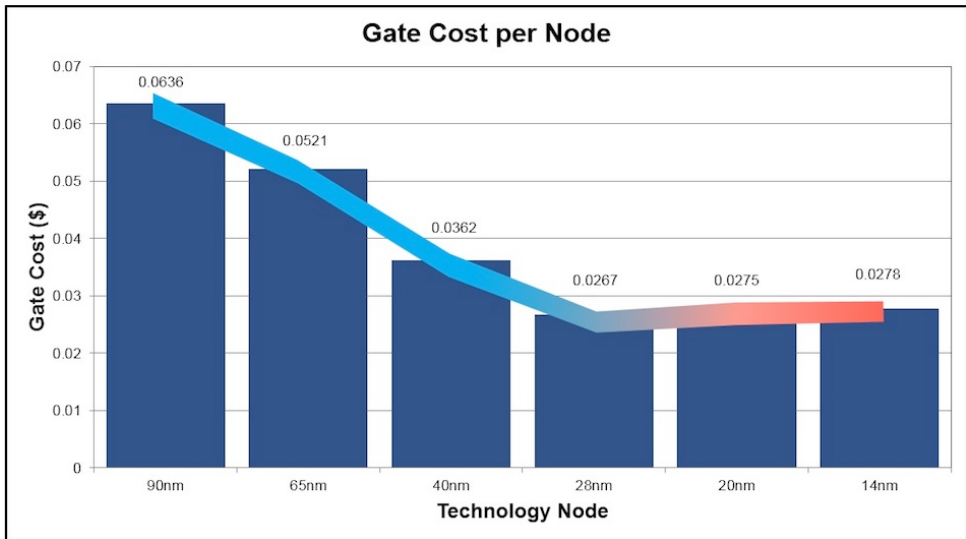
Σχήμα 1.5 Προβλέψεις του οργανισμού ITRS για τη κλιμάκωση της μέγιστης συχνότητας λειτουργίας ανά τακτές χρονολογικές μελέτες.



Σχήμα 1.6 Τάσεις της “Σκοτεινής Σιλικόνης” για διαφορετικούς κόμβους τεχνολογίας ολοκλήρωσης [14].

των 8nm, περισσότερο από το 50% της επιφάνειας πυριτίου αναμένεται να μην μπορεί να χρησιμοποιηθεί ταυτόχρονα με το υπόλοιπο πυρίτιο, όπως φαίνεται και στο Σχήμα 1.6 [13, 17].

Επίσης, ένας ακόμα περιορισμός της σύγχρονης βιομηχανίας ημιαγωγών, αναφέρεται στο οικονομικό κόστος κατασκευής ολοκληρωμένων κυκλωμάτων. Οι σύγχρονες σχεδιαστικές μέθοδοι λιθογραφίας απαιτήσαν συνεχείς καινοτόμες λύσεις για κάθε νέο τεχνολογικό κόμβο (high-k, Metal Gate, Strain, SiGe, Tri-gate, κλπ.) και αναμένεται οι επόμενοι κόμβοι να απαιτήσουν αντίστοιχες καινοτομίες έρευνας και υλοποίησης, αυξάνοντας το κατασκευαστικό κόστος [18]. Εκτός της ερευνητικής κοινότητας πολλές εταιρείες της βιομηχανίας ημιαγωγών έχουν αναγνωρίσει αυτό το πρόβλημα, π.χ. η Nvidia [19]. Το Σχήμα 1.7 απεικονίζει το σημερινό τυπικό κατασκευαστικό κόστος ανά πύλη για τους πρόσφατους κατασκευαστικούς κόμβους. Ως φαίνεται, κατά το παρελθόν (90 nm,



Σχήμα 1.7 Κλιμάκωση κατασκευαστικού κόστους ανά πύλη για διαφορετικούς κόμβους τεχνολογίας ολοκλήρωσης. Πηγή: IBS 2012

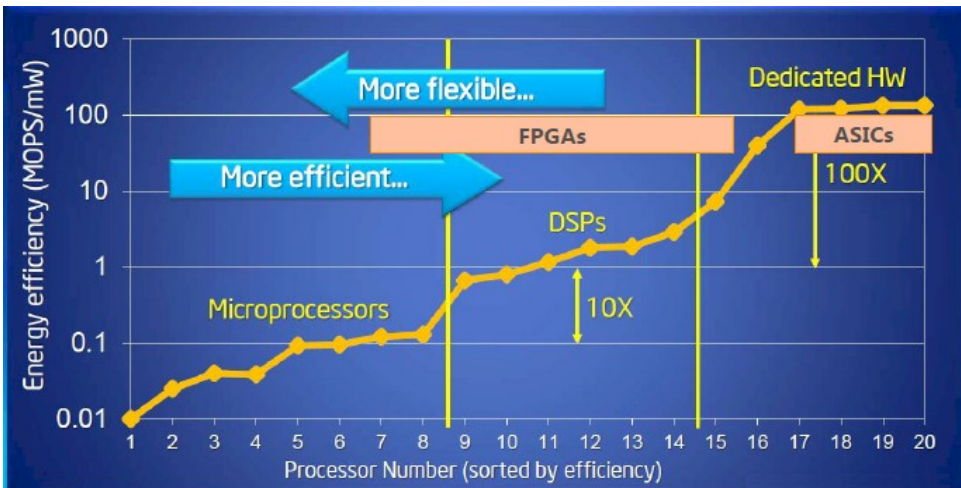
65 nm, 45 nm, 28 nm), το κόστος ανά πύλη σε κάθε κόμβο αυξανόταν. Ωστόσο ο αριθμός πυλών ανά wafer αυξανόταν. Η αντίθεση αυτή ήταν σχεδόν γραμμικά αμφιμονοσήμαντη, με αποτέλεσμα το μέσο κόστος ανά wafer να παραμένει σταθερό. Ωστόσο ο τεχνολογικός κόμβος των 20 nm φαίνεται ότι αυξάνει το κόστος σε σχέση με προηγούμενες τεχνολογικές γενεές, οπότε η κατασκευή ολοκληρωμένων κυκλωμάτων σε νέους κόμβους, πιθανών, είναι ασύμφορη σε σχέση με υφιστάμενους, ή ακόμα και παρελθόντες. Η τρισδιάστατη τεχνολογία ολοκλήρωσης (3-D) και η ετερογένεια έχουν χαρακτηριστεί ως λύσεις στο πρόβλημα της κλιμάκωσης του κόστους [20].

Η ερευνητική κοινότητα έχει ήδη προτείνει ένα αξιόλογο πλήθος εναλλακτικών προσεγγίσεων για το παραπάνω πρόβλημα, όπως οι ετερογενείς πολυεπεξεργαστικές αρχιτεκτονικές [21–26], η προσεγγιστική πληροφορική (Approximate Computing) [27–33] και η διαχείριση ενέργειας για αρχιτεκτονικές “Σκοτεινής” και “Αμυδρά Φωτεινής” σιλικόνης [34–40]. Από τα πρόσφατα δημοσιευμένα αποτελέσματα φαίνεται πως η πιο αποτελεσματική προσέγγιση είναι η δημιουργία ετερογενών και εξειδικευμένων επιταχυντών υλικού οι οποίοι αυξάνουν την υπολογιστική δύναμη ανά μονάδα κατανάλωσης ενέργειας και επιφάνειας πυριτίου. Ωστόσο σε αυτή τη κατεύθυνση, στα πλαίσια αυτής της διατριβής αναγνωρίζονται τέσσερις βασικοί περιορισμοί για αυτές τις πολυσύνθετες αρχιτεκτονικές: α) το πρόβλημα προγραμματισμού, β) το πρόβλημα κατάτμησης υλικού/λογισμικού, γ) το πρόβλημα κόστους/χρόνου σχεδίασης και δ) το πρόβλημα εύρεσης κατάλληλων εργαλείων CAD βέλτιστης σχεδίασης. Αναγνωρίζοντας αυτούς του περιορισμούς, προτείνεται η οργάνωση ενός πλαισίου που προσπαθεί μεθοδολογικά να συμβάλει στα διάφορα επίπεδα ανάπτυξης αποτελεσματικών

συστημάτων για την εποχή της “Σκοτεινής Σιλικόνης”.

1.3. Επισκόπηση Διδακτορικής Διατριβής

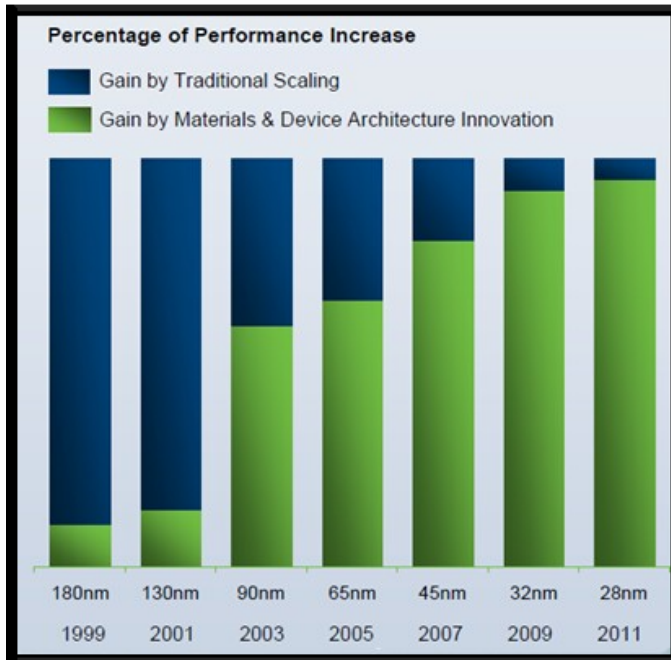
Το προηγούμενο εισαγωγικό εδάφιο αναδεικνύει την επιτακτική ανάγκη για μεθόδους, εργαλεία και αρχιτεκτονικές που προσφέρουν βιωσιμότητα στην συνεχή κλιμάκωση της τεχνολογίας και της ενέργειας. Το ιστορικό σχεδίασης υπολογιστικών συστημάτων των τελευταίων χρόνων έχει δείξει ότι κλιμάκωση στην ενέργεια και στην τεχνολογία μπορεί να επιτευχθεί με την εξειδίκευση της αρχιτεκτονικής και των υλικών/τεχνολογίας κατασκευής αντίστοιχα.



Σχήμα 1.8 Σύγκριση σχεδιασμών ολοκληρωμένων κυκλωμάτων διαφορετικής αρχιτεκτονικής προσέγγισης, αναφορικά προς την ενεργειακή τους απόδοση. Πηγή: Bob Broderson, Berkeley Wireless group, ISSCC proceedings, Microsoft, 2011.

Το Σχήμα 1.8 απεικονίζει ένα σύνολο σχεδιασμών ολοκληρωμένων κυκλωμάτων διαφορετικής αρχιτεκτονικής προσέγγισης, αναφορικά προς την ενεργειακή τους απόδοση. Είναι φανερό ότι οι εξειδικευμένες αρχιτεκτονικές μπορούν να οδηγήσουν σε αύξηση της επεξεργαστικής ισχύς ανά μονάδα ενέργειας (MOPS/mW) κατά παράγοντα 100x, σε σχέση με τους συμβατικούς επεξεργαστές γενικού σκοπού. Επίσης τονίζεται ότι οι συσκευές FPGAs αποτελούν μία ευέλικτη πλατφόρμα ενεργειακής απόδοσης, ανάλογα με το προγραμματισμό τους. Ωστόσο η μετάβαση σε εξειδικευμένες ή/και επαναδιαμορφούμενες αρχιτεκτονικές αντιστοιχεί σε αυξημένη δυσκολία σχεδίασης και προγραμματισμού λογισμικού αυτών.

Το Σχήμα 1.9 απεικονίζει το ποσοστό αύξησης απόδοσης που επιφέρει η συμβατική τεχνολογική κλιμάκωση και η καινοτομία αρχιτεκτονικής και υλικών για κάθε τεχνολογικό κόμβο. Ως φαίνεται, η βελτίωση απόδοσης των πρόσφατων τεχνολογικών κόμβων οφείλεται σε μικρό ποσοστό στη συμβατική τεχνολογική

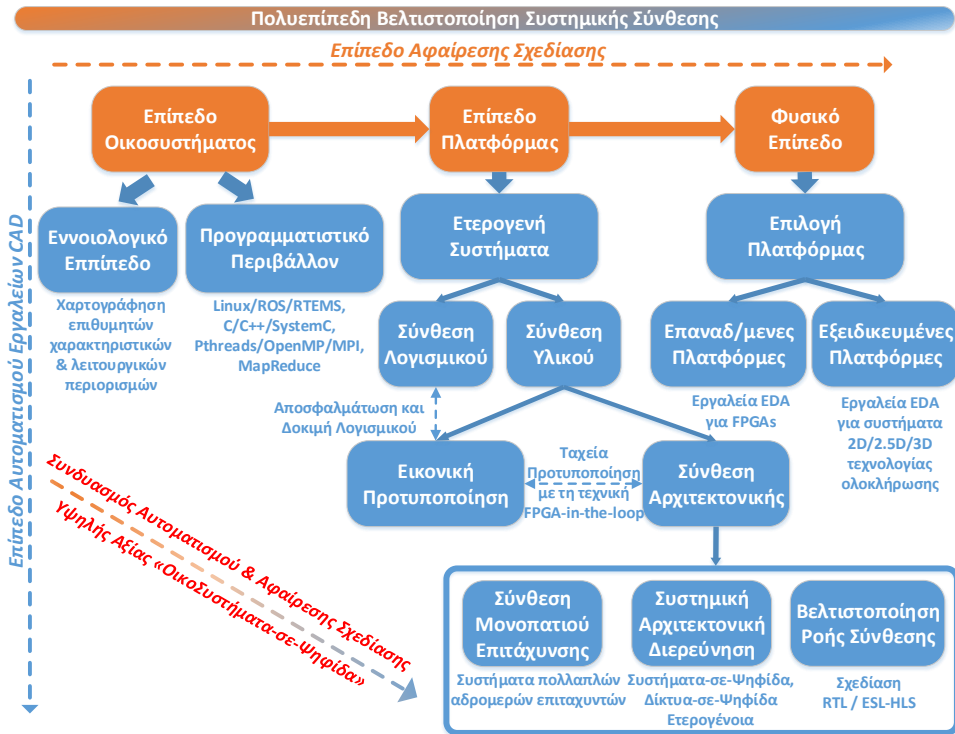


Σχήμα 1.9 Ποσοστό αύξησης απόδοσης της συμβατικής τεχνολογικής κλιμάκωσης και της καινοτομίας αρχιτεκτονικής και υλικών, για κάθε τεχνολογικό κόμβο. Πηγή: IBM Microelectronics, Intel IC Insights, 2012.

κλιμάκωση, ενώ το μεγαλύτερο κέρδος καταλογίζεται σε καινοτόμες τεχνικές βελτίωσης της αρχιτεκτονικής και των υλικών κατασκευής.

Προς τη κατεύθυνση των δύο προαναφερθέντων κινητήρων τοποθετείται η παρούσα διδακτορική διατριβή, ένα ολιστικό μεθοδολογικό πλαίσιο της οποίας αποτυπώνεται στο Σχήμα 1.10. Η στόχευση της διατριβής υποθέτει τη σύλληψη μιας νέας ιδέας, η οποία αναπτύσσεται στους πρόποδες ενός νέου προϊόντος/υπηρεσίας υλικού/λογισμικού. Αφετηρία των ανεπτυγμένων μεθοδολογικών συνιστωσών αποτελεί το εννοιολογικό επίπεδο, δηλαδή εκείνο το στάδιο στο οποίο αποτυπώνονται τα επιθυμητά χαρακτηριστικά ενός νέου προϊόντος ή υπηρεσίας πληροφοριακού συστήματος. Συνεχίζει δε, με τη χαρτογράφηση των επιθυμητών προδιαγραφών του ανεπτυγμένου προϊόντος καθώς και των λειτουργικών περιορισμών. Σημαντικό στάδιο σε αυτό το επίπεδο αποτελεί η οργάνωση των εργασιών. Τελευταία, έχει αναπτυχθεί πλήθος εξειδικευμένων εργαλείων διαχείρισης εργασιών, στα πλαίσια ενός έργου υλικού/λογισμικού, τα οποία ανάλογα με την πολυπλοκότητα, μπορούν να διαχειριστούν διάφορες φάσεις της παραγωγικής διαδικασίας όπως η εκτίμηση και ο σχεδιασμός, ο προγραμματισμός, ο έλεγχος του κόστους, η διαχείριση του προϋπολογισμού, η κατανομή των πόρων, η επικοινωνία των συνεργατών, η λήψη αποφάσεων, η διαχείριση ποιότητας και η οργάνωση των διοικητικών και τεχνικών εγγράφων (π.χ. [41, 42]).

Έχοντας προδιαγράψει τα ζητούμενα χαρακτηριστικά λειτουργίας, η ομάδα



Σχήμα 1.10 Επισκόπηση διδακτορικής διατριβής.

ανάπτυξης καλείται να τυποποιήσει το περιβάλλον λειτουργίας στο οποίο επρόκειτο να είναι λειτουργικό το προς-ανάπτυξη προϊόν. Αυτό το επίπεδο (Προγραμματιστικό Περιβάλλον) είναι αποτέλεσμα νέας αντιμετώπισης της παραγωγικής διαδικασίας και είναι υπεύθυνο για την προδιαγραφή όλων εκείνων των περιβαλλοντικών στοιχείων με τα οποία θα αλληλεπιδρά το ανεπτυγμένο προϊόν. Το Εννοιολογικό Επίπεδο μαζί με το Προγραμματιστικό Περιβάλλον συνιστούν το Επίπεδο Οικοσυστήματος το οποίο μέσα από τον πρόωρο εντοπισμό των προαναφερθέντων χαρακτηριστικών του τελικού συστήματος εξασφαλίζει τη προσαρμογή της ροής σχεδιασμού στις νέες τάσεις της αγοράς, όπου πλέον η ανάπτυξη οδηγείται από την ίδια την αγορά και όχι απαραίτητα από τις τεχνολογικές εξελίξεις [43].

Το επόμενο επίπεδο αφορά την ανάπτυξη της πλατφόρμας αναφοράς (Επίπεδο Πλατφόρμας), όπου πλέον σχηματίζονται οι απεικονίσεις των χαρακτηριστικών λειτουργίας στο πραγματικό σύστημα. Το στάδιο αυτό εκκινεί με το διαχωρισμό υλικού/λογισμικού βάση μιας διαδικασίας ανάλυσης (profiling) που στοιχειοθετεί τα εγγενή χαρακτηριστικά της εφαρμογής και βάση αυτών ταξινομεί τις συνολικές λειτουργικές διαδικασίες είτε σε εκτέλεση μέσω λογισμικού, είτε σε επιτάχυνση μέσω εξειδικευμένου υλικού. Ο συν-σχεδιασμός υλικού/λογισμικού εμπλουτίζεται με την ταχεία προτυποποίηση, μια διαδικασία που εξα-

σφαλίζει γρηγορότερη ανάπτυξη του προϊόντος καθώς επιτρέπει την ανάπτυξη του λογισμικού καθ' όλη τη διάρκεια ανάπτυξης του υλικού, κάνοντας χρήση εικονικών μονάδων υλικού που προσομοιώνουν το εξειδικευμένο υλικό. Σε σύγκριση με τις υφιστάμενες λύσεις προτυποποίησης, η προτεινόμενη μεθοδολογία υπερτερεί αναφορικά με την ταχύτητα, την ακρίβεια και την αποσφαλματωθείσα ολοκλήρωση του σχεδιασμού, καθώς εισαγάγει συσκευές FPGAs στη διαδικασία της προσομοίωσης. Αυτή η ενέργεια επιτρέπει μέρος του υλικού που αναπτύσσεται σε συστοιχία με το λογισμικό, να δοκιμάζεται απευθείας σε πρωτότυπο υλικό (FPGA) κατά τα πρώιμα στάδια σχεδίασης, όπου η συνολική αλυσίδα εκτέλεσης του συστήματος δεν είναι ολοκληρωμένη.

Η ανάπτυξη του υλικού λαμβάνει χώρα στο στάδιο της αρχιτεκτονικής σύνθεσης, και είναι ορθογωνίως ανεξάρτητη του σταδίου της εικονικής προτυποποίησης, με την έννοια ότι τα υποσυστήματα που σταδιακά συντίθεται κατά την αρχιτεκτονική σύνθεση μπορούν να χρησιμοποιηθούν στην εικονική προτυποποίηση αντικαθιστώντας τις εικονικές εκδόσεις τους. Με αυτόν τον τρόπο επιταχύνεται η διαδικασία της προσομοίωσης, αλλά και προσφέρεται μεγαλύτερη ακρίβεια στην αποτίμηση των μετρικών του συστήματος. Το στάδιο της αρχιτεκτονικής σύνθεσης συνίσταται στο διαχωρισμό τριών διακριτών μεθοδολογικών αναλύσεων, αυτό της σύνθεσης του μονοπατιού επιτάχυνσης (Acceleration Datapath Synthesis), της συστημικής αρχιτεκτονικής διερεύνησης (System Architecture Exploration) και της βελτιστοποίησης της ροής σύνθεσης (Synthesis Flow Optimization). Κάθε ένα από αυτά τα στάδια πραγματεύεται διαφορετικά προβλήματα της σύγχρονης σχεδίασης ολοκληρωμένων κυκλωμάτων και αφορά τόσο επαναδιαμορφούμενες πλατφόρμες (FPGAs) όσο και εξειδικευμένες (ASICs). Συγκεκριμένα, η σύνθεση του μονοπατιού επιτάχυνσης περιλαμβάνει τις τεχνικές επιτάχυνσης των υπολογιστικά κρίσιμων περιοχών του συστήματος. Προτείνεται η χρήση πολλαπλών επιταχυντών αδρομερούς υφής, καθώς προσφέρουν μεγαλύτερη παραγωγικότητα σε σχέση με την κατανάλωση και ως εκ τούτου αποτελούν ελκυστική λύση στο πρόβλημα της “σκοτεινής σιλικόνης” (Dark Silicon) [17, 44]. Η συστημική αρχιτεκτονική διερεύνηση αναφέρεται στην αναζήτηση των εναλλακτικών τρόπων διασύνδεσης των επί μέρους συστημάτων στο υπό-ανάπτυξη ολοκληρωμένο σύστημα. Στα πλαίσια της παρούσας διατριβής μελετώνται ετερογενή αρχιτεκτονικά πρότυπα που βασίζονται σε σύστημα-σε-ψηφίδα (SoC) καθώς και σε τοπολογίες δικτύου-σε-ψηφίδα (NoC). Το τελευταίο επίπεδο της αρχιτεκτονικής σύνθεσης αναφέρεται στη βελτιστοποίηση της ροής σύνθεσης. Η συγκεκριμένη διαδικασία προβλέπει την εισαγωγή νέων τεχνικών σχεδιασμού, αξιοποιώντας αφαιρετικές μεθόδους υψηλού επιπέδου σύνθεσης (HLS), οι οποίες επιταχύνουν περαιτέρω την αναπτυξιακή ροή, ενώ επιτρέπουν την προγραμματιστική ευελιξία στο μεγάλο πρόβλημα του προγραμματισμού των ετερογενών συστημάτων πολλαπλών επεξεργαστικών στοιχείων. Επιπλέον, σε αυτό το στάδιο προτείνεται ένας ταχύς και αυτοματοποιημένος τρόπος διερεύνησης του χώρου αρχιτεκτονικών λύσεων που συνδυάζει τεχνικές HLS και εικονικής προτυποποίησης προκειμένου να ανακαλύψει ένα σύνολο βέλτιστων λύσεων (Pareto optimal) βάσει ενός πολυ-κριτηριακού περιβάλλοντος ανάλυσης.

Το τελευταίο επίπεδο (Physical Layer) του ολιστικού προτεινόμενου μεθο-

δολογικού πλαισίου αναφέρεται στη φυσική απεικόνιση των ανεπτυγμένων μονάδων, μέσα από τις προηγούμενες διαδικασίες, σε πραγματικές πλατφόρμες υλοποίησης. Οι επί μέρους αναπτυχθείσες διαδικασίες επιτρέπουν την τοποθέτηση των σχεδιασμών είτε σε επαναδιαμορφούμενες πλατφόρμες, είτε σε εξειδικευμένες, ανάλογα με τα ζητούμενα χαρακτηριστικά που έχουν αποτυπωθεί στο επίπεδο οικοσυστήματος, αλλά και των περιορισμών που θέτει η προκείμενη αγορά. Συγκεκριμένα, η απεικόνιση σε FPGAs εμπλουτίζεται μέσα από εξειδικευμένους αλγορίθμους χωροθέτησης, οι οποίοι χρησιμοποιώντας τεχνικές από τη φυσική επιλογή και την εξελικτική θεωρία, προσφέρουν βέλτιστες λύσεις αναφορικά προς τη μέγιστη επιτευχθείσα συχνότητα λειτουργίας, το μήκος καλωδίου, την επιφάνεια πυριτίου και την κατανάλωση ενέργειας, σε μικρό χρόνο επεξεργασίας. Αυτό το αποτέλεσμα επιτυγχάνεται μέσω της οριζόντιας και κάθετης παραλληλοποίησης των προτεινόμενων αλγορίθμων και του πολυκριτηριακού μηχανισμού βελτιστοποίησής τους. Παράλληλα, η απεικόνιση σε εξειδικευμένες πλατφόρμες (ASICs) ενισχύεται μέσα από προτεινόμενες διαδικασίες, οι οποίες μεταβάλλοντας την μικρο-αρχιτεκτονική, επιτυγχάνουν τη μείωση της εκλυόμενης θερμοκρασίας και κατ'επέκταση την μείωση του ρυθμού γήρανσης του ολοκληρωμένου, με συνεπάγουσα αύξηση της αξιοπιστίας του κυκλώματος. Επιπλέον, μελετούνται τα κέρδη από την μετάβαση στην νέα υποσχόμενη τρισδιάστατη τεχνολογία ολοκλήρωσης (3D), για την οποία, μέχρι σήμερα, υπάρχει έλλειψη σχεδιαστικών εργαλείων, τόσο ακαδημαϊκών όσο και βιομηχανικών. Προς αυτή τη κατεύθυνση προτείνεται μία ροή σχεδιασμού που βασίζεται σε αξιόπιστα συμβατικά εργαλεία σχεδίασης (2D) και καταφέρνει να ποσοτικοποιήσει τα κέρδη υλοποίησης από τη μετάβαση στη τρισδιάστατη ολοκλήρωση. Τα αναπτυσσόμενα εργαλεία λογισμικού είναι συμβατά με τους τύπους αρχείων των βιομηχανικών προτύπων, οπότε και προσφέρουν μεγάλη ευελιξία στην υιοθέτησή τους από εναλλακτικές ροές σχεδίασης και εργαλεία διαφορετικών βιομηχανικών προμηθευτών.

Στο σύνολό της, η παρούσα διδακτορική διατριβή παρέχει μεθοδολογίες και εργαλεία σε διάφορα επίπεδα της σχεδίασης ολοκληρωμένων συστημάτων υλικού/λογισμικού. Μέσα από την δομημένη ροή σχεδίασης, γίνεται προσπάθεια συννεκτικής αλληλεπίδρασης των επί μέρους σχεδιαστικών μονοπατιών, προκειμένου να επιτευχθεί η ζεύξη στα χαρακτηριστικά που αναζητά η αγορά και σε αυτά που μπορεί να προσφέρει η τεχνολογία. Βασικός συνδετικός κρίκος σε αυτή τη προσπάθεια, αποτελεί ο σαφής εννοιολογικός διαχωρισμός των επιπέδων σχεδίασης και ο καθορισμός, μεταξύ αυτών, συγκεκριμένων και μη-επικαλυπτόμενων μηχανισμών βελτιστοποίησης. Η ολοκληρωμένη εισήγηση στη προτεινόμενη ροή σχεδιασμού στοχεύει στην επίτευξη σχεδίασης συστημάτων που μπορούν να αναπτυχθούν μέσα από ένα ταχύ, πολυ-κριτηριακά βελτιστοποιημένο περιβάλλον, ενσωματώνοντας μεγάλο βαθμό συστημικής πολυπλοκότητας. Ο βαθμός αυτός ολοκλήρωσης συντελεί στην ανάδειξη μια νέας γενιάς ολοκληρωμένων συστημάτων, στα οποία, στα πλαίσια της παρούσας διδακτορικής διατριβής, αποδίδεται ο όρος “οικοσυστήματα-σε-ψηφίδα” (Ecosystems-on-Chip, ESoC).

1.3.1. Ερευνητικά Προβλήματα και Συνεισφορά

Στο εδάφιο αυτό σημειώνονται τα ανοιχτά προβλήματα στις υφιστάμενες μεθοδολογίες σχεδιασμού και οι αντίστοιχες προτεινόμενες καινοτομικές λύσεις. Η ερευνητική εργασία στοχεύει στη παροχή μεθοδολογιών σε ένα ευρύ φάσμα της σχεδίασης εξειδικευμένων συστημάτων υλικού, εκκινώντας από τις απαιτήσεις του περιβάλλοντος λειτουργίας του υπό-σχεδιασμού συστήματος και καταλήγοντας στις φυσικές πλατφόρμες σχεδίασης (FPGAs/ASICs), λαμβάνοντας υπόψιν τους περιορισμούς που επιβάλλει η αρχιτεκτονική αυτών καθώς και η τεχνολογία ολοκλήρωσης. Σήμερα, οι υπάρχουσες μέθοδοι σχεδίασης αδυνατούν να αντεπεξέλθουν στις προκλήσεις που θέτει η αγορά και η τεχνολογία. Αυτός ο ισχυρισμός βασίζεται στα παρακάτω διακριτά ερευνητικά προβλήματα:

1. Οι αυστηρές απαιτήσεις της βιομηχανίας για μικρούς χρόνους εισαγωγής του προϊόντος στην αγορά (time-to-market) αποτρέπουν την ευρεία εισήγηση καινοτόμων αρχιτεκτονικών προτύπων, καθώς τα αντίστοιχα πολύπλοκα προγραμματιστικά περιβάλλοντα είναι περιορισμένης διάδοσης και δυνατοτήτων σε σχέση με τα συμβατικά περιβάλλοντα και αρχιτεκτονικές.
2. Οι λύσεις προτυποποίησης αδυνατούν να συν-προσομοιώσουν εικονικό και πραγματικό υλικό. Συνεπώς παρόλο που δίνουν τη δυνατότητα για δοκιμή του αναπτυχθέντος λογισμικού σε μια αφαιρετική εικονική μηχανή, υπάρχει έλλειψη δοκιμής σε πραγματικό σύστημα, κατά τη διάρκεια που αυτό σταδιακά αναπτύσσεται με το λογισμικό. Η προσέγγιση αυτή συχνά οδηγεί σε μεγάλες αποκλίσεις του υλικού/λογισμικού από τις επιθυμητές προδιαγραφές και προκαλεί αλλαγές στα τελικά στάδια της διαδικασίας σχεδίασης. Το αποτέλεσμα αυτό είναι ευρέως διαδεδομένο με τον χαρακτηρισμό “Engineering Change Orders” (ECO) και αποτελεί μεγάλο κίνδυνο για τα περιθώρια κέρδους της βιομηχανίας συστημάτων υπολογιστών.
3. Η ανάγκη για αυξημένη πολυπλοκότητα στα σύγχρονα ολοκληρωμένα συστήματα απαιτεί εργαλεία σχεδίασης με αντίστοιχη ικανότητα διαχείρισης της πολυπλοκότητας. Ωστόσο οι αλγόριθμοι σχεδίασης παρόλο που συνεχώς ενσωματώνουν νέα χαρακτηριστικά βελτιστοποίησης στη σχεδίαση, όπως π.χ. η αξιοπιστία, παραμένουν συντηρητικοί όσον αφορά τους μηχανισμούς κλιμάκωσης στην αυξημένη πολυπλοκότητα. Έτσι προκύπτει το παράδοξο γεγονός, ότι ενώ υπάρχει πληθώρα σύγχρονων πολυπύρηνων επεξεργαστικών συστημάτων, εντούτοις, τα σχεδιαστικά εργαλεία δεν τα αξιοποιούν αποδοτικά. Μία από τις βασικές αιτίες για αυτό το παράδοξο είναι η σειριακή φύση των αλγορίθμων βελτιστοποίησης.
4. Οι τάσεις της αγοράς επιβάλλουν νέα κριτήρια σχεδίασης, τα οποία συχνά είναι αντικρουόμενα μεταξύ τους. Αυτό συμβαίνει διότι τα διαφορετικά κριτήρια σχεδιασμού ικανοποιούνται από εναλλακτικές αρχιτεκτονικές και φυσικές υλοποιήσεις, με μικρό βαθμό συσχέτισης μεταξύ τους. Παράλληλα, τα υπάρχοντα εργαλεία σχεδίασης επεξεργάζονται ομοιογενώς τις εφαρμογές, χωρίς να αναγνωρίζουν τα ιδιαίτερα χαρακτηριστικά

τους και συνεπώς αδυνατούν να προσφέρουν οικουμενικές λύσεις που ικανοποιούν πολλά κριτήρια σχεδιασμού ταυτόχρονα, καταλήγοντας συνήθως σε υπό-βέλτιστες λύσεις αναφορικά προς το χώρο σχεδιασμού.

5. Η ιεραρχική μεμονωμένη συν-σχεδίαση υλικού/λογισμικού, όπως έχει καθιερωθεί τα τελευταία χρόνια, βασίζεται αποκλειστικά στη πληροφορία του επεξεργαστικού φόρτου, ενώ επίσης αδυνατεί να μεταφέρει τους περιορισμούς στα διάφορα επίπεδα σχεδιασμού. Αυτό οδηγεί συχνά σε αναποτελεσματική διαδικασία διαχωρισμού του προγραμματιστικού μέρους που πρέπει να απεικονισθεί σε υλικό για την συνολική επιτάχυνση του συστήματος.
6. Το σύνθετες προγραμματιστικό μοντέλο υλοποίησης επιταχυντών υλικού ακολουθεί πιστά τα αντίστοιχα περιβάλλοντα λογισμικού, αδιαφορώντας για τις ιδιαιτερότητες των φυσικών φορέων υλοποίησης (FPGAs, ASICs), των αρχιτεκτονικών υλικού αλλά και της τεχνολογίας ολοκλήρωσης. Οι περιορισμοί που θέτουν αυτοί οι παράγοντες συνήθως λαμβάνονται υπόψιν κατά τις τελευταίες φάσεις του σχεδιασμού, αποκρύπτοντας την ευελιξία που μπορεί να προσφερθεί από την υψηλού-επιπέδου αφαιρετική αντιμετώπισή τους.

Στα Κεφάλαια 2-4 παρουσιάζονται οι λύσεις που προτείνονται στα πλαίσια της παρούσας έρευνας για την επιτυχή αντιμετώπιση των προαναφερθέντων προβλημάτων.

1.3.2. Οργάνωση Κεφαλαίων

Τα προτεινόμενα μεθοδολογικά πλαίσια και τα αντίστοιχα εργαλεία σχεδίασης, τα οποία αναπτύσσονται στους ερευνητικούς άξονες της παρούσας διατριβής, οργανώνονται σύμφωνα με την παρακάτω δομή:

- Στο Κεφάλαιο 1, δηλαδή στο παρόν κεφάλαιο, παρουσιάζονται οι εξελίξεις της τεχνολογίας και οι τάσεις της δυναμικά εξελισσόμενης αγοράς των συστημάτων πληροφορικής και υπολογιστών. Παράλληλα αναλύονται οι περιορισμοί των υφιστάμενων μεθόδων σχεδίασης για την σχεδίαση συστημάτων που προσφέρουν μεσοπρόθεσμη και μακροπρόθεσμη βιωσιμότητα στις αναφερόμενες τεχνολογικές κλιμακώσεις και αγοραστικές τάσεις. Το κεφάλαιο συνοψίζει τους βασικούς άξονες ανάπτυξης της διατριβής, μέσα από την παρουσίαση ενός ολιστικού μεθοδολογικού πλαισίου που ιεραρχεί σε διακριτά επίπεδα τις επί μέρους συνεισφορές και περιγράφει τις συνδετικές σχέσεις μεταξύ αυτών.
- Στο Κεφάλαιο 2, παρουσιάζεται η προτεινόμενη μεθοδολογία ταχείας προτυποποίησης. Εισηγείται η έννοια της εικονικής προτυποποίησης με την τεχνική “FPGA-in-the-loop”, η οποία επιταχύνει τη διαδικασία προσομοίωσης και βοηθά στην αποδοτικότερη διερεύνηση του χώρου λύσεων κατά τη βελτιστοποίηση της αρχιτεκτονικής.

- Στο Κεφάλαιο 3, παρουσιάζεται η πολυεπίπεδη σύνθεση ετερογενών αρχιτεκτονικών εξειδικευμένης αποστολής. Αναφορικά προς την ολιστική μεθοδολογία της διατριβής (Σχήμα 1.10), το κεφάλαιο περιέχει το στάδιο της σύνθεσης του μονοπατιού επιτάχυνσης, της συστημικής αρχιτεκτονικής διερεύνησης και της βελτιστοποίησης της ροής σύνθεσης.
- Στο Κεφάλαιο 4, παρουσιάζονται τα προτεινόμενα εργαλεία σχεδιασμού για επαναδιαμορφούμενες πλατφόρμες. Αναλύεται η τεχνική χαρακτηρισμού και ταξινόμησης των εφαρμογών βάση των εγγενών χαρακτηριστικών τους, κατά τη φάση της σύνθεσης. Επίσης περιγράφεται αναλυτικά ο γενετικός αλγόριθμος τοποθέτησης πολυ-κριτηριακής βελτιστοποίησης, ο οποίος αξιοποιεί τον προαναφερθέν χαρακτηρισμό προκειμένου να κατευθύνει την τοποθέτηση προς βέλτιστα σημεία Pareto ανάλογα με την ταξινόμηση της εφαρμογής.
- Στο Κεφάλαιο 5, προτείνονται νέα αρχιτεκτονικά πρότυπα για κλιμακούμενες υπολογιστικές λύσεις. Στόχευση αυτών των αρχιτεκτονικών είναι η αύξηση της επεξεργαστικής ισχύος ανά μονάδα κατανάλωσης ενέργειας. Παράλληλα με τα αρχιτεκτονικά πρότυπα, προτείνονται τα αντίστοιχα περιβάλλοντα προγραμματισμού αυτών. Το κεφάλαιο 5 συνδέεται με τη γενικότερη μεθοδολογία της διατριβής που παρουσιάζεται στην Ενότητα 1.3, ως προς τη συνεισφορά του στη “Σύνθεση Μονοπατιού Επιτάχυνσης” και τη “Βελτιστοποίηση Ροής Σύνθεσης”
- Στο Κεφάλαιο 6, συνοψίζονται τα ευρήματα της παρούσας διατριβής και τονίζονται τόσο οι σημερινές ερευνητικές ενέργειες καθώς και οι μελλοντικές επεκτάσεις που ανακύπτουν από τη χρήση των προτεινόμενων μεθοδολογιών.

Βιβλιογραφία

- [1] B. Horst-Dieter, *Einführung in das antike theaterwesen - Εισαγωγή στο Αρχαίο Θέατρο* (Μορφωτικό ίδρυμα εθνικής τραπέζης, 1992).
- [2] S. Lonsdale, *Dance and Ritual Play in Greek Religion*, Ancient Society and History (Johns Hopkins University Press, 2000).
- [3] H. C. Baldry, *Το τραγικό θέατρο στην αρχαία Ελλάδα* (Καρδαμίτσα, 2008).
- [4] B. Horst-Dieter, *L' Homme Grec - Ο Έλληνας Άνθρωπος* (Ελληνικά Γράμματα, 1996).
- [5] K. Kotsanas, *Οι εφευρέσεις των αρχαίων Ελλήνων (Αρχαία Ελληνική Τεχνολογία)* (Kostas Kotsanas, 2011).
- [6] D. Kalligeropoulos, *Αυτοματοποιητική Ήρωνα του Αλεξανδρινού* (1996).
- [7] Statista, *Statista statistics portal, global semiconductor sales from october 2011 to february 2015 (in billion u.s. dollars)*, .
- [8] Gartner, *Gartner inc. technology research*, .
- [9] G. Moore, *Cramming more components onto integrated circuits*, *Proceedings of the IEEE* **86**, 82 (1998).
- [10] R. Dennard, F. Gaensslen, H.-N. YU, V. Rideout, E. BASSOUS, and A. R. LEBLANC, *Design of ion-implanted mosfet's with very small physical dimensions*, *Proceedings of the IEEE* **87**, 668 (1999).
- [11] CNET.com, *End of moore's law: It's not just about physics*, .
- [12] Recode.net, *Moore's law hits 50, but it may not see 60*, .
- [13] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, *Power challenges may end the multicore era*, *Commun. ACM* **56**, 93 (2013).
- [14] M. Shafique, S. Garg, J. Henkel, and D. Marculescu, *The eda challenges in the dark silicon era: Temperature, reliability, and variability perspectives*, in *Proceedings of the 51st Annual Design Automation Conference*, DAC '14 (ACM, New York, NY, USA, 2014) pp. 185:1–185:6.
- [15] C. Mack, *Fifty years of moore's law*, *Semiconductor Manufacturing, IEEE Transactions on* **24**, 202 (2011).
- [16] M. Taylor, *Is dark silicon useful? harnessing the four horsemen of the coming dark silicon apocalypse*, in *Design Automation Conference (DAC), 2012 49th ACM/EDAC/IEEE* (2012) pp. 1131–1136.

- [17] N. Hardavellas, M. Ferdman, B. Falsafi, and A. Ailamaki, *Toward dark silicon in servers*, *Micro, IEEE* **31**, 6 (2011).
- [18] A. Mallik, J. Ryckaert, A. Mercha, D. Verkest, K. Ronse, and A. Thean, *Maintaining moore's law: enabling cost-friendly dimensional scaling*, (2015) pp. 94221N–94221N–12.
- [19] J. Hruska, *Nvidia deeply unhappy with tsmc, claims 20nm essentially worthless*, .
- [20] M. D. I. Zvi Or-Bach, *Is the cost reduction associated with ic scaling over?* .
- [21] J. Cong, M. Ghodrati, M. Gill, B. Grigorian, and G. Reinman, *Architecture support for accelerator-rich cmps*, in *Design Automation Conference (DAC), 2012 49th ACM/EDAC/IEEE* (2012) pp. 843–849.
- [22] J. Cong and B. Xiao, *Optimization of interconnects between accelerators and shared memories in dark silicon*, in *Computer-Aided Design (ICCAD), 2013 IEEE/ACM International Conference on* (2013) pp. 630–637.
- [23] N. Goulding-Hotta, J. Sampson, G. Venkatesh, S. Garcia, J. Auricchio, P. Huang, M. Arora, S. Nath, V. Bhatt, J. Babb, S. Swanson, and M. Taylor, *The greendroid mobile application processor: An architecture for silicon's dark future*, *Micro, IEEE* **31**, 86 (2011).
- [24] Y. Turakhia, B. Raghunathan, S. Garg, and D. Marculescu, *Hades: Architectural synthesis for heterogeneous dark silicon chip multi-processors*, in *Design Automation Conference (DAC), 2013 50th ACM / EDAC / IEEE* (2013) pp. 1–7.
- [25] G. Venkatesh, J. Sampson, N. Goulding-Hotta, S. K. Venkata, M. B. Taylor, and S. Swanson, *Qscores: Trading dark silicon for scalable energy efficiency with quasi-specific cores*, in *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO-44 (ACM, New York, NY, USA, 2011) pp. 163–174.
- [26] M. J. Lyons, M. Hempstead, G.-Y. Wei, and D. Brooks, *The accelerator store: A shared memory framework for accelerator-based systems*, *ACM Trans. Archit. Code Optim.* **8**, 48:1 (2012).
- [27] M. R. Choudhury and K. Mohanram, *Approximate logic circuits for low overhead, non-intrusive concurrent error detection*, in *Proceedings of the Conference on Design, Automation and Test in Europe*, DATE '08 (ACM, New York, NY, USA, 2008) pp. 903–908.
- [28] V. Gupta, D. Mohapatra, S. P. Park, A. Raghunathan, and K. Roy, *Impact: Imprecise adders for low-power approximate computing*, in *Proceedings of the 17th IEEE/ACM International Symposium on Low-power Electronics and Design*, ISLPED '11 (IEEE Press, Piscataway, NJ, USA, 2011) pp. 409–414.

- [29] J. Han and M. Orshansky, *Approximate computing: An emerging paradigm for energy-efficient design*, in *Test Symposium (ETS), 2013 18th IEEE European* (2013) pp. 1–6.
- [30] P. Kulkarni, P. Gupta, and M. Ercegovac, *Trading accuracy for power with an underdesigned multiplier architecture*, in *VLSI Design (VLSI Design), 2011 24th International Conference on* (2011) pp. 346–351.
- [31] D. Mohapatra, V. Chippa, A. Raghunathan, and K. Roy, *Design of voltage-scalable meta-functions for approximate computing*, in *Design, Automation Test in Europe Conference Exhibition (DATE), 2011* (2011) pp. 1–6.
- [32] A. Verma, P. Brisk, and P. Ienne, *Variable latency speculative addition: A new paradigm for arithmetic circuit design*, in *Design, Automation and Test in Europe, 2008. DATE '08* (2008) pp. 1250–1255.
- [33] H. Esmaeilzadeh, A. Sampson, L. Ceze, and D. Burger, *Architecture support for disciplined approximate programming*, in *ACM SIGPLAN Notices*, Vol. 47 (ACM, 2012) pp. 301–312.
- [34] J. Allred, S. Roy, and K. Chakraborty, *Designing for dark silicon: A methodological perspective on energy efficient systems*, in *Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design, ISLPED '12* (ACM, New York, NY, USA, 2012) pp. 255–260.
- [35] R. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, *Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits*, *Proceedings of the IEEE* **98**, 253 (2010).
- [36] V. Hanumaiah, S. Vrudhula, and K. Chatha, *Performance optimal online ddfs and task migration techniques for thermally constrained multi-core processors*, *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* **30**, 1677 (2011).
- [37] V. Hanumaiah and S. Vrudhula, *Energy-efficient operation of multicore processors by ddfs, task migration, and active cooling*, *Computers, IEEE Transactions on* **63**, 349 (2014).
- [38] U. Karpuzcu, A. Sinkar, N. S. Kim, and J. Torrellas, *Energysmart: Toward energy-efficient manycores for near-threshold computing*, in *High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on* (2013) pp. 542–553.
- [39] D. Markovic, C. Wang, L. Alarcon, T.-T. Liu, and J. Rabaey, *Ultralow-power design in near-threshold region*, *Proceedings of the IEEE* **98**, 237 (2010).
- [40] L. Wang and K. Skadron, *Implications of the power wall: Dim cores and reconfigurable logic*, *Micro, IEEE* **33**, 40 (2013).
- [41] Microsoft, *Microsoft project, project management software*, (1992-2015).

- [42] Trello, *Trello, web-based project management software*, (2011-2015).
- [43] Synopsys, *Changing market drivers, synopsys insight newsletter, issue 1*, (2015).
- [44] H. Esmailzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, *Dark silicon and the end of multicore scaling*, in *Proceedings of the 38th Annual International Symposium on Computer Architecture*, ISCA '11 (ACM, New York, NY, USA, 2011) pp. 365–376.

2

Ταχεία Προτυποποίηση με Υβριδικά Εικονικά Συστήματα-σε-Ψηφίδα

Στον πεδίο των ενσωματωμένων συστημάτων αναζητείται συνεχώς μεγαλύτερη ευελιξία στην ανάπτυξη εφαρμογών. Η τάση αυτή, έχει καθιερώσει την εικονική προτυποποίηση ως μία ευέλικτη μεθοδολογία για ταχεία συστημική προσομοίωση. Μεταξύ άλλων, μια τέτοια προσέγγιση, τίθεται υπέρ του ταυτόχρονου σχεδιασμού υλικού/λογισμικού, επιτρέποντας την ανάπτυξη, δοκιμή και επαλήθευση του ενσωματωμένου λογισμικού, ουσιαδώς ταχύτερα σε σχέση με παρελθούσες τεχνολογίες. Υπό το πρίσμα αυτό, στο παρόν κεφάλαιο παρουσιάζεται ένα νέο πλαίσιο, υπό την ονομασία «Plug&Chip», το οποίο στοχεύει στην ταχεία προτυποποίηση των δισδιάστατων και τρισδιάστατων 2-D και 3-D ψηφιακών συστημάτων. Σημειώνεται, ότι εν αντιθέσει προς τις υφιστάμενες σχετικές προσεγγίσεις, η προτεινόμενη λύση παρέχει υψηλότερο επίπεδο ευελιξίας, επιτρέποντας ταχύ σταδιακό σχεδιασμό του συστήματος, ενώ δίνει ακόμα τη δυνατότητα της υλοποίησης σε πλατφόρμες που αναπτύσσονται με τη χρήση της νέας υποσχόμενης τρισδιάστατης τεχνολογίας ολοκλήρωσης (3-D).

2.1. Εισαγωγή - Κίνητρο έρευνας για ταχεία εικονική προτυποποίηση

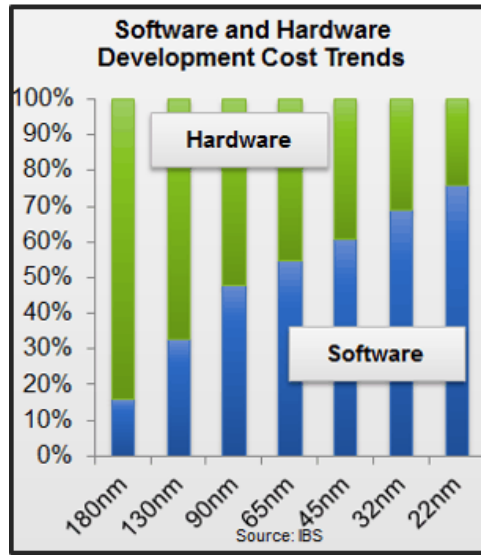
Με την αυξανόμενη ανάγκη για πολυπλοκότητα και λειτουργικότητα, ειδικά κατά την εποχή σχεδίασης στην κλίμακα μερικών νανομέτρων, ο σχεδιασμός ολοκληρωμένων κυκλωμάτων αποτελεί ιδιαίτερη πρόκληση για τους μηχανικούς και αρχιτέκτονες υπολογιστικών συστημάτων. Πέραν τούτου, η συνεχώς αυξα-

νόμενη ζήτηση για ακόμη υψηλότερες επιδόσεις, αναφορικά προς τη συχνότητα λειτουργίας, την κατανάλωση ενέργειας, κλπ, καθιστά απολύτως απαραίτητες νέες τεχνικές σχεδιασμού, τόσο σε φυσικό επίπεδο σχεδιασμού όσο και σε μεθοδολογικές προσεγγίσεις. Αναφορικά με την πρώτη τεχνική, η τρισδιάστατη ολοκλήρωση (3-D), η οποία υπόσχεται πολλαπλά επίπεδα πυριτίου στον τρισδιάστατο χώρο, δύναται να αντιμετωπίσει την εν λόγω πρόκληση, δεδομένου ότι βελτιώνει σημαντικά την απόδοση και τη λειτουργικότητα των κυκλωμάτων, ενώ ταυτόχρονα μειώνει την απόσταση μεταξύ των λογικών δομών ενός ολοκληρωμένου κυκλώματος [4].

Εκτός από τις παραμέτρους τεχνολογικής φύσεως, οι οποίες επηρεάζουν την αποτελεσματικότητα και την ευελιξία ενός ψηφιακού συστήματος, τα στενά χρονικά περιθώρια σχετικά με τις απαιτήσεις της αγοράς ωθούν τους συμβατικούς τρόπους για την ανάπτυξη του προϊόντος (π.χ. την έναρξη ανάπτυξης λογισμικού μετά την οριστικοποίηση του υλικού) συνήθως στην απώλεια αγοραστικών ευκαιριών και των εντεύθεν εσόδων. Ως εκ τούτου, τίθεται μια αρχική και απόλυτη προϋπόθεση για έναρξη του σχεδιασμού λογισμικού πριν την ολοκλήρωση του υλικού, π.χ. σε επίπεδο RTL (Επίπεδο Μεταφοράς Καταχωρητή). Η σπουδαιότητα του εν λόγω προβλήματος καθίσταται εμφανέστερη, λαμβάνοντας υπόψη ότι το κόστος του λογισμικού μπορεί να αντιστοιχεί σε ποσοστό της τάξεως του 80%, ή και περισσότερο, του συνολικού κόστους ανάπτυξης των σύγχρονων ενσωματωμένων συστημάτων, σύμφωνα με μελέτες του Διεθνούς Οργανισμού Ημιαγωγών [2], καθιστώντας τον συμβατικό τρόπο σχεδιασμού ανεπαρκή. Επίσης, ο ίδιος οργανισμός προβλέπει ότι το κόστος ανάπτυξης του λογισμικού θα αυξηθεί ισόποσα προς το κόστος του υλικού, παρά την έλευση των πολυνηματικών εργαλείων ανάπτυξης λογισμικού που αξιοποιούν την παραλληλία που προσφέρουν συστήματα πολλαπλών επεξεργαστών [2].

Σε συνάρτηση με την προαναφερθείσα παρατήρηση, μία ακόμη παράμετρος σχεδιασμού ολοκληρωμένων κυκλωμάτων αποτελεί η κλιμάκωση του κόστους σχεδίασης αναφορικά προς τον τεχνολογικό κόμβο ολοκλήρωσης. Όπως φαίνεται στο Σχήμα 5.1 το κόστος ανάπτυξης λογισμικού αυξάνεται συνεχώς κατά τη μετάβαση σε μικρότερους τεχνολογικούς κόμβους. Σχετική πρόσφατη έρευνα έδειξε ότι τα περισσότερα έργα καθυστερούν κυρίως λόγω της ανάπτυξης λογισμικού, ενώ το μεγαλύτερο ποσοστό του ανθρώπινου δυναμικού της ομάδας σχεδίασης ασχολείται με το λογισμικό [3].

Δια την εύρεση βέλτιστων σχεδιαστικών λύσεων, σημαντικό ρόλο έχουν τα εργαλεία σχεδίασης. Οι υφιστάμενες ροές Αυτόματου Ηλεκτρονικού Σχεδιασμού (Electronic Design Automation - EDA) εδράζονται στη θεμελιώδη παραδοχή ότι τα μοντέλα σχεδίασης έχουν ανοιχτά πρότυπα ενσωμάτωσης και επομένως είναι ελεύθερα εναλλάξιμα μεταξύ των κατασκευαστών (vendors), ενώ παράλληλα παρουσιάζουν διαλειτουργικότητα μεταξύ τους. Αυτή η παραδοχή, στη φάση της σχεδίασης, σημαίνει ότι είτε τα μοντέλα σχεδιάζονται εξ αρχής από την ομάδα σχεδίασης, είτε αυτά προέρχονται από άλλους κατασκευαστές, τότε θα είναι επεξεργάσιμα από οποιαδήποτε εργαλείο σχεδίασης για την εκτέλεση διαφόρων σταδίων των φυσικής προτυποποίησης (π.χ. αρχιτεκτονική ανάλυση, εξομοίωση, σύνθεση, κ.λπ.). Ακόμη και αν η ιδέα αυτή φαίνεται απλή και πολλά



Σχήμα 2.1 Τάσεις κόστους ανάπτυξης υλικού και λογισμικού, αναφορικά προς τη κλιμάκωση της τεχνολογίας ολοκλήρωσης.

υποσχόμενη, παραμένει ασύλληπτη στον κόσμο των Εργαλείων Σχεδίασης σε Επίπεδο Συστήματος (Electronic System Level, ESL). Ειδικότερα, οι υπάρχουσες λύσεις ESL δεν παρέχουν τη διαλειτουργικότητα των μοντέλων, αλλά ούτε και την ανεξαρτησία μεταξύ μοντέλου και εργαλείου λογισμικού. Κατά συνέπεια, η υιοθέτηση κοινών προτύπων, μοντέλων και μεθοδολογιών ESL μεταξύ διαφόρων κατασκευαστών, είναι ένα επιθυμητό χαρακτηριστικό της σύγχρονης σχεδίασης.

Προς αυτή την κατεύθυνση και καθώς η έρευνα πιέζει για καλύτερα προγραμματιστικά μοντέλα ενσωματωμένων συστημάτων με πολλαπλούς επεξεργαστές και πυρήνες, οι Εικονικές Πλατφόρμες (VP) επιλύουν μία από τις μεγαλύτερες σύγχρονες προκλήσεις στο φυσικό σχεδιασμό, δηλαδή αυτή της ύπαρξης μοντέλων για επαρκή ανάπτυξη λογισμικού, αποσφαλμάτωση και επαλήθευση, πριν το τελικό προϊόν υλικού γίνει διαθέσιμο. Ειδικότερα, με τη δυνατότητα της εικονικοποίησης, είναι δυνατόν να προτυποποιηθεί μία πλατφόρμα υλικού, αποτελούμενη από διαφορετικούς πυρήνες επεξεργασίας, μνήμες, περιφερειακά συστήματα, καθώς και συστήματα διασύνδεσης, υπό τη μορφή ενός προσομοιωτή, π.χ. [4]. Επιπλέον, όσο η σχεδίαση του υλικού επιτυγχάνεται προοδευτικά, καθίσταται εφικτή η ενημέρωση των εκδόσεων Εικονικών Πλατφόρμων (VP), στην ομάδα ανάπτυξης λογισμικού, ώστε να βελτιωθεί έτι περαιτέρω η περιγραφή της αρχιτεκτονικής του υπό-ανάπτυξη συστήματος.

Η ιδέα της εικονικοποίησης είναι επίσης εξέχουσας σημασίας για τους προγραμματιστές υλικού, δεδομένου ότι επιτρέπει ευκολότερη επαλήθευση των πυρήνων Πνευματικής Ιδιοκτησίας (Intellectual Properties, IP). Αυτή η διαδικασία μπορεί να εφαρμοστεί τόσο στην περίπτωση που μόνο μερικές από τις εφαρμο-

γές πυρήνων πρέπει να αναπτυχθούν στο υλικό, όσο και στην σταδιακή καθολική ανάπτυξη του συστήματος. Σε αμφότερες τις περιπτώσεις, η λειτουργία της εικονικοποίησης παρέχει όλους τους απαραίτητους μηχανισμούς για την εκτέλεση προσομοίωσης σε πολύπλοκες σχεδιάσεις και επαλήθευση μεταξύ των πυρήνων Πνευματικής Ιδιοκτησίας (Intellectual Property, IP) που αναπτύχθηκαν στο Επίπεδο Καταχωρητή (Register Transfer Level, RTL) και της υπόλοιπης λειτουργικότητας της εφαρμογής που εκτελείται στις εικονικές πλατφόρμες.

Στο παρόν κεφαλαίο, παρουσιάζεται μια μεθοδολογία, βασισμένη σε νέα εργαλεία λογισμικού, για την εκτέλεση ταχείας προτυποποίησης σύνθετων 2-D και 3-D συστημάτων SoCs. Για το σκοπό αυτό, έχουν αυτοματοποιηθεί οι σχετικές με την μοντελοποίηση και προσομοίωση διαδικασίες μεταξύ των πυρήνων σχεδιασμού που ανατίθενται σε διαφορετικές πλατφόρμες σχεδίασης. Στα πλαίσια αυτής της εργασίας ορίζεται ένα ευρύ και γενικό μοντέλο εικονικής προτυποποίησης, το οποίο περιλαμβάνει τον υπολογιστή υποδοχής (host PC), την εικονική πλατφόρμα (VP) και το σύστημα φυσικής προτυποποίησης, π.χ. FPGA, ASIC κτλ. Δεδομένου ότι για τους σκοπούς της παρούσας ανάλυσης επιστρατεύονται πολλαπλές ετερογενείς πλατφόρμες, έχει αναπτυχθεί ένα γενικό σύστημα επικοινωνίας, το οποίο παρέχει την μεταφορά των επιθυμητών δεδομένων με διάφορες παραμέτρους (π.χ. το μέγεθος του πακέτου, μέγιστη μονάδα μετάδοσης (MTU), κλπ). Μία ακόμα πολλά υποσχόμενη λειτουργία η οποία αντιμετωπίζεται υπό την ερευνητική αυτή προσέγγιση, αναφέρεται στα πιθανά οφέλη από τη μετάβαση του σχεδιασμού SoC στην τρισδιάστατη 3-D τεχνολογία ολοκλήρωσης. Λαμβάνοντας υπόψιν το γεγονός ότι τα φυσικά εργαλεία υλοποίησης για το σχεδιασμό 3-D SoCs [5], [6] έχουν αυξημένο κόστος αγοράς και σχεδίασης, το προτεινόμενο πλαίσιο παρέχει μία προσεγγιστική εκτίμηση της αποδοτικότητας των σχεδιασμών από την υιοθέτηση μιας τέτοιας προηγμένης τεχνολογίας διασύνδεσης, χρησιμοποιώντας συμβατικά εργαλεία σχεδίασης 2-D.

Το υπόλοιπο τμήμα του παρόντος κεφαλαίου δομείται ως εξής: Στην ενότητα 2.2 συνοψίζεται η υφιστάμενη βιβλιογραφία στον τομέα της εικονικής προτυποποίησης. Με βάση αυτή την ανάλυση, δίδεται έμφαση στα ανοικτά ζητήματα, καθώς και στα κίνητρα αυτής της εργασίας. Εν συνεχεία στην ενότητα 2.3 αναλύεται η προτεινόμενη μεθοδολογία ταχείας προτυποποίησης, ενώ στην ενότητα 2.4, παρέχεται μια σειρά από πειραματικά αποτελέσματα που αποδεικνύουν την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας, ενώ, εν τέλει, στην ενότητα 2.5 συνοψίζονται τα συμπεράσματα της εργασίας.

2.2. Υφιστάμενες Προσεγγίσεις Προτυποποίησης Συστημάτων με Εικονικές Πλατφόρμες

Η παρούσα ενότητα συνοψίζει ένα αντιπροσωπευτικό αριθμό από σύγχρονες προσεγγίσεων στον τομέα της εικονικοποίησης πλατφορμών. Ο σκοπός της παρούσας ανάλυσης είναι η επισήμανση των περιορισμών των υφιστάμενων λύσεων, καθώς και η μετενέργεια αυτών στην επαρκή ταχεία προτυποποίηση των

σύνθετων ψηφιακών συστημάτων. Επιπρόσθετα, τα ανοικτά ζητήματα, ως αυτά συναφθούν συμπερασματικά από την ανάλυση, θα χρησιμοποιηθούν ως κίνητρο για την παρουσίαση του νέου πλαισίου, υπό την ονομασία «Plug&Chip».

Δεδομένου ότι οι αυστηρές απαιτήσεις για την ακρίβεια και το χρόνο ανάπτυξης δεν μπορούν ευχερώς να επιτευχθούν με τη χρήση συμβατικών συστημικών πρωτοτύπων, όπως οι Εικονικές Πλατφόρμες (VP), η χρήση των υβριδικών πρωτοτύπων, αποδεικνύεται η πλέον βιώσιμη λύση. Ειδικότερα, υπό τον όρο υβριδική προτυποποίηση νοείται το σενάριο εκείνο, όπου ένα μέρος του σχεδιασμού τοποθετείται σε μια φυσική συσκευή (δηλαδή FPGA, ASIC, GPU, κλπ), ένα άλλο μέρος προσομοιώνεται σε μια εικονική πλατφόρμα η οποία εκτελείται στο σύστημα υποδοχής (host), ενώ τέλος, το υπόλοιπο σύστημα εκτελείται εγγενώς στο σύστημα υποδοχής. Όπως ορίζεται στη βιβλιογραφία, με τον τρόπο αυτό δίδεται η δυνατότητα συνδυασμού της ακρίβειας και της ταχύτητας των πλατφορμών υλικού με την ευελιξία και την αφαιρετικότητα που προσφέρουν οι εικονικές πλατφόρμες [7]. Δεδομένου, όμως, ότι η εν λόγω προσέγγιση ενσωματώνει διαφορετικές μεθοδολογίες, οι οποίες θα πρέπει να συνδυαστούν, επιβάλλεται όπως αναπτυχθούν οι κατάλληλες διασυνδέσεις μεταξύ διαδοχικών τεχνικών και εργαλείων. Πρόσφατα, τέτοιου είδους πλαίσια έχουν αποκτήσει ιδιαίτερο ενδιαφέρον τόσο στην ακαδημαϊκή, όσο και στην βιομηχανική κοινότητα, καθώς η πολυπλοκότητα των ψηφιακών συστημάτων αυξάνεται εκθετικά [2]. Επιπλέον, η ευελιξία που παρέχεται από τα υφιστάμενα πλαίσια, να ξεκινήσουν τη φυσική υλοποίηση από διαφορετικά αφαιρετικά επίπεδα σε σχέση με το επίπεδο RTL (π.χ. SystemC) επιβάλλει την αντίστοιχη προσαρμογή των εργαλεία λογισμικού, τα οποία αυτοματοποιούν τις διαδικασίες προσομοίωσης, αποσφαλμάτωσης και επαλήθευσης.

Μια τυπική διασύνδεση για υβριδικά πρωτότυπα είναι η λύση SCE-MI [8], η οποία υποστηρίζεται από την εταιρεία Accelera και η οποία παρέχει μια διεπαφή επικοινωνίας μεταξύ του συστήματος υποδοχής και ενός εξομοιωτή γλώσσας HDL. Πιο αναλυτικά, το σύστημα υποδοχής εκτελεί ένα λογισμικό ελέγχου (testbench) για το χειρισμό της διεπαφής επικοινωνίας. Οι πυρήνες IP που αναπτύσσονται σε γλώσσα HDL, περιλαμβάνουν επίσης την απαραίτητη διασύνδεση για την επικοινωνία με το σύστημα υποδοχής.

Στις επόμενες υποενότητες λαμβάνει χώρα λεπτομερέστερη παρουσίαση των σχετικών προσεγγίσεων για την υποστήριξη της επικοινωνίας μεταξύ του Συστήματος Υποδοχής και Εικονικής πλατφόρμας (Host2VP), καθώς και Εικονικής Πλατφόρμας και Υλικού (VP2HW).

2.2.1. Επικοινωνία Μεταξύ Συστήματος Υποδοχής και Εικονικής πλατφόρμας (Host2VP)

Σε αυτήν την ενότητα αναλύεται ποιοτικά το σκέλος της επικοινωνίας του υπολογιστή υποδοχής και της εικονικής πλατφόρμας (Host2VP), ενός περιβάλλοντος εικονικοποίησης για την υποστήριξη υβριδικής προτυποποίησης. Τα αποτελέσματα αυτής της ανάλυσης συνοψίζονται στον Πίνακα 2.1, ενώ για λόγους πληρότητας, η τελευταία στήλη του πίνακα αναφέρεται στο νέο προτεινόμενο πλαίσιο Plug&Chip, το οποίο παρουσιάζεται στο παρόν κεφάλαιο.

Η πρώτη προσέγγιση που απεικονίζεται στον Πίνακα 2.1 περιλαμβάνει έναν εξομοιωτή παρεχόμενο από την εταιρεία Altera[9]. Αυτός ο εξομοιωτής επιτρέπει την ταχεία ανάπτυξη λογισμικού χρησιμοποιώντας τον ενσωματωμένο επεξεργαστή ARM Cortex-A9 καθώς και τα περιφερειακά που εμπεριέχονται στα Altera SoC FPGA. Για την υποστήριξη υβριδικής προτυποποίησης, ο εξομοιωτής συνδέεται με το σύστημα υποδοχής μέσω ενός μηχανισμού επικοινωνίας και του αντίστοιχου οδηγού λογισμικού (software driver). Παρόλο που η λύση της Altera [9] υποστηρίζει γρήγορη λειτουργική εξομοίωση, η χαμηλού χρονισμού ακρίβεια, εφόσον χρησιμοποιεί μόνο εξομοίωση σε επίπεδο περιγραφής, εισάγει περιορισμούς στη λεπτομερή αξιολόγηση του λογισμικού.

Η δεύτερη στήλη του Πίνακα 2.1 ανταποκρίνεται σε μια προσέγγιση όπου ένας QEMU x86 εξομοιωτής συνδυάζεται με ένα SystemC μοντέλο [10], [11]. Παρόλο που η γλώσσα SystemC διαβεβαιώνει την καθολικότητα αυτής της λύσης, καθώς επίσης και την απεριόριστη υποστήριξη για επεξεργαστές (CPU cores) στο μέρος της εικονικής πλατφόρμας VP, υπάρχουν περιορισμοί απόδοσης που σχετίζονται με την εκτέλεση του λογισμικού στον x86 επεξεργαστή, εφόσον πρόκειται για προσομοίωση (emulation). Επίσης, δεν υπάρχουν τυποποιημένοι μηχανισμοί συγχρονισμού που να διασφαλίζουν συνεχείς και μη-διακεκομμένες μεταφορές δεδομένων. Όσον αφορά τη προσέγγιση HySim (Hybrid Processor Simulation) [12], η επικοινωνία με τον υπολογιστή υποδοχής γίνεται με ένα εξειδικευμένο λογισμικό (wrapper), το οποίο λειτουργεί σαν ένας εξομοιωτής αφαιρετικού επιπέδου για το τμήμα λογισμικού που δεν χρίζει προσομοίωσης. Το εργαλείο αυτό επιτρέπει την εκτέλεση εξομοίωσης σε επίπεδο περιγραφής (functional) του τελικού συστήματος, ενώ, όσον αφορά την ακρίβεια της χρονικής προσομοίωσης, εξαρτάται από το επιλεγόμενο σετ εντολών του εξομοιωτή (Instruction Set Simulator, ISS) και το επίπεδο αφαίρεσης. Ένα μειονέκτημα αυτής της τεχνικής είναι η επιπρόσθετη προσπάθεια που απαιτείται από τον σχεδιαστή έτσι ώστε να τροποποιήσει το λογισμικό που εκτελείται τόσο στην πλευρά του υπολογιστή υποδοχής, όσο και στις εικονικής πλατφόρμας VP.

Πίνακας 2.1 Ποιοτική σύγκριση για μηχανισμούς επικοινωνίας Host-to-VP.

	Altera Virtual Target [9]	QEMU/SystemC [40] [11]	HySim [12]	Semi-hosting [4]	Προτεινόμενη Προσέγγιση
Επίπεδα αφαίρεσης	Functional	Functional, TLM, RTL	Timed Functional	Functional	TLM, RTL
Ακρίβεια VP	Χαμηλή	Επίπεδο Κύκλου	Επίπεδο Κύκλου ¹	Χαμηλή	Επίπεδο Κύκλου ¹
Συγχρονισμός Δεδομένων	Ενσωματωμένη	Manual	Ενσωματωμένη	Manual	Ενσωματωμένη
Εκτέλεση SW στον Host	Γρήγορη	Μέτρια	Γρήγορη	Εξάρτηση πακέτου ²	Γρήγορη
Εξομοίωση στη VP	Γρήγορη	Μέτρια	Μέτρια	Εξάρτηση πακέτου ²	Μέτρια
Πλατφόρμα για Host	Φυσική Πλατφόρμα	QEMU	Φυσική Πλατφόρμα	Φυσική Πλατφόρμα	Φυσική Πλατφόρμα
CPUs για VP	ARM	Οποιαδήποτε	Οποιαδήποτε ¹	Οποιαδήποτε	Οποιαδήποτε
Αριθμός CPUs στην VP	Max 2	Απεριόριστος	Απεριόριστος	Απεριόριστος	Απεριόριστος
Μηχανισμοί Επικοινωνίας	Driver-based ³	Ενσωματωμένοι ⁴ ή IPC	Ειδικός wrapper	Σκληρός Δίσκος	IPC, Οδηγός
Παρεμβατικό στον Host	Όχι	Όχι	Ναι	Ναι	Όχι
Παρεμβατικό στη VP	Όχι	Όχι	Ναι	Ναι	Όχι
Υποστήριξη περιφερ. TLM	Όχι	Ναι	Όχι	Όχι	Ναι
Πολλαπλά περιβάλλοντα VP	Όχι	Ναι	Ναι	Όχι	Ναι

Η επόμενη προσέγγιση περιλαμβάνει τη χρήση της τεχνικής semi-hosting, η οποία αφορά τη χρήση των κοινά χρησιμοποιούμενων κλήσεων συστήματος (e.g. *read()* or *write()*) στο σύστημα υποδοχής, στη περίπτωση που δεν υπάρχει λειτουργικό σύστημα στην εικονική πλατφόρμα VP. Αντιπροσωπευτική λύση του semi-hosting είναι το λογισμικό OVP [4]. Εφόσον το OVP υποστηρίζει μόνο εξομοίωση σε επίπεδο περιγραφής, είναι κατάλληλο κυρίως για μοντελοποίηση του συστήματος σε υψηλά επίπεδα αφαίρεσης, όπου δε χρειάζεται ακρίβεια χρονισμού. Όσον αφορά την επικοινωνία μεταξύ Host και VP, στο OVP, δεν υπάρχει συγκεκριμένος μηχανισμός συγχρονισμού καθώς κλήση συστήματος *fcntl()*, η οποία χρησιμοποιείται για προστασία κλειδώματος αρχείων (file locking), δεν υποστηρίζεται σε κάθε μοντέλο επεξεργαστή που διαθέτει το OVP. Πιο συγκεκριμένα, αυτή η επικοινωνία μπορεί να επιτευχθεί μέσω ενός μέσου αποθήκευσης, π.χ. το σκληρό δίσκο του συστήματος υποδοχής, το οποίο επιβάλλει ότι η ταχύτητα εξομοίωσης εξαρτάται από τη ποιότητα των μεταφερόμενων δεδομένων. Παρόλο που αυτός ο περιορισμός μπορεί να αντιμετωπιστεί με μηχανισμούς Unix IPC (π.χ. named pipes), το semi-hosting υποστηρίζει μόνο αρχεία συμβατά με τη σειρά προτύπων POSIX.

¹Σε εξάρτηση από τον ISS (Instruction Set Simulator).

²Μεγάλη για μεγάλα πακέτα δεδομένων, χαμηλή για μικρά πακέτα δεδομένων.

³In a driver-based communication scheme, the simulator is accessed through device driver.

⁴Ο QEMU ISS χρησιμοποιείται σαν μονάδα SystemC.

Η τελευταία στήλη σε αυτό τον πίνακα συνοψίζει τα χαρακτηριστικά του προτεινόμενου πλαισίου, παρέχοντας την επικοινωνία μεταξύ συστήματος υποδοχής και VP. Πιο συγκεκριμένα, όντας μια προσέγγιση βασισμένη στη γλώσσα SystemC, αποτελεί μια καθολική λύση που υποστηρίζει διαφορετικά επίπεδα αφαίρεσης (π.χ. TLM) και απεριόριστο αριθμό από επεξεργαστές CPUs. Εξαρτώμενο από το επιλεγμένο σετ εντολών εξομοίωσης ISS, το πλαίσιο Plug&Chip μπορεί να παρέχει ακόμα και ακριβής προσομοίωση σε επίπεδο κύκλου ρολογιού (cycle-accurate). Επιπλέον, έχει ενσωματωμένους μηχανισμούς συγχρονισμού, ενώ οι μηχανισμοί επικοινωνίας υποστηρίζουν μια μη-παρεμβατική επίπτωση στο χρόνο εξομοίωσης, λόγω της χρήσης τεχνικών IPC (Inter-Process Communication) αλλά και του εξειδικευμένου οδηγού λογισμικού.

Πίνακας 2.2 Ποιοτική σύγκριση για μηχανισμούς επικοινωνίας VP-to-Hardware.

	Synopsys HAPS [13]	Xilinx Co-Simulation [14]	Altera Virtual Target [9]	Semi-hosting ⁵ [4]	Proposed
Επίπεδα αφαίρεσης	TLM	RTL	RTL	Functional	TLM
Ακρίβεια VP	Επίπεδο Κύκλου	Επίπεδο Κύκλου	Επίπεδο Κύκλου	Μικρή	Επίπεδο Κύκλου ⁶
Συγχρονισμός Δεδομένων	Ενσωματωμένη	Ενσωματωμένη	Ενσωματωμένη	Εξωτερικός	Ενσωματωμένη
Εξομοίωση στη VP	Γρήγορη	Αργή	Γρήγορη	Γρήγορη	Γρήγορη
Υποστηριζόμενο Υλικό	FPGA	FPGA	FPGA	Οποιαδήποτε	Οποιαδήποτε
Υποστηριζόμενες CPUs για VP	ARM	Οποιαδήποτε	ARM	Οποιαδήποτε	Οποιαδήποτε
Αριθμός CPUs στη VP	Απεριόριστος	Απεριόριστος	Max 2 cores	Απεριόριστος	Απεριόριστος
Πρωτόκολλο Επικοινωνίας	AMBA, Raw data	Raw data	Raw data	Raw data	Raw data
Μηχανισμοί Επικοινωνίας	Οδηγός	Οδηγός	Οδηγός	Σκληρός Δίσκος	Οδηγός
Επεκτάσιμη Βιβλιοθήκη	Όχι	Ναι	Όχι	Ναι	Ναι
Υποστήριξη περιφερ. TLM	Εξωτερικές Βιβλιοθήκες	Όχι	Όχι	Όχι	Ναι
Πολλαπλά Περιβάλλοντα VP	Όχι	Όχι	Όχι	Όχι	Ναι

⁵VP communicates with a host proxy which forwards data to/from the FPGA.

⁶Σε εξάρτηση από τον ISS.

2.2.2. Επικοινωνία Μεταξύ Εικονικής πλατφόρμας και Υλικού

Παρόλο που η σύνδεση στη πλατφόρμα υλικού μπορεί εύκολα να επιτευχθεί μέσω ενός οδηγού συσκευής, ο μηχανισμός επικοινωνίας θα πρέπει επίσης να κάνει ικανό το λογισμικό της εικονικής πλατφόρμας να στείλει/λάβει δεδομένα σε/από την πλατφόρμα υλικού. Αυτό είναι το κύριο χαρακτηριστικό ενός τέτοιου σχήματος επικοινωνίας εικονικής πλατφόρμας και υλικού (VP2HW). Ο Πίνακας 2.2 συνοψίζει τον αριθμό των αντιπροσωπευτικών λύσεων που παρέχουν τέτοια είδους επικοινωνία. Ομοίως με τη προηγούμενη περίπτωση, η τελευταία στήλη αυτού του πίνακα αντιστοιχεί στη προτεινόμενη Plug&Chip μεθοδολογία.

Οι πρώτες τρεις προσεγγίσεις που αναφέρονται στον Πίνακα 2.2 είναι προϊόντα εμπορίου, τα οποία παρέχουν ακριβή εξομοίωση, σε επίπεδο κύκλου ρολογιού, ενώ επίσης δίνουν τη δυνατότητα συν-εξομοίωσης με τη χρήση FPGAs. Πιο συγκεκριμένα, η πλατφόρμα Synopsys HAPS [13] διαθέτει μια διεπαφή TLM με την εικονική μηχανή προκειμένου να επικοινωνεί με την πλατφόρμα υλικού. Εκτός του να επιτρέπει τη μεταφορά ακατέργαστων (raw) δεδομένων μέσω φυσικών συνδέσμων (π.χ. USB, Ethernet, κτλ), η πλατφόρμα HAPS επίσης υποστηρίζει ένα ευρέως γνωστό πρωτόκολλο μεταφοράς για ενσωματωμένα συστήματα (AMBA), στη περίπτωση που αυτό είναι διαθέσιμο και στο τελικό σύστημα. Παρόλο που η πλατφόρμα HAPS φαίνεται μια υποσχόμενη λύση, βασίζεται σε περιορισμένης άδειας βιβλιοθήκες TLM και επομένως είναι δύσκολα επεκτάσιμο.

Μία εναλλακτική προσέγγιση, το πλαίσιο Xilinx Co-Simulation Flow [14], εγκαθιστά μια σύνδεση μεταξύ του RTL-εξομοιωτή και του FPGA, μέσω ενός φυσικού διαύλου (π.χ. USB, Ethernet, κτλ). Όπως και στην περίπτωση της πλατφόρμας HAPS, η λύση της Xilinx δεν υποστηρίζει υψηλότερα επίπεδα αφαίρεσης κατά το σχεδιασμό νέων πυρήνων IP. Συνεπώς οι IP πυρήνες, όπως και ολόκληρη η πλατφόρμα, πρέπει να αναπτυχθούν αποκλειστικά σε επίπεδο RTL. Αυτός ο περιορισμός επίσης επιβάλλει ότι την μικρότερη ταχύτητα προσομοίωσης συγκριτικά με παρόμοιες προσεγγίσεις που βασίζονται σε μοντέλα SystemC. Μια σχετική λύση περιλαμβάνει η πλατφόρμα Altera Virtual Target [9], η οποία, ομοίως με το πλαίσιο Xilinx Co-Simulation Flow, επιτρέπει την ενσωμάτωση ενός FPGA κατά τη φάση προτυποποίησης, προσφέροντας την ταυτόχρονη εξομοίωση πυρήνων IP, περιγραφόμενων σε γλώσσα HDL, μαζί με συστήματα που περιέχουν ARM επεξεργαστές. Ωστόσο ο αριθμός αυτών των επεξεργαστών περιορίζεται στους δύο, ενώ δεν προσφέρεται η χρήση διαφορετικού τύπου σετ εντολών και κατ' επέκταση επεξεργαστών.

Η τέταρτη προσέγγιση παρουσιάζει τα πλεονεκτήματα και μειονεκτήματα της τεχνικής semi-hosting. Στη περίπτωση αυτή, το σύστημα υποδοχής γίνεται μία ενδιάμεση πλατφόρμα επικοινωνίας μεταξύ της εικονικής μηχανής και της πλατφόρμας υλικού, καθώς αυτή είναι η μόνη λύση που μπορεί να υποστηρίξει την πρόσβαση σε φυσικές συσκευές (character devices) μέσα από το λογισμικό της εικονικής μηχανής. Η τεχνική semi-hosting μπορεί να θεωρηθεί επεκτάσιμη σε διάφορα επίπεδα σχεδίασης, αρκεί το πρωτόκολλο επικοινωνίας να υλοποιείται σε καθένα από αυτά. Ωστόσο αυτή η επεκτασιμότητα είναι σπάνια αποδοτική λόγω της έλλειψης υποστήριξης στα χαμηλά επίπεδα αφαίρεσης. Επιπλέον,

δεν υπάρχει δεδομένος τρόπος υποστήριξης διεπαφής TLM στην τεχνική semi-hosting.

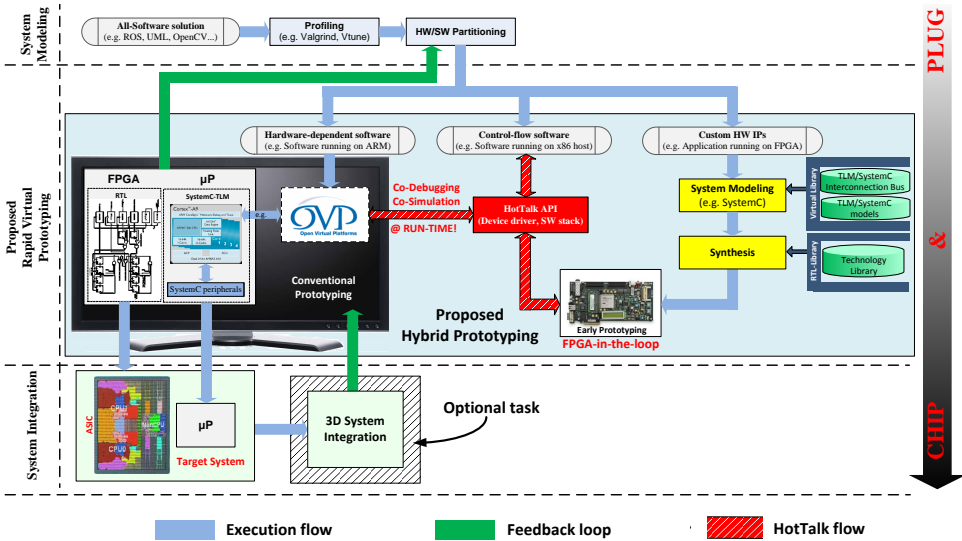
Το προτεινόμενο σχήμα επικοινωνίας VP2HW του πλαισίου Plug&Chip μπορεί να χειριστεί ένα πλήθος επεξεργαστικών πυρήνων, ενώ η τελική πλατφόρμα υλοποίησης μπορεί να είναι μια συσκευή οποιασδήποτε σύγχρονης αρχιτεκτονική (π.χ. FPGA, ASIC, GPU, CPU, κτλ.), με την προϋπόθεση ότι παρέχεται το κατάλληλο λογισμικό οδηγού. Επιπλέον, αναφέρεται ότι παρόλο που στα πλαίσια αυτής της έρευνας η μεταφορά δεδομένων με τη πλατφόρμα υλικού γίνεται μέσω ενός απλού πρωτοκόλλου μη επεξεργασμένων δεδομένων, η προτεινόμενη VP2HW βιβλιοθήκη επικοινωνίας είναι πλήρως επεκτάσιμη με την έννοια ότι ο σχεδιαστής μπορεί εύκολα να χρησιμοποιήσει οποιοδήποτε πρωτόκολλο επικοινωνίας, επεκτείνοντας τα δεδομένα κεφαλίδας (header data) του απλού πρωτοκόλλου (raw data).

2.3. Μεθοδολογία Προτυποποίησης

Παρά το γεγονός ότι έχουν αναπτυχθεί πολλά εργαλεία σχεδιασμού τα οποία αντιμετωπίζουν μεμονωμένα τα προβλήματα λογισμικού (SW) και υλικού (HW), οι προσεγγίσεις που επιλαμβάνονται μεικτών προβλημάτων ολοκληρωμένου λογισμικού και εξειδικευμένου είναι ελάχιστες και τούτο διότι στην περίπτωση αυτή πρέπει να αντιμετωπιστούν προκλήσεις οι οποίες προκύπτουν από τη σχεδίαση σε επίπεδο συστήματος. Παρόλα αυτά, υφίστανται σε περιορισμένο εύρος και αριθμό προσεγγίσεις εργαλείων EDA που επιλύουν το πρόβλημα της συστημικής σχεδίασης με το συνδυασμό εργαλείων προσομοίωσης RTL, εξομίωσης και περιβάλλοντα προτυποποίησης. Ωστόσο, αυτές οι λύσεις είναι συχνά πολύ περίπλοκες, χρονοβόρες και δαπανηρές και ως επί το πλείστον υστερούν αναφορικά προς την επικοινωνία μεταξύ του συστήματος υποδοχής και του συστήματος προτυποποίησης.

Στην παρούσα ενότητα παρουσιάζεται ένα πλαίσιο με το οποίο δίδεται η δυνατότητα στους προγραμματιστές λογισμικού και τους μηχανικούς υλικού της από κοινού ανάπτυξης κατά τη φάση σχεδίασης. Το Σχήμα 2.2 απεικονίζει τη συνολική ροή της προτεινόμενης Plug&Chip μεθοδολογίας, η οποία απαρτίζεται από τρεις διαδοχικές φάσεις σχεδιασμού: (i) την μοντελοποίηση του συστήματος, (ii) την ταχεία εικονική προτυποποίηση και (iii) την συστημική ολοκλήρωση. Το ανταγωνιστικό πλεονέκτημα αυτού του πλαισίου είναι η συν-προσομοίωση μεταξύ συστήματος υποδοχής και πραγματικής πλατφόρμας υλικού, η οποία συμβιβάζει την ταχύτητα, μέσω προσομοίωσης σε επίπεδο συστήματος με αφαιρετικά μοντέλα σχεδίασης και την ακρίβεια, μέσω προσομοίωσης σε επίπεδο ακρίβειας κύκλου, ανάλογα με τις σχεδιαστικές απαιτήσεις.

Η μοντελοποίηση του συστήματος περιγράφει το στάδιο, όπου η ομάδα ανάπτυξης παρέχει μια αφηρημένη περιγραφή της αρχιτεκτονικής του συστήματος και ξεκινά τη σχεδίαση του τρόπου δια του οποίου τα ζητούμενα χαρακτηριστικά διαμορφώνονται σε πυρήνες λογισμικού και υλικού. Επίσης, κατά το στάδιο αυτό αποτιμώνται όλοι οι περιορισμοί αναφορικά με την επίτευξη της απόδοσης του τελικού συστήματος. Δεδομένου ότι αυτό είναι ένα πρώιμο στά-



Σχήμα 2.2 Το προτεινόμενο μεθοδολογικό πλαίσιο Plug&Chip.

διο της ροής σχεδιασμού, η συνολική λειτουργικότητα του τελικού συστήματος περιγράφεται με μία λύση αποκλειστικού λογισμικού (all-software solution).

Εκκινώντας από αυτή τη λύση του “πλήρους λογισμικού”, δίδεται η δυνατότητα αρχικά να χαρακτηρισθούν λεπτομερώς οι αλγόριθμοι της εφαρμογής, ώστε να καθορισθούν εκείνοι οι πυρήνες που επηρεάζουν ιδιαίτερα την απόδοση του συστήματος. Προς το σκοπό αυτό, οι πυρήνες διακριτοποιούνται σε τρεις κύριες κατηγορίες, στους πυρήνες αυξημένης υπολογιστικής πολυπλοκότητας, στους πυρήνες με αυξημένες απαιτήσεις για επικοινωνία (I/O), καθώς και στους πυρήνες που μπορούν να εκτελεστούν πιο γρήγορα, εάν αξιοποιηθούν τεχνικές παραλληλίας στον πηγαίο κώδικά τους. Για το σκοπό αυτό, μπορεί να χρησιμοποιηθεί μια σειρά εργαλείων λογισμικού (π.χ. Valgrid, VTune, κλπ). Κατά συνέπεια, με βάση τα συμπεράσματα που προέκυψαν από το στάδιο του χαρακτηρισμού των εφαρμογών, καθίσταται δυνατή ο συν-σχεδιασμός υλικού-λογισμικού, συμφώνως προς τα κριτήρια που τέθηκαν ανωτέρω. Το αποτέλεσμα του συν-σχεδιασμού ταξινομεί τη λύση του “πλήρους λογισμικού” σε τρεις κατηγορίες: (i) στην κατηγορία του - εξαρτώμενου από το υλικό - λογισμικού (HW-dependent software) , (ii) σε αυτήν του λογισμικού ελέγχου ροής (control-flow SW) και (iii) των εξειδικευμένων πυρήνων υλικού (custom HW IPs).

Το σύνολο των εξειδικευμένων πυρήνων υλικού αναφέρεται σε περιφερειακές μονάδες και επιταχυντές υλικού, που παρέχουν επικοινωνία με τον εξωτερικό κόσμο του ολοκληρωμένου συστήματος και επιτάχυνση σε υλικό, αντίστοιχα. Με τον όρο “εξαρτώμενο από το υλικό, λογισμικό“, αναφέρονται οι αλγόριθμοι που εκτελούνται στον -ενσωματωμένο ή συμβατικό- επεξεργαστή του τελικού συστήματος. Επίσης, υπάρχει και μια στοιβία λογισμικού, η οποία εκτελείται στο σύστημα υποδοχής (π.χ. συμβατό με την αρχιτεκτονική x86) και η οποία είναι

υπεύθυνη για την υποστήριξη της επικοινωνίας μεταξύ υλικού και "εξαρτώμενου από το υλικό λογισμικού", καθώς και για την παροχή όλων των απαραίτητων εργασιών συγχρονισμού των υπολογιστικών διαδικασιών στα διάφορα επίπεδα. Αυτή η στοίβα λογισμικού αναφέρεται στο Σχήμα 2.2 ως "το λογισμικό ελέγχου ροής". Εκτός από το στάδιο της σχεδίασης, το αναφερόμενο σύστημα υποδοχής (host PC), μπορεί επίσης να είναι και μέρος του τελικού συστήματος. Σε αυτή τη περίπτωση, το λογισμικό ελέγχου ροής, περιλαμβάνει επίσης και το λογισμικό το οποίο εκτελείται στο σύστημα υποδοχής.

Το προτεινόμενο πλαίσιο αντιμετωπίζει επίσης τους περιορισμούς που τίθενται κατά το στάδιο χαρακτηρισμού της εφαρμογής. Συγκεκριμένα, στην περίπτωση κατά την οποία η διαδικασία χαρακτηρισμού εκτελείται σε διαφορετική πλατφόρμα από αυτήν που θα εκτελείται το τελικό σύστημα, τότε είναι πιθανή η εισήγηση ανακριβειών στα εξαγόμενα συμπεράσματα. Για παράδειγμα, η εκτέλεση σε μια διαφορετική πλατφόρμα επιφέρει αλλαγές στο σύνολο εντολών αρχιτεκτονικής (Instruction Set Architecture-ISA), την μικροαρχιτεκτονική, και τον μεταγλωττιστή, με αποτέλεσμα την εισαγωγή αποκλίσεων των μετρικών απόδοσης κατά τη φάση χαρακτηρισμού των εφαρμογών. Συνεπώς ακριβής πληροφορία χαρακτηρισμού μπορεί να αποτιμηθεί κατά το στάδιο που η πρώτη εκδοχή του πρωτότυπου υλικού έχει σχεδιαστεί. Ωστόσο, συνήθως, όταν ένα έργο έχει φτάσει σε αυτό το στάδιο, το μεγαλύτερο μέρος του κόστους προϋπολογισμού έχει απορροφηθεί και η κατάσταση της ροής σχεδίασης χαρακτηρίζεται ως "μη αναστρέψιμη", καθώς ο επανασχεδιασμός, στη περίπτωση που τα επιθυμητά χαρακτηριστικά λειτουργίας δεν έχουν επιτευχθεί, μπορεί να τοποθετήσει το έργο εκτός προϋπολογισμού, ή εκτός χρονοδιαγράμματος εναπόθεσης στην αγορά (time-to-market). Προς την αντιμετώπιση αυτών των δυσχερών καταστάσεων, το προτεινόμενο πλαίσιο προσφέρει την ευελιξία του σταδιακά αναπτυσσόμενου υλικού και τη δοκιμή του συστήματος υπό πραγματικές συνθήκες, από τα πρώιμα στάδια σχεδίασης, έτσι ώστε να καταστούν διαθέσιμες ακριβείς πληροφορίες χαρακτηρισμού κατά το στάδιο της ανάπτυξης και να αξιοποιηθούν για τη αυτο-βελτιστοποίηση του υπό-ανάπτυξη συστήματος, μέσα από ένα βρόχο ανάδρασης.

Το δεύτερο στάδιο του προτεινόμενου πλαισίου ασχολείται με την "Υβριδική Εικονική Προτυποποίηση", η οποία αποτελεί και το βασικό στάδιο της προτεινόμενης μεθοδολογίας. Προκειμένου να υποστηριχθεί η αλληλεπίδραση μεταξύ των προγραμματιστών λογισμικού και των μηχανικών υλικού, έχει υιοθετηθεί η χρήση των μοντέλων TLM-SystemC. Για το σκοπό αυτό, μπορούν να χρησιμοποιηθούν διαφορετικές Στρατηγικές Πληροφοριακών Συστημάτων (Information Systems Strategy, ISS), ωστόσο, στα πλαίσια αυτής της έρευνας, το πλαίσιο Plug&Chip βασίζεται, μεταξύ άλλων, στο OVP [4], δεδομένου ότι είναι δημόσια διαθέσιμο στην κοινότητα και εύκολα προσβάσιμο. Επιπλέον, η αυξημένη ταχύτητα προσομοίωσης που παρέχει το περιβάλλον εξομίωσης OVPSim, διασφαλίζει την απεικόνιση των σύνθετων συστημάτων σε εύλογο χρονικό διάστημα (π.χ. δίνει τη δυνατότητα επεξεργασίας εκατοντάδων εκατομμύρια εντολών προσομοίωσης ανά δευτερόλεπτο). Καθώς τα μοντέλα OVP είναι προσχεδιασμένα, υποστηρίζουν πλήρως τη λειτουργική προσομοίωση ενός πλήρους ενσωματωμένου

συστήματος. Επίσης, δεδομένου ότι τα μοντέλα αυτά είναι συμβατό σε επίπεδο εκτέλεσης με το προσομοιωμένο υλικό, το αναπτυχθέν σύστημα μπορεί να εκτελεστεί στο τελικό σύστημα προορισμού χωρίς λειτουργικές τροποποιήσεις. Με τον τρόπο αυτό, είναι δυνατή ταχύτερος κύκλος ανάπτυξης για τους προγραμματιστές λογισμικού.

Ομοίως, οι μηχανικοί υλικού επωφελούνται από την Υβριδική Εικονική Πλατφόρμα, για την οποία γίνεται λόγος στην παρούσα ενότητα. Δεδομένου ότι η πλατφόρμα αυτή συντίθεται από το περιβάλλον προσομοίωσης OVP και τα μοντέλα TLM/SystemC, επιδεικνύει αυξημένη ευελιξία, η οποία με τη σειρά της αντιμετωπίζει επιτυχώς ένα πλήθος περιορισμών που προκύπτουν στους σχεδιαστές κατά τη φάση του αρχιτεκτονικού σχεδιασμού. Ειδικότερα, τα μοντέλα του περιβάλλοντος OVP χρησιμοποιούνται για να περιγράψουν το βασικό σύστημα προορισμού (π.χ. έναν ενσωματωμένο επεξεργαστή), το οποίο συντονίζει τον έλεγχο ροής της επιθυμητής λειτουργικότητας, ενώ τα μοντέλα TLM/SystemC παρέχουν τη λειτουργικότητα που υλοποιείται στους εξειδικευμένους πυρήνες υλικού, μετά τις αποφάσεις συν-σχεδιασμού του συστήματος.

Μετά την ολοκλήρωση της μοντελοποίησης του συστήματος σε υψηλό επίπεδο και την εξασφάλιση ότι πληροί τις σχεδιαστικές προδιαγραφές, έπεται η ανάπτυξη των εξειδικευμένων πυρήνων υλικού σε γλώσσα HDL. Κατά τη διάρκεια υλοποίησης αυτών των πυρήνων, οι μηχανικοί υλικού δύνανται να τις δοκιμάσουν σταδιακά με την αντικατάσταση της λειτουργικότητας του αντίστοιχου μοντέλου SystemC/TLM, με το αντίστοιχο HDL πρωτότυπο που έχει αναπτυχθεί. Η επικοινωνία μεταξύ της εικονικής πλατφόρμας VP και του FPGA διασφαλίζεται μέσω ενός νέου διαύλου επικοινωνίας, το HotTalk API. Συγκεκριμένα, αυτός ο δίαυλος επικοινωνίας παρέχει τη συνδεσιμότητα μεταξύ της εικονικής πλατφόρμας VP και του εξειδικευμένου υλικού, καθώς και μεταξύ του συστήματος υποδοχής και της εικονικής πλατφόρμας VP, μέσω μίας φυσική διασύνδεσης με τον κεντρικό υπολογιστή (π.χ. Ethernet, PCI, USB, κλπ). Στις επόμενες υποενοτήτες περιγράφεται με περισσότερη λεπτομέρεια η λειτουργικότητα του διαύλου επικοινωνίας HotTalk.

Με τον εν λόγω δίαυλο επικοινωνίας HotTalk API, το προτεινόμενο πλαίσιο παρέχει μια ευρεία στοίβα υλισμικού, δηλαδή υλικού και του συνοδευόμενου λογισμικού διεπαφής χαμηλού επιπέδου, αποτελούμενη από προγράμματα οδήγησης συσκευών στο σύστημα υποδοχής, βιβλιοθήκες στο περιβάλλον OVP και πυρήνες συναλλαγής (transactors) στο FPGA, έτσι ώστε οι σχεδιαστές να ελέγχουν με αποτελεσματικότητα ολόκληρο το σύστημα από την αρχική σχεδιαστική προσέγγιση έως την τελική επαλήθευση του συστήματος, υπό πραγματικές συνθήκες ελέγχου. Όπως αναφέρθηκε προηγουμένως, μια τέτοια σταδιακή σχεδιαστική ροή, παρέχει όλες τις απαραίτητες πληροφορίες σχετικά με τις τεχνικές προδιαγραφές του συστήματος και κατά συνέπεια, δύναται να βελτιστοποιηθεί έτι περαιτέρω το σύστημα είτε με προσθήκη νέων περιορισμών, είτε με την εκ νέου συν-σχεδίασης υλικού/λογισμικού, μέσω του βρόχου ανάδρασης.

Στο τελευταίο στάδιο της προτεινόμενης μεθοδολογίας λαμβάνει χώρα η ολοκλήρωση του συστήματος, όπου οι διάφοροι πυρήνες του συστήματος προορισμού, που περιλαμβάνουν μεταξύ άλλων τον ενσωματωμένο επεξεργαστή, το

επαναδιαμορφούμενο σύστημα υλικού, καθώς και τις μονάδες μνήμης, συντίθενται σε φυσικό επίπεδο. Σε αυτό το στάδιο, το πλαίσιο προβλέπει την αξιολόγηση νέων τεχνολογιών ολοκλήρωσης, όπως η τρισδιάστατη τεχνολογία, ώστε να αποτιμηθούν τα κέρδη ποιότητας αλλά και οι περιορισμοί από τη μετάβαση σε μία τέτοια σύγχρονη πολλά υποσχόμενη τεχνολογία. Προς το σκοπό τούτο, η μεθοδολογία μας επιστρατεύει μια προσέγγιση ποσοτικοποίησης - με αποδεκτή πιστότητα - αναφορικά προς τα δυνητικά οφέλη από τη χρήση αυτής της τεχνολογίας. Σε αντίθεση με υφιστάμενες λύσεις, οι οποίες βασίζονται κυρίως σε ακαδημαϊκά εργαλεία, η αξιολόγηση της τεχνολογίας 3-D, υπό το πρίσμα της παρούσας εργασίας, καθίσταται δυνατή με τη χρήση των αξιόπιστων βιομηχανικών εργαλείων της εταιρείας Cadence.

Η προτεινόμενη μεθοδολογία, η οποία απεικονίζεται στο Σχήμα 2.2, είναι αυτοματοποιημένη με ένα πλήθος εργαλείων CAD. Πρόσθετες πληροφορίες σχετικά με τις προτεινόμενες Host-to-VP και VP-to-HW βιβλιοθήκες επικοινωνίας, καθώς και την αξιολόγηση των 3-D ενσωματωμένων συστημάτων, ακολουθούν στις επόμενες υποενότητες.

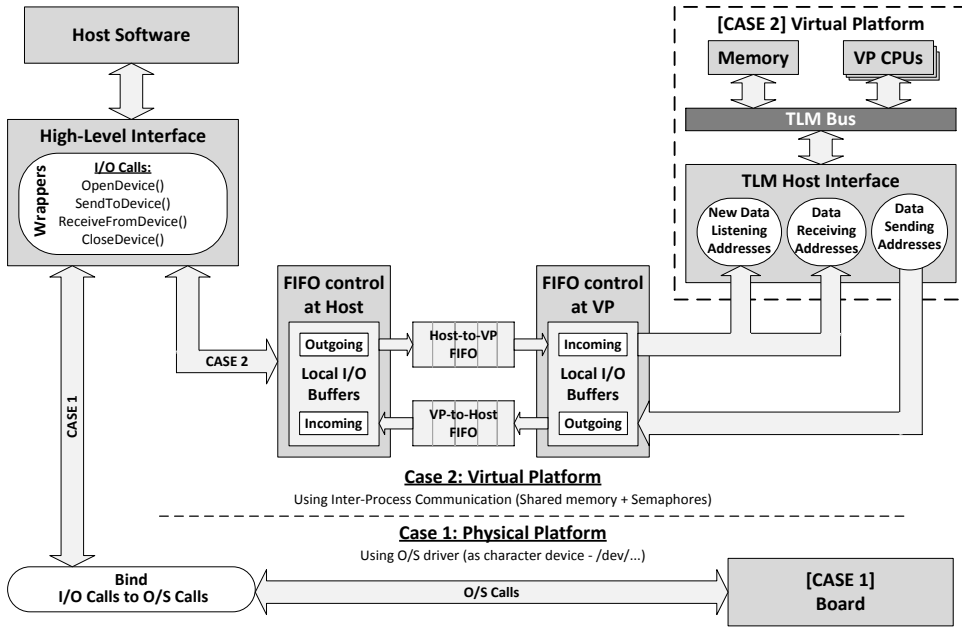
2.3.1. HotTalk API: Διεπαφή Επικοινωνίας Host2VP και VP2HW

Αυτή η υποενότητα περιγράφει με περισσότερες λεπτομέρειες το προτεινόμενο σύστημα επικοινωνίας μεταξύ του υπολογιστή υποδοχής και της εικονικής μηχανής VP, καθώς και μεταξύ της εικονικής μηχανής VP και της πλατφόρμας υλικού, η οποία στα πλαίσια αυτής της εργασίας είναι η συσκευή FPGA. Αυτές οι δύο βιβλιοθήκες, που ονομάζονται Host2VP και VP2HW αντίστοιχα, αποτελούν τον πυρήνα του διαύλου HotTalk API. Επιπλέον, HotTalk API παρέχει ένα πυρήνα συναλλαγής στο FPGA, δηλαδή μια μονάδα εξειδικευμένου υλικού για την επαναδιαμορφούμενη πλατφόρμα, η οποία πραγματοποιεί τη φυσική σύνδεση με τη συσκευή FPGA.

2.3.1.1. Host2VP

Το Σχήμα 2.3 απεικονίζει μία λειτουργική επισκόπηση των εργασιών που υλοποιεί η βιβλιοθήκη Host2VP. Αυτή η βιβλιοθήκη παρέχει μία καθολική διεπαφή προγραμματισμού στο σύστημα υποδοχής προκειμένου να αποφεύγονται οι τροποποιήσεις όταν η εικονική μηχανή αντικαθίσταται με μία πραγματική πλατφόρμα υλοποίησης, όπως συμβαίνει κατά τα τελευταία στάδια σχεδίασης. Προς την κατεύθυνση αυτή, η βιβλιοθήκη παρέχει ένα περιβάλλον υψηλού επιπέδου για την υλοποίηση τεσσάρων βασικών κλήσεων συστήματος (system calls): (i) καθίδρυση επικοινωνίας με συσκευή, (ii) κλείσιμο επικοινωνίας με συσκευή, (iii) αποστολή δεδομένων στη συσκευή και (iv) λήψη δεδομένων από τη συσκευή. Αυτές οι κλήσεις συστήματος μπορούν να χρησιμοποιηθούν σε οποιοδήποτε διαδικασία επικοινωνίας μεταξύ του υπολογιστή υποδοχής και του ενσωματωμένου συστήματος, είτε εικονικού, είτε πραγματικού, χωρίς την τροποποίηση του λογισμικού του συστήματος υποδοχής. Ειδικότερα, στην περίπτωση χρήσης μιας πραγματικής πλατφόρμας (περίπτωση 1 του Σχήματος 2.3), η οποία αναγνωρίζεται ως "συσκευή χαρακτήρων" (character device) από το λειτουργικό σύστημα, οι κλήσεις εισόδου/εξόδου εξυπηρετούνται μέσω των κλήσεων συστήματος του

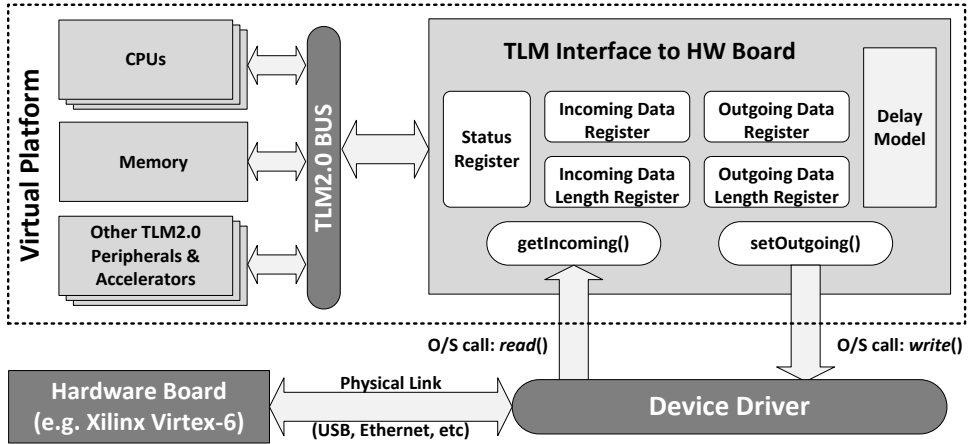
λειτουργικού συστήματος (*open()*, *close()*, *write()* και *read()*, αντίστοιχα), ενώ η συνολική επικοινωνία πραγματοποιείται μέσω των προγραμμάτων οδήγησης της πλατφόρμας.



Σχήμα 2.3 Μηχανισμοί επικοινωνία μεταξύ συστήματος υποδοχής και εικονικής πλατφόρμας VP.

2.3.1.2. VP2HW

Η λειτουργικότητα της βιβλιοθήκης VP2HW, η οποία υλοποιεί την επικοινωνία μεταξύ εικονικής μηχανής VP και της φυσικής πλατφόρμας υλοποίησης, απεικονίζεται στο Σχήμα 2.4. Ομοίως με τη βιβλιοθήκη Host2VP, η βιβλιοθήκη VP2HW υποστηρίζει ένα πλήθος πλατφορμών υλικού. Για αυτό το σκοπό, η πλατφόρμα υλικού αναγνωρίζεται ως μία συσκευή χαρακτηρισμών (character device). Η επιλογή αυτή παρέχει τον κατάλληλο συγχρονισμό μέσω του λειτουργικού συστήματος του υπολογιστή υποδοχής και εγγυάται σταθερή μετάδοση δεδομένων με επιλογές ανίχνευσης και διόρθωσης λαθών. Παρόλο που στα πλαίσια της συγκεκριμένης εργασίας επιλέχθηκε η σύνδεση Ethernet ως μέσο και πρωτόκολλο επικοινωνίας της βιβλιοθήκης VP2HW, ωστόσο, οποιαδήποτε άλλο πρωτόκολλο μπορεί να χρησιμοποιηθεί με ελάχιστη προγραμματιστική προσπάθεια. Επιπλέον, η βιβλιοθήκη υποστηρίζει μετάδοση και λήψη δεδομένων με ελάχιστη πλήθος bytes στις κεφαλίδες του πρωτοκόλλου Ethernet (raw ethernet), προκειμένου να επιτύχει την ελάχιστη δυνατή επιβάρυνση στην επικοινωνία μεταξύ VP και πλατφόρμας υλικού.



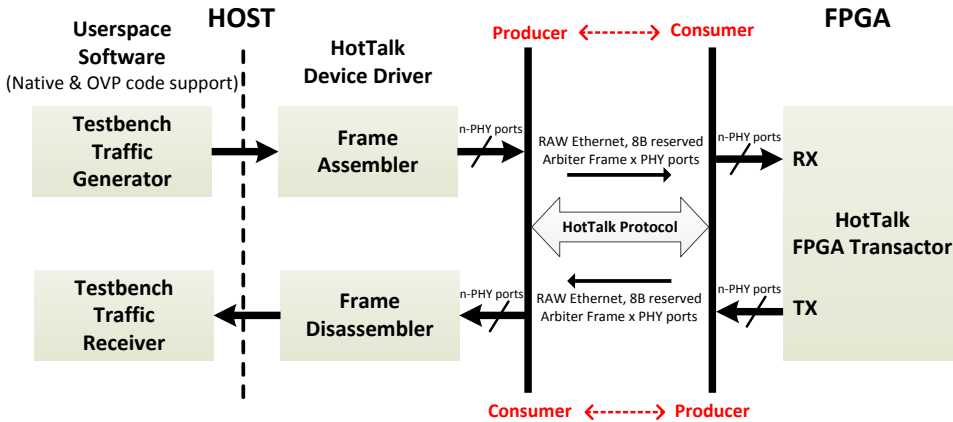
Σχήμα 2.4 Μηχανισμοί επικοινωνία μεταξύ εικονικής πλατφόρμας VP και φυσικής πλατφόρμας υλοποίησης.

2.3.1.3. Υλοποίηση του διαύλου επικοινωνίας HotTalk

Προκειμένου να επιτευχθεί η επικοινωνία μεταξύ εικονικής μηχανής και υλικού, VP2HW, το προγραμματιστικό περιβάλλον του διαύλου HotTalk παρέχει ένα πρωτόκολλο επικοινωνίας το οποίο υλοποιείται από μία εξειδικευμένη μονάδα υλικού στο FPGA, τον πυρήνα συναλλαγής (transactor). Αυτή η μονάδα χειρίζεται τα εισερχόμενα και εξερχόμενα δεδομένα, δημιουργώντας έτσι τη σύνδεση μεταξύ των πυρήνων υλικού και της θύρα επικοινωνίας, όπου στην προτεινόμενη προσέγγιση χρησιμοποιείται το η φυσική ζεύξη μέσω Ethernet. Καθώς οι μονάδες υλικού αναπτύσσονται, είναι δυνατόν να εξακριβωθεί η λειτουργικότητά τους καθώς και η ποιότητας επεξεργασίας τους σε πραγματικά σενάρια λειτουργίας, χρησιμοποιώντας τον πυρήνα συναλλαγής HotTalk. Η προσέγγιση αυτή αποδείχθηκε ότι ανιχνεύει σφάλματα και προβλήματα απόδοσης, τα οποία δεν μπορούν να προβλεφθούν όταν χρησιμοποιούνται σενάρια ελέγχου που είναι ανεπτυγμένα αποκλειστικά σε λογισμικό.

Η συνολική ροή επικοινωνίας του πυρήνα συναλλαγής βασίζεται στη λογική παραγωγού-καταναλωτή και απεικονίζεται στο Σχήμα 2.5. Συγκεκριμένα, τα σενάρια ελέγχου (testbenches) στο σύστημα υποδοχής παράγουν τα αναγκαία δεδομένα εισόδου τα οποία προορίζονται για τις μονάδες υλικού. Στη συνέχεια, το πρόγραμμα οδήγησης συσκευής ομαδοποιεί τα δεδομένα εισόδου σε πλαίσια πρωτοκόλλου. Το πρωτόκολλο είναι σχεδιασμένο στην βάση των πρωταρχικών πλαισίων Ethernet (raw Ethernet packets). Προκειμένου να μεγιστοποιηθεί η απόδοση επικοινωνίας, το μέγεθος των πακέτων κεφαλίδας περιορίζεται σε 8 bytes, τα οποία αφορούν τη διεύθυνση προορισμού (6 bytes) και το μήκος του πλαισίου (2 bytes). Το υπόλοιπο πακέτο, το οποίο χρησιμοποιείται για τη μεταφορά δεδομένων, έχει μήκος MTU-8 bytes, όπου η ποσότητα MTU (maximum transmission unit) εκφράζει το μέγεθος της μεγαλύτερης ποσότητας δεδομένων που μπορούν να μεταδοθούν με μία χειραψία και για τη φυσική ζεύξη του

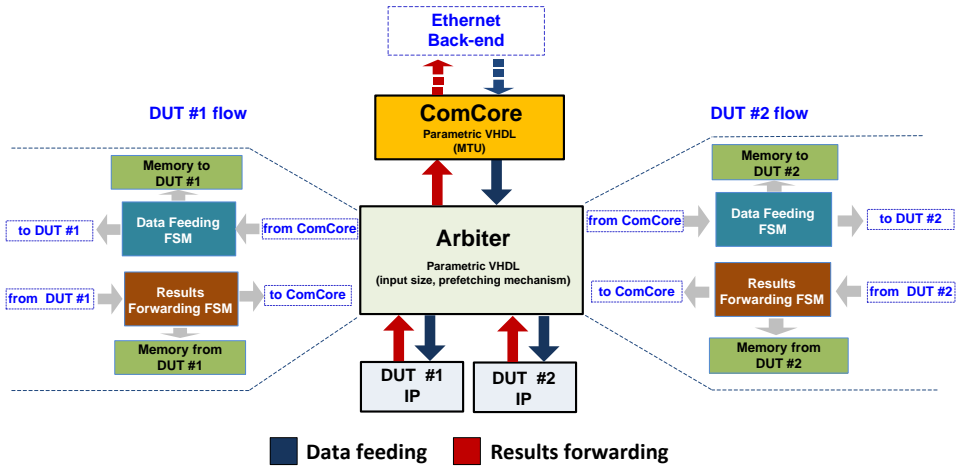
Ethernet έχει τυπικές τιμές έως 1500 bytes (Ethernet II) αλλά μπορεί να φτάσει έως 9198 bytes (Ethernet Jumbo Frames). Στη μετρήσεις που ακολουθούν στη συγκεκριμένη εργασία, χρησιμοποιείται MTU μεγέθους 1500 bytes. Στην πλευρά του FPGA, ο πυρήνας συναλλαγής HotTalk είναι υπεύθυνος για την εξυπηρέτηση των αιτήσεων αποστολής/λήψης από τον ελεγκτή διεπαφής της κάρτας δικτύου (NIC) με τους πυρήνες υλικού. Μια παρόμοια προσέγγιση ακολουθείται για την αντίστροφη κατεύθυνση δεδομένων, στην περίπτωση δηλαδή που το FPGA επιστρέφει τα αποτελέσματα από τους πυρήνες υλικού στο σύστημα υποδοχής.



Σχήμα 2.5 Η ροή επικοινωνίας στον πυρήνα συναλλαγής HotTalk.

Όπως έχει ήδη αναφερθεί, η επικοινωνία μεταξύ των ανεπτυγμένων πυρήνων υλικού IP και της εικονικής μηχανής έχει σημαντικό ρόλο στην απόδοση του συνολικού συστήματος. Για να βελτιστοποιηθεί η μεταφορά δεδομένων, σχεδιάστηκε ένα σύστημα επικοινωνίας, το οποίο απεικονίζεται στο Σχήμα 2.6, το οποίο συνδυάζει υφιστάμενους και νέους εξειδικευμένους πυρήνες υλικού ανοικτού λογισμικού για την υλοποίηση της επικοινωνίας με σύνδεση raw Ethernet. Το σύστημα αυτό υποστηρίζεται από τρία διακριτά υποσυστήματα, (i) τη μονάδα ComCore, (ii) τη μονάδα διαιτησίας και (iii) τη μονάδα επαλήθευσης DUT (Design Under Test).

Ξεκινώντας την περιγραφή ιεραρχικά από το κατώτερο συστημικό επίπεδο, βρίσκεται το ολοκληρωμένο κύκλωμα PHY chip το οποίο υλοποιεί τις λειτουργίες φυσικού επιπέδου, όπως είναι η κωδικοποίηση χαρακτήρων, καθώς και η μετάδοση, λήψη και αποκωδικοποίηση αυτών. Στην παρούσα εργασία, χρησιμοποιήθηκε το ολοκληρωμένο chip Marvell 88E111 το οποίο είναι τοποθετημένο στη συσκευή HTG-V6-PCIE-L240T-2 [15] η οποία χρησιμοποιήθηκε για το στάδιο της επαλήθευσης του προτεινόμενου πλαισίου. Στο επόμενο επίπεδο έχει αναπτυχθεί ένας εξειδικευμένος ελεγκτής Rx/Tx ο οποίος βασίζεται στο πυρήνα υλικού ETHMAC, έναν πυρήνα Ethernet MAC διαθέσιμο από το αποθετήριο OpenCores [16], και ο οποίος υλοποιεί τις λειτουργίες επιπέδου δικτύου CSMA/CD LAN, σύμφωνα με το πρότυπο IEEE 802.3. Τα επόμενα επίπεδα



Σχήμα 2.6 Η αρχιτεκτονική του προτεινόμενου πυρήνα συναλλαγής.

αφορούν την υποστήριξη των πυρήνων υλικού μέσω της διανομής και του συγχρονισμού των πακέτων δεδομένων και των αποτελεσμάτων. Συγκεκριμένα, η μονάδα “ComCore” είναι υπεύθυνη για το διαχωρισμό/σύνθεση μεγάλων πακέτων δεδομένων σε πλαίσια μεγέθους 1500 bytes (MTU). Επιπλέον, η μονάδα ComCore χειρίζεται αφαιρετικά τη σηματοδότηση του ελεγκτή από το φυσικό επίπεδο MAC, ώστε να παρέχει μία απλοποιημένη διεπαφή Wishbone στις υπόλοιπες μονάδες του FPGA. Αυτή η αφαιρετικότητα βοηθάει τους σχεδιαστές στην συγγραφή πυρήνων σε γλώσσα HDL με απλοποιημένο πρωτόκολλο διεπαφής, σε σχέση με τη διαχείριση όλου του πρωτοκόλλου Wishbone, το οποίο χρησιμοποιείται από το ETHMAC.

Η μονάδα διαιτησίας (Arbiter) ελέγχει το κανάλι επικοινωνίας για την αποφυγή συγκρούσεων μεταξύ των μονάδων επεξεργασίας. Μεταξύ άλλων, η μονάδα αυτή επίσης παρέχει προτεραιότητα στην επικοινωνία για τους πυρήνες υλικού, η οποία βασίζεται σε αλγόριθμο round-robin. Παράλληλα η μονάδα διαιτησίας παρέχει υποστήριξη για αλυσιδωτή επεξεργασία των πυρήνων υλικού, μέσω της επιλογής για προανάκληση δεδομένων (prefetching). Η προανάκληση δεδομένων μπορεί να γίνει την ώρα που οι υπόλοιποι πυρήνες υλικού επεξεργάζονται δεδομένα, οπότε και η μονάδα διαιτησίας είναι αδρανής. Αυτή η λειτουργία εκμεταλλεύεται τις δυνατότητες άμεσης πρόσβασης μνήμης (Direct memory access, DMA) τις οποίες διαθέτουν οι περισσότερες σύγχρονες κάρτες δικτύου NIC. Με αυτό το τρόπο οι λειτουργίες αποστολής/λήψης μπορούν να εκτελούνται παράλληλα από εξειδικευμένα ολοκληρωμένα συστήματα PHY chips, ενώ στο επίπεδο του χρήστη, διαφορετικά νήματα εκτέλεσης μπορούν να εκτελούν ανεξάρτητα τμήματα ενός αλγόριθμου, τα οποία έχουν χρονο-προγραμματιστεί δυναμικά από το λειτουργικό σύστημα. Επιπλέον, τονίζεται ότι το προτεινόμενο σύστημα μπορεί να υποστηρίξει πολλαπλά κανάλια επικοινωνίας Ethernet μεταξύ της εικονικής μηχανής και του FPGA. Συγκεκριμένα μπορεί να υλοποι-

ηθεί ένα κανάλι ανά μονάδα επαλήθευσης DUT, ανάλογα με το πλήθος των διαφορετικών θυρών Ethernet που υπάρχουν διαθέσιμες στο υπό-ανάπτυξη σύστημα/FPGA. Για παράδειγμα, το Σχήμα 2.6 απεικονίζει μία μονάδα διαίτησίας με δύο ξεχωριστές μονάδες επαλήθευσης DUT.

Σε αυτή τη περίπτωση, η μονάδα διαίτησίας χρησιμοποιεί κεφαλίδες στο μεταδιδόμενο πακέτο, προκειμένου να ορίσει συγκεκριμένη μονάδα λήψης/μετάδοσης από/προς την εικονική μηχανή. Όσον αφορά την αρχιτεκτονική της μονάδας διαίτησίας, χρησιμοποιούνται δύο ξεχωριστές μηχανών πεπερασμένων καταστάσεων (Finite-State Machines, FSMs), κάθε μία από τις οποίες είναι αφιερωμένη στη διαδικασία μετάδοσης και λήψης αντίστοιχα. Αυτή η προσέγγιση σχεδίασης επιτρέπει ανεξάρτητη τροφοδότηση δεδομένων και διαβίβαση αποτελεσμάτων, υπό την προϋπόθεση ότι η μνήμη FIFO που είναι δεσμευμένη σε κάθε FSM, είναι αρκετά μεγάλη προκειμένου να αποθηκεύσει τα σχετικά δεδομένα. Συνεπώς, η μονάδα διαίτησίας μπορεί να υποστηρίξει έναν παραμετρικό μηχανισμό προανάκλησης, έτσι ώστε η μονάδα επαλήθευσης DUT να λειτουργεί σε υψηλό ποσοστό χρησιμοποίησης χωρίς να περιμένει την επικοινωνία VP2HW σε κάθε βήμα επανάληψης του αλγορίθμου.

Σε σύγκριση με τις υφιστάμενες υλοποιήσεις, σε επίπεδο υλικού, για την επικοινωνία PC-FPGA [16], η προτεινόμενη μονάδα ComCore, η οποία αναπτύχθηκε για το πλαίσιο Plug&Chip, προσφέρει συνεχή (batch) τρόπο μετάδοσης-λήψης των δεδομένων καθώς και ανίχνευση και επιδιόρθωση σφαλμάτων. Σε συνδυασμό με ένα εξειδικευμένο πρόγραμμα οδήγησης συσκευής για τον πυρήνα Linux, το οποίο αναπτύχθηκε αποκλειστικά για το προτεινόμενο πλαίσιο, προσφέρεται επίσης η δυνατότητα ανάκτησης/επανεκκίνησης της ροής επικοινωνίας, σε περίπτωση σφάλματος στην επικοινωνία του συστήματος. Αυτό το πρόγραμμα οδήγησης συσκευής, το οποίο παρέχει την υποστήριξη της επικοινωνίας μεταξύ της κάρτας δικτύου Ethernet (χρησιμοποιείται το πρότυπο MII) και της εικονικής μηχανής VP, ουσιαστικά, προσφέρει το χειρισμό διακοπών της διεπαφής δικτύου NIC προκειμένου να καταστεί δυνατή η ασύγχρονη επικοινωνία μέσα στο σύστημα, παρέχοντας ταυτόχρονα ένα απλοποιημένο και αφαιρετικό περιβάλλον προγραμματισμού στο σχεδιαστή. Αναλυτικότερα, έχουν υλοποιηθεί συναρτήσεις ανάγνωσης και εγγραφής από/σε ένα ειδικό αρχείο συστήματος, το οποίο σε επίπεδο προγραμματισμού στον πυρήνα αποτελεί μία συσκευή χαρακτήρα του Linux (character device file). Με αυτόν τον τρόπο είναι δυνατή η απλοποιημένη μεταφορά δεδομένων κατά την οποία η ομάδα σχεδιασμού χρησιμοποιεί αποκλειστικά τις προαναφερθείσες συναρτήσεις μεταφοράς δεδομένων και το υπόλοιπο σύστημα αναλαμβάνει την αυτούσια μεταφορά τους στις μνήμες FIFO του FPGA.

2.3.2. Αξιολόγηση της τρισδιάστατης τεχνολογίας ολοκλήρωσης για ενσωματωμένα συστήματα

Το πλαίσιο Plug&Chip πέραν της ευελιξίας που προσφέρει για ταχεία ανάπτυξη συστημάτων μέσω της εικονικής προτυποποίησης σε επαναδιαμορφούμενες πλατφόρμες, προτείνει επίσης ένα μεθοδολογικό πλαίσιο για την υλοποίηση των σχεδιασμών, που έχουν προτυποποιηθεί από τα προηγούμενα στάδια σχε-

δίασης, σε εξειδικευμένα ολοκληρωμένα συστήματα ASIC. Η προσέγγιση αυτή αποτελεί ένα επιπλέον βήμα σχεδιαστικής ολοκλήρωσης, κατά το οποίο μπορούν να επιτευχθούν καλύτερες μετρικές QoR, αντισταθμίζοντας το υψηλό κόστος φυσικού επιπέδου σχεδίασης, όταν το μέγεθος παραγωγής είναι αρκετά μεγάλο. Προς αυτή τη κατεύθυνση αναζητήθηκε η περισσότερο σύγχρονη και ταυτόχρονα πολλά υποσχόμενη τάση στη σχεδίαση φυσικού επιπέδου, η οποία ακολουθώντας και τις τεχνολογικές προβλέψεις [2] είναι η τρισδιάστατη ολοκλήρωση 3-D. Με αυτό το τρόπο γίνεται προσπάθεια διερεύνησης της εφαρμοσιμότητας και της αποτελεσματικότητας σχεδιασμού από την υιοθέτησή της τρισδιάστατης τεχνολογίας. Σημειώνεται επίσης ότι η παρούσα υποενοότητα αποτελεί μια επισκόπηση εφαρμογής της γενικότερης μεθοδολογίας για σχεδίαση ολοκληρωμένων κυκλωμάτων 3-D που προτείνει αυτή η διδακτορική διατριβή και η οποία παρουσιάζεται αναλυτικότερα στην ενότητα 3.5 του Κεφαλαίου 3.

Παρόλο που τα πλεονεκτήματα της 3-D τεχνολογίας σχεδίασης φαίνονται προφανή από μία θεωρητική σκοπιά [2], [1], η αποτελεσματικότητα αυτής της τεχνολογίας στηρίζεται σε μεγάλο βαθμό στην απόδοση των αντίστοιχων εργαλείων σχεδίασης CAD. Ακόμη και πριν τη διάθεση εμπορικών εργαλείων για φυσική σχεδίαση 3-D υπάρχει ανάγκη για εργαλεία εύρεσης της κατάλληλης αρχιτεκτονικής. Τέτοια εργαλεία επιτρέπουν στους σχεδιαστές την εξερεύνηση μιας σειράς από διαφορετικές εναλλακτικές υλοποιήσεις 3-D, οι οποίες μέσα από μία διαδικασία αξιολόγησης παρέχουν τις βέλτιστες λύσεις Pareto αναφορικά προς τα επιθυμητά χαρακτηριστικά. Αυτή η διαδικασία συμβαίνει συνήθως σε πρώιμα στάδια του κύκλου σχεδιασμού ενός νέου προϊόντος, όπου οι σχεδιαστές καθορίζουν αποφάσεις υψηλού επιπέδου όπως η επιλογή της τεχνολογίας ολοκλήρωσης, η αρχιτεκτονική υψηλού επιπέδου αναφορικά με τους περιορισμούς του φυσικού σχεδιασμού κτλ.

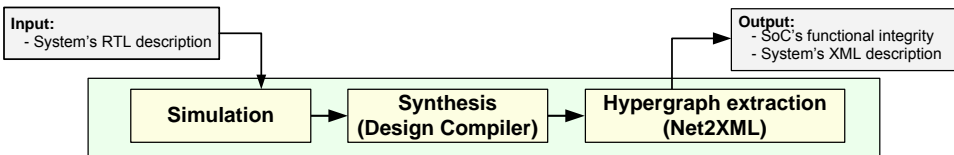
Σε αυτή την υποενοότητα περιγράφεται το τμήμα του προτεινόμενου πλαισίου Plug&Chip το οποίο ασχολείται με την εκτίμηση των μετρικών απόδοσης για εξειδικευμένα συστήματα SoC, υιοθετώντας τη σύγχρονη τεχνολογία ολοκλήρωσης 3-D. Κατά τη διάρκεια αυτού του σταδίου ένα σύνολο από σχεδιαστικές αποφάσεις πρέπει να διερευνηθούν ως προς της αποτελεσματικότητά τους. Μεταξύ αυτών, στα πλαίσια αυτής της εργασίας, μελετάται η επιλογή των σημείων κατάτμησης της αρχιτεκτονικής, η κατανομή των πυρήνων υλικού στα επίπεδα πυριτίου και η επιλογή της διασύνδεσης των σημάτων μεταξύ των επιπέδων πυριτίου. Με την εισηγούμενη συστηματική και διακριτή ροή σχεδιασμού καθίσταται δυνατή η εκτίμηση αποδοτικότητας για σχεδιασμούς 3-D κυκλωμάτων. Η προτεινόμενη μεθοδολογία υποστήριξης ταχείας αξιολόγησης των 3-D SoCs αποτελείται από τρία διακριτά βήματα τα οποία συνοψίζονται ως εξής:

- *Προ-επεξεργασία:* Επαλήθευση της λειτουργικής ακεραιότητας του σχεδιασμού και εξαγωγή σε περιγραφή IP-XACT [17]. Η περιγραφή IP-XACT είναι μια μορφή XML που ορίζει και περιγράφει ηλεκτρονικές μονάδες και ιεραρχικούς σχεδιασμούς.
- *Δημιουργία 3-D στοίβας:* Δημιουργία της 3-D στοίβας και καθορισμός επικοινωνίας μεταξύ των επιπέδων πυριτίου.

- *Συστημική προτυποποίηση 3-D*: Υλοποίηση σε φυσικό επίπεδο μάσκας πυριτίου του 3-D SoC και επαλήθευση ορθής λειτουργίας του τελικού συστήματος.

2.3.2.1. Προ-επεξεργασία

Το πρώτο βήμα της προτεινόμενης μεθοδολογία απεικονίζεται στο Σχήμα 2.7. Είσοδος της ροής εργασιών αποτελεί η περιγραφή, σε συνθέσιμη μορφή, του σχεδιασμού σε μία γλώσσα HDL. Ο σχεδιασμός εξομοιώνεται με διάφορους περιορισμούς (όπως π.χ. η περίοδος του ρολογιού, η οργάνωση των μνημών, η ιεραρχία του σχεδιασμού κτλ.) προκειμένου να επαληθευτεί η λειτουργικότητα του συστήματος.



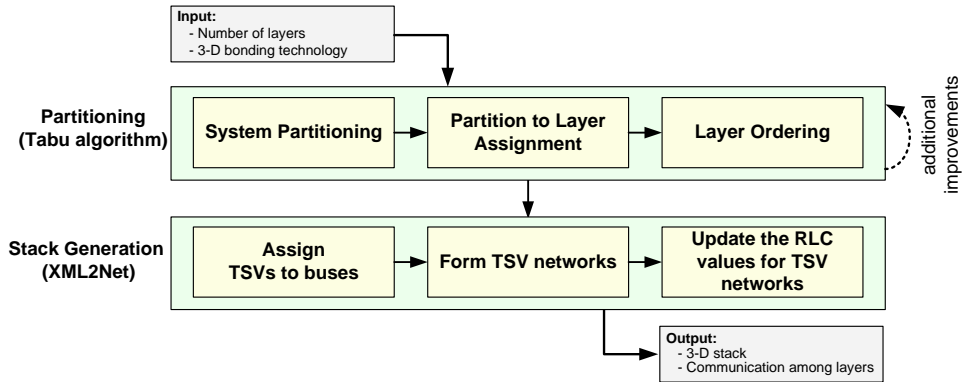
Σχήμα 2.7 Διεργασίες κατά το στάδιο προ-επεξεργασίας.

Στη συνέχεια, καθορίζεται η ιεραρχία της 3-D αρχιτεκτονικής. Διαφορετικά σενάρια επιπέδων ιεραρχίας υποστηρίζονται από το προτεινόμενο πλαίσιο, το καθένα από τα οποία παρουσιάζει πλεονεκτήματα και μειονεκτήματα. Για παράδειγμα, η περιγραφή ενός συστήματος σε επίπεδο λειτουργικής μονάδας οδηγεί σε αδρομερή ιεραρχία, ενώ η περιγραφή σε επίπεδο πύλης οδηγεί σε λεπτομερή ιεραρχία. Όσο πιο λεπτομερής είναι η ιεραρχία του σχεδιασμού, τόσο μεγαλύτερη ευελιξία μπορεί να επιτευχθεί από τα εργαλεία σχεδίασης, καθώς όλος ο σχεδιασμός επεξεργάζεται εξολοκλήρου δίχως επί μέρους περιορισμούς. Ωστόσο αυτή η προσέγγιση αυξάνει την υπολογιστική πολυπλοκότητα για την εξερεύνηση της αρχιτεκτονικής 3-D. Πρέπει να σημειωθεί ότι λόγω της ιδιαιτερότητας αυτής της επιλογής, πρέπει να εφαρμοστεί προσεκτική μελέτη για την επιλογή της ιεραρχίας, καθώς μία υπο-βέλτιστη λύση θα μπορούσε να επισκιασει τη βελτίωση των επιδόσεων που προσφέρει η 3-D ολοκλήρωση. Για τη παρούσα εργασία, στο πλαίσιο Plug&Chip, έχει επιλεγεί, χωρίς να επηρεάζεται η γενικότητα εφαρμογής του προτεινόμενου πλαισίου, η διατήρηση της ιεραρχίας του συστήματος μεταξύ ανομοιογενών μονάδων, όπως για παράδειγμα μεταξύ λογικών μονάδων και μνήμης, ενώ κάθε μονάδα αποδομείται εσωτερικά από οποιαδήποτε ιεραρχία (design flatten), προκειμένου να μεγιστοποιηθεί η βελτίωση των επιδόσεων.

Έπειτα από τον καθορισμό της ιεραρχίας του συστήματος, ο σχεδιασμός συντίθεται με το εργαλείο *Synopsys Design Compiler*. Εφόσον σε αυτό το στάδιο επιτευχθούν οι ζητούμενοι περιορισμοί, η έξοδος της σύνθεσης μεταφράζεται σε μία αντίστοιχη περιγραφή XML. Αυτή η διαδικασία υλοποιείται μέσω ενός νέου εργαλείου, υπό την ονομασία *Net2XML*.

2.3.2.2. Δημιουργία 3-D στοιβάς

Η εξαγόμενη περιγραφή XML, η οποία αναπαριστά την netlist μετά το στάδιο της σύνθεσης, τροφοδοτείται στο δεύτερο στάδιο της προτεινόμενης μεθοδολογίας, το οποίο απεικονίζεται στο Σχήμα 2.8 και είναι υπεύθυνο για τη δημιουργία της 3-D στοιβάς, δηλαδή της βασικής αρχιτεκτονικής του σχεδιασμού, πριν το τελικό φυσικό στάδιο υλοποίησης.



Σχήμα 2.8 Διεργασίες κατά το στάδιο της δημιουργίας 3-D στοιβάς.

Αρχικά, ο σχεδιασμός σε επίπεδο RTL netlist κατατέμνεται σε έναν αριθμό υποσυνόλων (system partitioning). Διαφορετικοί στόχοι βελτιστοποίησης μπορούν να τεθούν κατά τη διάρκεια αυτής της διαδικασίας, όπως για παράδειγμα η ελαχιστοποίηση των συνδέσεων μεταξύ των επιπέδων, τηρώντας ταυτόχρονα κάποιους περιορισμούς, όπως είναι η διατήρηση της DRAM και της λογικής σε διαφορετικά διαμερίσματα. Προηγούμενες μελέτες έχουν δείξει ότι η διαδικασία της κατάτμησης μπορεί να προσφέρει αυξημένη ευελιξία στη συνολική σχεδίαση, όταν ο αριθμός των υποσυνόλων είναι υψηλότερη σε σύγκριση με τον αντίστοιχο αριθμό των επιπέδων πυριτίου [18].

Το επόμενο βήμα είναι η ανάθεση των υποσυνόλων στα επίπεδα πυριτίου του 3-D συστήματος (partitioning to layer assignment). Κατά τη διάρκεια αυτού του σταδίου, σημαντικό ρόλο έχουν η διαδικασία παραγωγής μέσω της αξιοποιήσιμης ποσότητας πυριτίου ανά παραγωγική μονάδα (yield) και οι παράμετροι κόστους μέσω του αριθμού των TSVs. Συγκεκριμένα, σε κάθε επίπεδο πυριτίου μπορούν να ανατεθούν μόνο τεχνολογικά συμβατές μονάδες, ενώ κάθε ένα από τα επίπεδα πρέπει να υφίσταται επαρκής αξιοποιήσιμη επιφάνεια πυριτίου. Ακολουθεί η κατασκευή ενός πρωτότυπου του σχεδιασμού εφόσον ληφθεί απόφαση για τη διάταξη της σειράς των επιπέδων στη 3-D στοιβά (layer ordering) και για την επιλογή της 3-D τεχνολογίας συγκόλλησης. Παρόλο που το προτεινόμενο πλαίσιο μπορεί να αξιολογήσει εναλλακτικές τεχνολογίες συγκόλλησης, όπως η TSV (Face-to-back), η Face-to-Face, καθώς και η wire-bond, στη παρούσα εργασία παρέχονται αποτελέσματα μόνο για την προσέγγιση TSV, λόγω του ότι οδηγεί στη μέγιστη βελτίωση των επιδόσεων [2], [1].

Οι διαδικασίες της κατάτμησης του σχεδιασμού, της ανάθεσης των υπο-

μονάδων στα επίπεδα πυριτίου και της διάταξης της σειράς των επιπέδων στη 3-D στοίβα, υποστηρίζονται από νέα εργαλεία που βασίζονται στον προτεινόμενο αλγόριθμο *Tabu*. Μια πρώτη έκδοση αυτού του αλγορίθμου παρουσιάστηκε στην εργασία [19], αλλά για τις ανάγκες της παρούσας εργασίας έχει επεκταθεί προκειμένου να καλύπτει τα ειδικά χαρακτηριστικά βελτιστοποίησης των 3-D ASIC πλατφορμών.

Η διαδικασία κατάτμησης παρέχει ως έξοδο πληροφορίες σχετικά με το μέρος του σχεδιασμού που υλοποιείται σε κάθε επίπεδο πυριτίου, καθώς και την αντίστοιχη συνδεσιμότητα μεταξύ των επιπέδων. Αυτές οι πληροφορίες εισέρχονται στο νέο εργαλείο *XML2Net*, το οποίο αναθέτει μια σειρά από TSVs στα σήματα που διατρέχουν πολλαπλά επίπεδα πυριτίου. Σημειώνεται ότι για την δρομολόγηση ενός σήματος μεταξύ δύο διαφορετικών επιπέδων πυριτίου i και j , δεσμεύεται σημαντική επιφάνεια πυριτίου και στα δύο επίπεδα. Το προτεινόμενο πλαίσιο μπορεί να διαχειριστεί τόσο διακριτές συνδέσεις TSVs όσο και ομαδοποιημένες. Στην παρούσα εργασία μελετούνται οι δεύτερες, καθώς θέτουν λιγότερους περιορισμούς στο εργαλείο δρομολόγησης και επομένως προσφέρουν ταχύτερη και αποδοτικότερη ροή σχεδίασης. Οι συστοιχίες TSV που παρέχουν δυνατότητα ομαδικής σύνδεσης μεταξύ των γειτονικών επιπέδων i και j , συνδέονται μέσω μονοπατιών ειδικού σκοπού δρομολόγησης, τα οποία ονομάζονται δίκτυα TSV, (*TSV networks*). Όπως αναλύεται στην επόμενη υποενότητα, τα δίκτυα σχεδιάζονται στα συμβατικά εργαλεία με πρόσθετα στρώματα μετάλλου, τα οποία εμφανίζουν παραμετροποιησιμα χαρακτηριστικά *RLC*, έτσι ώστε να προσομοιώνουν τα αντίστοιχα ηλεκτρικά χαρακτηριστικά των TSVs συγκεκριμένων βιβλιοθηκών και μοντέλων 3-D τεχνολογιών.

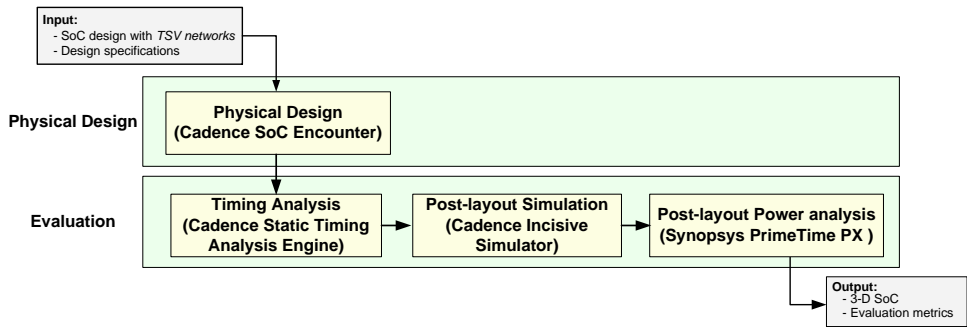
2.3.2.3. Συστημική προτυποποίηση 3-D

Το τελευταίο στάδιο του προτεινόμενου πλαισίου, το οποίο απεικονίζεται στο Σχήμα 2.9, ασχολείται με την υλοποίηση σε φυσικό επίπεδο μάσκας πυριτίου της στοίβας 3-D. Συγκεκριμένα, κατά τη διάρκεια αυτού του σταδίου πραγματοποιείται η χωροθέτηση των μονάδων του σχεδιασμού (*floor-planning*), η δημιουργία του δικτύου διανομής ενέργειας στο τσιπ (*power/ground network generation*), η τοποθέτηση των κελιών του σχεδιασμού, βάση της βιβλιοθήκης της τεχνολογίας (*physical library cells placement*) και η γενική/ειδική δρομολόγηση των σημάτων του σχεδιασμού (*global/detail signal routing*). Αυτές οι διαδικασίες επιτελούνται από το εργαλείο *Cadence SoC Encounter*. Στο προτεινόμενο πλαίσιο χρησιμοποιούνται συμβατικά βιομηχανικά εργαλεία σχεδίασης 2-D, και συνεπώς δεν είναι δυνατόν να τροποποιηθεί ο πηγαίος κώδικάς τους και κατ'επέκταση η λειτουργικότητά τους προκειμένου να υποστηρίξουν 3-D σχεδιασμούς. Ως εκ τούτου, εισάγονται νέες έννοιες, εργαλεία και βιβλιοθήκες σχεδίασης έτσι ώστε να υποστηρίξουν την τρίτη διάσταση σχεδίασης. Για το σκοπό αυτό εισάγονται οι παρακάτω έννοιες:

- *Εικονικά επίπεδα (Virtual layers)*: Το προτεινόμενο πλαίσιο ορίζει την έννοια του εικονικού επιπέδου προκειμένου να χαρακτηρίσει τις μονάδες υλικού που τοποθετούνται σε διαφορετικά επίπεδα του 3-D SoC. Συνεπώς καθώς το εργαλείο *Cadence SoC Encounter* αναγνωρίζει μόνο ένα επίπεδο

πυριτίου, η έννοια των εικονικών επιπέδων επιτρέπει την συνύπαρξη στον ίδιο 2-D σχεδιασμό πολλαπλών επιπέδων πυριτίου, προσομοιώνοντας ένα 3-D σχεδιασμό.

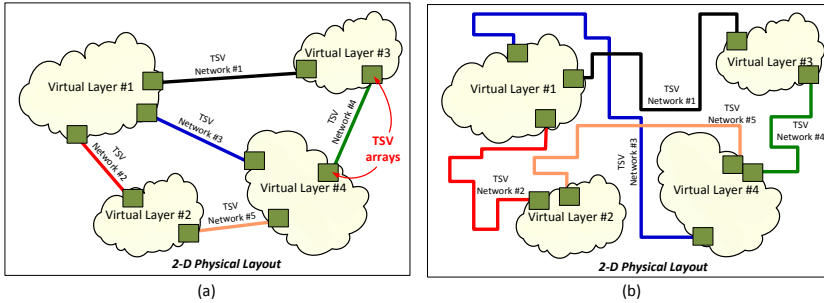
- **Δίκτυα TSV:** Αυτά τα δίκτυα αναφέρονται στα μονοπάτια δρομολόγησης των μονάδων υλικού που ανήκουν σε διαφορετικά επίπεδα πυριτίου του 3-D σχεδιασμού. Εφόσον το εργαλείο *Cadence SoC Encounter* αναγνωρίζει μόνο ένα επίπεδο πυριτίου, οι συνδέσεις μεταξύ των εικονικών επιπέδων πραγματοποιούνται μέσω των ειδικών δικτύων TSV. Αυτό επιτυγχάνεται μέσω της διατήρησης της τοποθέτησης των δικτύων TSV στην ίδια σχετική γεωμετρική θέση (x, y) ανάμεσα στα γειτονικά εικονικά επίπεδα πυριτίου. Πρακτικά τα δίκτυα TSV υλοποιούνται ως επιπλέον επίπεδα μετάλλου στην βιβλιοθήκη τεχνολογίας, στα οποία αποδίδονται συγκεκριμένα ηλεκτρικά χαρακτηριστικά αντίστασης (R), χωρητικότητας (C) και επαγωγής (L) ανάλογα με τις παραμέτρους RLC της τεχνολογίας TSV που χρησιμοποιείται. Στα πλαίσια αυτής της εργασίας χρησιμοποιήθηκαν τα ηλεκτρικά χαρακτηριστικά TSV που βρίσκονται στην εργασία [20].



Σχήμα 2.9 Διεργασίες κατά το στάδιο της συστημικής προτυποποίησης 3-D.

Προς καλύτερη διευκρίνηση των παραπάνω όρων, το Σχήμα 2.10 απεικονίζει ένα παράδειγμα σχεδίασης στο οποίο ένας σχεδιασμός μοντελοποιείται ως ένα 3-D chip, αποτελούμενο από τέσσερα φυσικά επίπεδα πυριτίου. Συγκεκριμένα, το Σχήμα 2.10(α) απεικονίζει την κατάτμηση του σχεδιασμού σε τέσσερα επίπεδα, καθώς και τις συνδέσεις μεταξύ των επιπέδων με πέντε εικονικά δίκτυα TSV. Αντίστοιχα, το Σχήμα 2.10(b) αντιστοιχεί στον ίδιο σχεδιασμό μετά την ολοκλήρωση της λεπτομερούς δρομολόγησης (detailed routing) με το εργαλείο *Cadence nanoprouter*. Όπως έχει ήδη αναφερθεί, τα ηλεκτρικά χαρακτηριστικά αυτών των δικτύων επεξεργάζονται από νέα εργαλεία κατά το στάδιο της τοποθέτησης, προκειμένου να ανταποκριθούν κατάλληλα στα πραγματικά ηλεκτρικά χαρακτηριστικά της τεχνολογίας TSV. Συνεπώς παρόλο που με την προτεινόμενη προσέγγιση τα μονοπάτια δρομολόγησης διαφορετικών TSV δικτύων του Σχήματος 2.10(b) έχουν διαφορετικό μήκος καλωδίου στο επίπεδο σχεδίασης, εντούτοις, τροποποιούνται κατάλληλα προκειμένου να εμφανίζουν όλα τις ίδιες

παραμέτρους RLC , ανεξαρτήτως της φυσικής τους θέσης και χωροθέτησης εντός του σχεδιασμού.



Σχήμα 2.10 Παράδειγμα αξιολόγησης ενός 3-D σχεδιασμού με τέσσερα επίπεδα πυριτίου: (α) σχεδιασμός με τέσσερα εικονικά επίπεδα πυριτίου και δίκτυα TSV και (β) σχεδιασμός μετά από λεπτομερή δρομολόγηση.

Μετά από το στάδιο της προτυποποίηση 3-D, ο σχεδιασμός αξιολογείται ως προς την λειτουργικότητά του με τη διαδικασία της ανάλυσης χρονισμού (timing analysis). Για το σκοπό αυτό χρησιμοποιείται το εργαλείο *Cadence Static Timing Analysis Engine*, ενώ για λόγους πληρότητας η ανάλυση γίνεται τόσο εκ των προτέρων, καθώς και μετά τη σύνθεση του δικτύου διανομής ρολογιού (clock tree synthesis) και τη δρομολόγηση του σχεδιασμού. Σε περίπτωση που ο σχεδιασμός 3-D δεν πληροί τους περιορισμούς/προδιαγραφές του συστήματος, μπορούν να εφαρμοσθούν σχεδιαστικές βελτιστοποιήσεις κατά τον ίδιο τρόπο με τη συμβατική 2-D σχεδίαση.

2.4. Πειραματικά Αποτελέσματα

Σε αυτήν την ενότητα παρέχεται μία σειρά πειραματικών αποτελεσμάτων, τα οποία αποδεικνύουν την αποτελεσματικότητα του προτεινόμενου πλαισίου Plug&Chip. Για το σκοπό της επίδειξης, θα παρατεθεί αυτή η ανάλυση σε δύο συμπληρωματικά στάδια, παρόμοια με αυτά που παρουσιάζονται στην μεθοδολογία μας. Ειδικότερα, αρχικά ποσοτικοποιείται η αποτελεσματικότητα που προέρχεται από το δίαυλο επικοινωνίας μεταξύ του συστήματος υποδοχής και της πλατφόρμας υλικού. Στη συνέχεια παρέχονται αποτελέσματα σχετικά με την εφαρμογή του προτεινόμενου πλαισίου για τον φυσικό σχεδιασμό του ευρέως χρησιμοποιούμενου επεξεργαστή OpenRISC [21] με τη χρήση τρισδιάστατης τεχνολογίας ολοκλήρωσης 3-D. Επιπλέον, προκειμένου να αποδειχθεί η χρησιμότητα του προτεινόμενου πλαισίου στη σχεδίαση σύνθετων πραγματικών συστη-

μάτων, παρουσιάζονται ακολούθως, στην υποενότητα XXXXX τα πειραματικά αποτελέσματα από την εφαρμογή της μεθοδολογίας Plug&Chip επί του έργου του Ευρωπαϊκού Οργανισμού Διαστήματος (ESA), το οποίο στοχεύει στην ανάπτυξη υλικού αλγορίθμων μηχανικής όρασης για την υποστήριξη αυτόνομης ρομποτικής πλοήγησης.

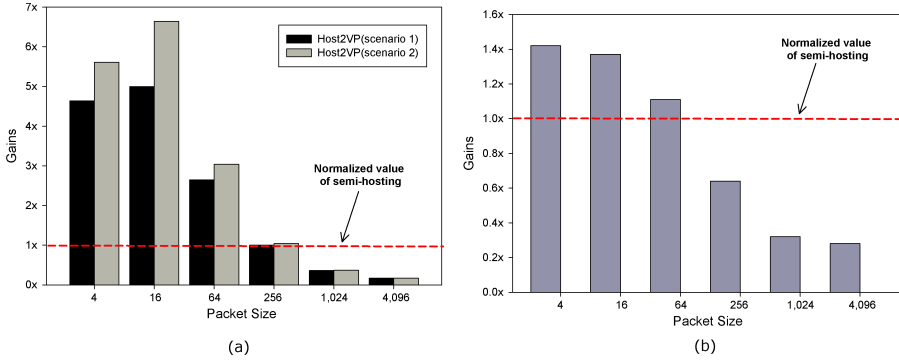
2.4.1. Αξιολόγηση του Διαύλου Επικοινωνίας

Η παρούσα υποενότητα επικεντρώνεται στην ποσοτικοποίηση της απόδοσης του προτεινόμενου διαύλου επικοινωνίας μεταξύ του συστήματος υποδοχής και της εικονικής πλατφόρμας VP, καθώς και μεταξύ της εικονικής πλατφόρμας VP και της πλατφόρμας υλικού. Δεδομένου ότι ο κύριος στόχος κατά την ανάπτυξη του προτεινόμενου πλαισίου είναι η επίτευξη της κατά το δυνατόν μικρότερης επιβάρυνσης στην επικοινωνία, διατηρώντας ταυτόχρονα, μεταξύ άλλων, την υποστήριξη πολλαπλών περιβαλλόντων λογισμικού εικονικής προτυποποίησης και τη δυνατότητα επέκτασης, η εν λόγω αξιολόγηση είναι ιδιαίτερα αντιπροσωπευτική για την αποδοχή μίας καθολικής λύσης στο πεδίο της συστημικής εικονικοποίησης. Επιπλέον, με δεδομένο ότι πρώτιστο ενδιαφέρον έχει η ποσοτικοποίηση του συστήματος επικοινωνίας, αγνοώντας τους πιθανούς περιορισμούς που θέτει το υπολογιστικό μέρος του συστήματος προορισμού, οι υποθέσεις εργασίας, που παρουσιάζονται στο υπόλοιπο του παρόντος, έχουν επιλεγεί κατάλληλα, ώστε να επιτευχθεί η κατά το δυνατόν μικρότερη επιβάρυνση σε απόλυτο χρόνο ρολογιού.

Για να ποσοτικοποιηθεί η απόδοση των βιβλιοθηκών Host2VP και VP2HW, έχουν αναπτυχθεί δύο προγράμματα σε γλώσσα C, τα οποία με επαναληπτικό τρόπο στέλνουν και λαμβάνουν πακέτα δεδομένων πλήθους 10^6 ακεραίων, συγκεκριμένου μεγέθους (που κυμαίνονται από 4 έως 4.096 bytes) μεταξύ του υπολογιστή υποδοχής και της εικονικής πλατφόρμας VP, καθώς και της εικονικής πλατφόρμας VP και του εξειδικευμένου υλικού, αντίστοιχα. Η επιβάρυνση της προτεινόμενης βιβλιοθήκης στην επικοινωνία αποτιμάται μέσω της αποστολής ενός πακέτου 10^6 ακεραίων. Τα αποτελέσματα της εν λόγω ανάλυσης απεικονίζονται στο Σχήμα 2.11. Ο οριζόντιος άξονας στο σχήμα αυτό αντιστοιχεί στο μέγεθος του πακέτου δεδομένων που ανταλλάσσεται μεταξύ του υπολογιστή υποδοχής και της VP (Σχήμα 2.11(a)), όπως επίσης μεταξύ της VP και του υλικού (Σχήμα 2.11(b)), ενώ ο κατακόρυφος άξονας απεικονίζει τα κέρδη επιτάχυνσης, σε απόλυτο χρόνο ρολογιού, σε σύγκριση με την προσέγγιση semi-hosting. Σε αυτό το σχήμα, απεικονίζεται περαιτέρω με διακεκομμένη γραμμή η λύση που αντιστοιχεί στην προσέγγιση της τεχνικής semi-hosting.

Σχετικά με τη μεταφορά δεδομένων, χαρακτηρίζονται δύο διαφορετικά σενάρια, τα οποία απεικονίζονται στο Σχήμα 2.11(a), και συνοψίζονται ως εξής:

- **Σενάριο 1:** Η μεταφορά δεδομένων γίνεται σε πακέτα, των οποίων το μέγεθος είναι ίσο προς which το μέγεθος πακέτου του μηχανισμού επικοινωνίας Host2VP, όπως αυτό έχει οριστεί κατά το χρόνο σχεδίασης.
- **Σενάριο 2:** Η μεταφορά δεδομένων γίνεται σε πακέτα μεγέθους 4,096 bytes, δηλαδή 1,024 ακεραίων σε ένα συμβατικό σύστημα αρχιτεκτονικής



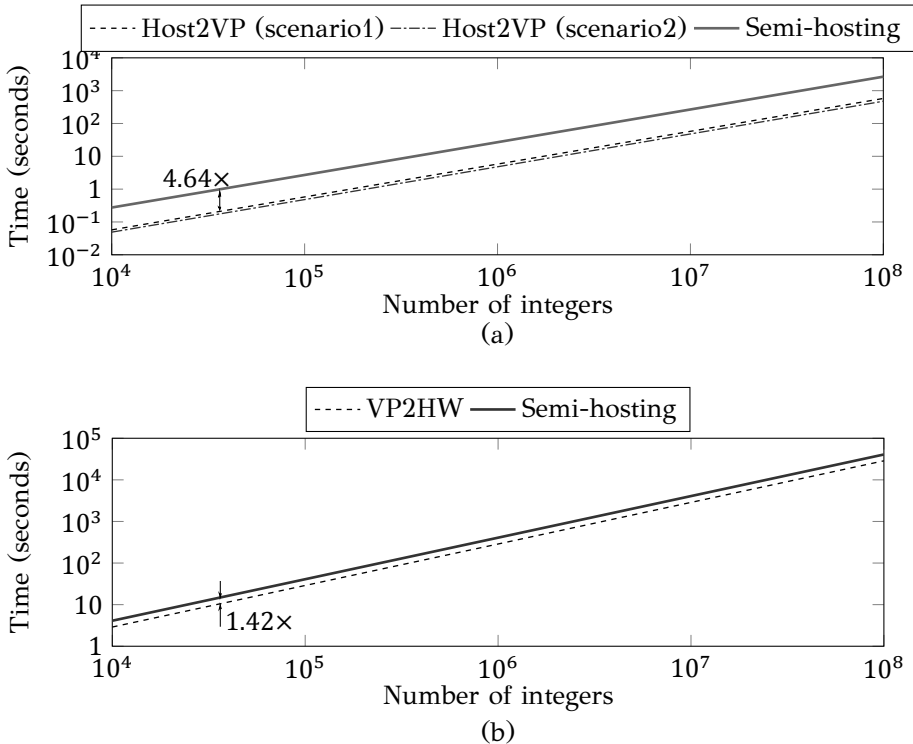
Σχήμα 2.11 Κέρδη επιτάχυνσης σε απόλυτο χρόνο ρολογιού για: (α) την προτεινόμενη βιβλιοθήκη Host2VP και (β) την προτεινόμενη βιβλιοθήκη VP2HW, συγκριτικά με την προσέγγιση semi-hosting[4].

x86).

Με βάση τα αποτελέσματα των Σχημάτων 2.11(α),(β), συμπεραίνεται ότι για μικρού και μεσαίου μεγέθους πακέτα δεδομένων, η προτεινόμενη λύση είναι έως 5.61x ταχύτερη σε σχέση με την προσέγγιση της τεχνική semi-hosting Host2VP και έως 1.42x ταχύτερη στην περίπτωση της VP2HW προσέγγισης. Ειδικότερα, σχετικά με την προσέγγιση Host2VP, η προτεινόμενη βιβλιοθήκη αποδίδει υψηλότερη απόδοση για πακέτα μεγέθη μικρότερα (ή ίση) με 256 bytes, ενώ η το αντίστοιχο μέγεθος πακέτου για τη προσέγγιση VP2HW είναι 64 bytes. Ο μικρή απόδοση της προσέγγισης semi-hosting για μικρού και μεσαίου μεγέθους πακέτα δεδομένων, μπορεί να εξηγηθεί λαμβάνοντας υπόψη ότι η μεταφορά δεδομένων με την προσέγγιση αυτή πραγματοποιείται στα συστήματα μαζικής αποθήκευσης, π.χ. στο σκληρό δίσκο, όπου κάθε μεταφορά δεδομένων γίνεται σε πακέτα των 4-Kbyte, όπως έχει ήδη αναφερθεί στους Πίνακες 2.1 και 2.2. Ως εκ τούτου, αναγκάζοντας τα τμήματα του σκληρού δίσκου να προσπελαστούν περισσότερες φορές, η τεχνική semi-hosting οδηγεί σε αξιοσημείωτες επιβαρύνσεις καθυστέρησης. Αντίθετα, σχετικά με μεγαλύτερου μεγέθους πακέτα δεδομένων, η τεχνική semi-hosting σημειώνει καλύτερα αποτελέσματα από την προτεινόμενη

λύση, δεδομένου ότι εκτελεί λιγότερες προσπελάσεις στο σκληρό δίσκο. Παρά την ανωτέρω ανάλυση, ένα πακέτο μεγέθους 4 byte ανταποκρίνεται σε πιο ρεαλιστικό σενάριο στα πλαίσια της παρούσας εργασίας, καθώς το συγκεκριμένο OVP μοντέλο που χρησιμοποιείται (επεξεργαστής OpenRISC), υποστηρίζει μεταφορές μόνο μικρού μήκους, παρά το γεγονός ότι η τεχνική TLM μπορεί να υποστηρίξει, εν δυνάμει, τη μεταφορά οποιουδήποτε μήκους.

Στο Σχήμα 2.12 απεικονίζονται οι σχετικές επιβαρύνσεις επικοινωνίας για τα δύο σενάρια που μελετούνται στην παρούσα ενότητα, όταν το μέγεθος του πακέτου ορίζεται σταθερά ίσο προς 4 bytes. Ο οριζόντιος άξονας στο σχήμα αυτό, αντιστοιχεί στον αριθμό των ακεραίων που αποστέλλονται μέσω του διαύλου επικοινωνίας, ενώ οι κάθετοι άξονες απεικονίζουν, σε λογαριθμική κλίμακα, το χρόνο που απαιτείται για αυτή τη μεταφορά δεδομένων. Με αυτό το σχήμα εμφανίζεται η αναλογική αύξηση του χρόνου σε απόλυτο χρόνο ρολογιού, με τον αριθμό των ακεραίων που μεταφέρονται μέσω του διαύλου επικοινωνίας. Συνεπώς είναι δυνατή η εκτίμηση της χρονικής επιβάρυνσης για μεταφορά δεδομένων οποιουδήποτε πλήθους. Συμπεραίνεται λοιπόν ότι τα κέρδη σε απόλυτο χρόνο ρολογιού είναι ανεξάρτητα προς το συνολικό όγκο δεδομένων.



Σχήμα 2.12 Αξιολόγηση της επιβάρυνσης επικοινωνίας μεταξύ: (α) συστήματος υποδοχής και εικονικής πλατφόρμας VP, (β) εικονικής πλατφόρμας VP και υλικού, για μέγεθος πακέτου 4 bytes.

Με βάση το Σχήμα 2.12, ακολούθως, εξάγονται συμπεράσματα σχετικά προς

την αποδοτικότητα του προτεινόμενου διαύλου επικοινωνίας, σε σύγκριση με την τεχνική semi-hosting, όπως αναλύεται στο Σχήμα 2.11, για την περίπτωση του πακέτου μεγέθους 4-byte. Η παραμετροποίηση του διαμεσολαβητή (transactor) του διαύλου επικοινωνίας HotTalk επηρεάζει σημαντικά την απόδοση της συν-προσομοίωσης, καθώς ο δίαυλος επικοινωνίας μπορεί να έχει μεγαλύτερη καθυστέρηση εξυπηρέτησης σε σχέση με τις ανάγκες επικοινωνίας των εξειδικευμένων πυρήνων IP που υλοποιούνται στο FPGA. Ως εκ τούτου, προκειμένου να μεγιστοποιηθεί η αποτελεσματικότητα της συν-προσομοίωσης, οι σχεδιαστές πρέπει να χρησιμοποιήσουν ένα κατάλληλο διαμεσολαβητή αναφορικά προς την επιλογή του μεγέθους της ανατροφοδοτούμενης μνήμης, δηλαδή της ποσότητας δεδομένων με την οποία το σύστημα εποπτείας (arbiter) τροφοδοτεί τους πυρήνες υλικού IP σε κάθε συναλλαγή. Σε αυτή την επιλογή πρέπει επίσης να ληφθεί υπόψη το γεγονός ότι το μεγαλύτερο μέγεθος της μνήμης ανατροφοδοσίας επιβάλλει χρήση πρόσθετων πόρων υλικού κατά την υλοποίηση του διαμεσολαβητή, λόγω των αυξημένων μονάδων μνήμης του FPGA, καθώς και των ευρύτερων διαύλων διευθύνσεως για τις εν λόγω μνήμες.

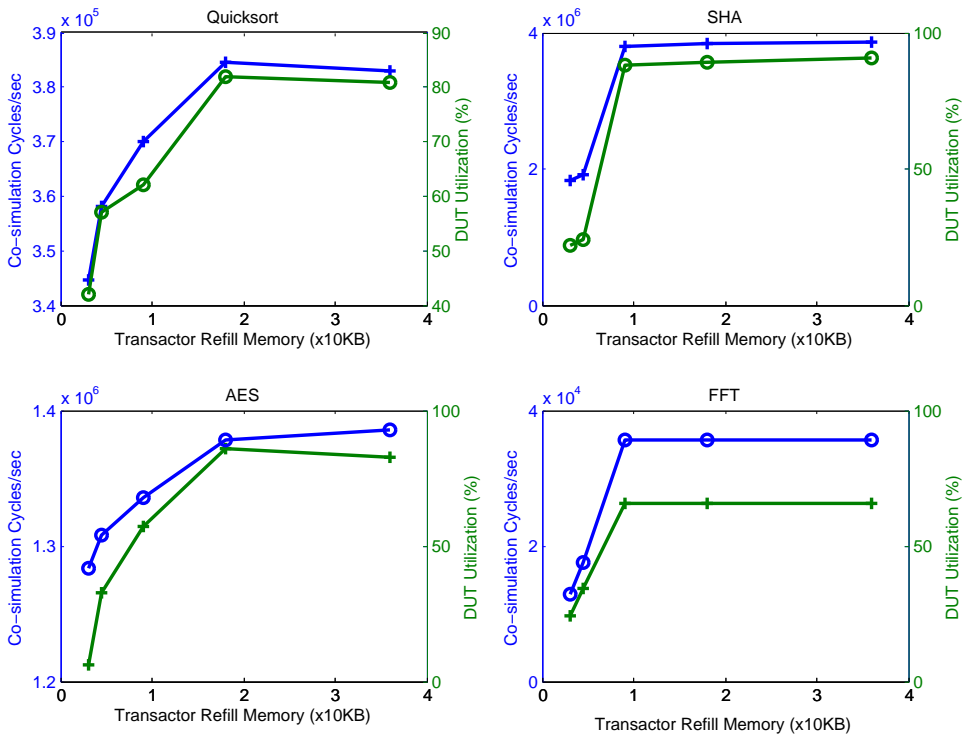
Αντιθέτως, δεν υφίσταται τέτοιου είδους μονότονη σχέση μεταξύ του χρόνου εκτέλεσης της συν-προσομοίωσης και του μεγέθους της ανατροφοδοτούμενης μνήμης του διαμεσολαβητή. Αναφορικά με τη λύση που προτείνει το πλαίσιο Plug&Chip, ο ρυθμός δεδομένων του διαύλου επικοινωνίας HotTalk API προσαρμόζεται στην κατάλληλη τιμή, μέσω της ρύθμισης της ανατροφοδοτούμενης μνήμης, ανάλογα με τις απαιτήσεις της εκάστοτε εφαρμογής. Για παράδειγμα, δεδομένης της υπόθεσης ότι μία εφαρμογή πρέπει να επεξεργαστεί ένα σύνολο n -bytes και ο διαμεσολαβητής είναι ρυθμισμένος έτσι ώστε να τροφοδοτεί το FPGA με $m < n$ bytes, τότε στην περίπτωση αυτή, ο αριθμός των επαναλήψεων που πρέπει να εκτελεστούν ισούται με $\lceil \frac{n}{m} \rceil$ και με τον τρόπο αυτό, υποβαθμίζεται η ταχύτητα της συν-προσομοίωσης, εφόσον το σύστημα επαλήθευσης (Design Under Test, DUT) θα βρίσκεται σε κατάσταση στασιμότητας του διαύλου επικοινωνίας για παράδοση των δεδομένων. Η αξιοποίηση του DUT (DUT utilization) αναφέρεται στην αναλογία του χρόνου υπό τον οποίο το σύστημα επαλήθευσης δεν παραμένει στάσιμο, λόγω της μεταφοράς δεδομένων, επί του συνολικού χρόνου προσομοίωσης.

Η απόδοση του διαμεσολαβητή της βιβλιοθήκης HotTalk ποσοτικοποιήθηκε με τέσσερα προγράμματα από τη βιβλιοθήκη εφαρμογών MiBench [22]. Η επιλογή των προγραμμάτων έγινε δεδομένου ότι η ροή εκτέλεσης για καθένα από αυτά επιβάλλει διαφορετικούς ρυθμούς δεδομένων εισόδου/εξόδου. Αναφορικά με την ανάλυση αυτή, ο επεξεργαστής OpenRISC υλοποιήθηκε στην συσκευή Virtex-6 FPGA (HTG-V6-PCIE-L240T-2), ενώ το σύστημα εποπτείας συνδέθηκε άμεσα με τη μνήμη του επεξεργαστή, έτσι ώστε η εισαγωγή δεδομένων στα προγράμματα να γίνεται απευθείας στη μνήμη του επεξεργαστή.

Τα αποτελέσματα αυτής της ανάλυσης, τα οποία απεικονίζουν τον τρόπο με τον οποίο η παραμετροποίηση του διαμεσολαβητή επηρεάζει τη συνολική ταχύτητα του συν-προσομοίωσης, συνοψίζονται στο Σχήμα 2.13. Οι κάθετοι άξονες σε αυτό το σχήμα αντιπροσωπεύουν την απόδοση της συν-προσομοίωσης σε μετρική αριθμού κύκλων ανά δευτερόλεπτο (αριστερός άξονας) και το ποσοστό

χρησιμοποίησης του DUT (δεξιός άξονας), ενώ ο οριζόντιος άξονας αντιστοιχεί στο μέγεθος της μνήμης του διαμεσολαβητή. Η απόδοση των συν-προσομοίωσης μπορεί να εκτιμηθεί ως το σύνολο των κύκλων εκτέλεσης του διαμεσολαβητή, συμπεριλαμβανομένου του χρόνου επικοινωνίας και του χρόνου εκτέλεσης του συστήματος επαλήθευσης (DUT). Επίσης, το ποσοστό χρησιμοποίησης του DUT καθορίζει το ποσοστό του χρόνου που το DUT δεν βρίσκεται σε κατάσταση στασιμότητας λόγω μεταφοράς δεδομένων

Με βάση το Σχήμα 2.13, συμπεραίνεται ότι η απόδοση του διαμεσολαβητή της βιβλιοθήκης HotTalk εξαρτάται σε μεγάλο βαθμό από τις εγγενείς απαιτήσεις της εκάστοτε εφαρμογής για είσοδο/έξοδο δεδομένων. Πιο αναλυτικά, καθώς αυξάνεται η μνήμη ανατροφοδότησης, παρατηρείται μια αναλογική αύξηση στην αποδοτικότητα του διαμεσολαβητή, μέχρι ένα σημείο κορεσμού. Για παράδειγμα, όσον αφορά τα προγράμματα που μελετήθηκαν σε αυτή την εργασία, αυτό το σημείο κορεσμού είναι σχεδόν 10KB για τα προγράμματα *SHA* και *FFT* και 20KB για τα προγράμματα *Quicksort* και *AES* αντίστοιχα. Πέραν αυτών των σημείων κορεσμού, δεν μπορεί να επιτευχθεί περαιτέρω κέρδος, καθώς ακόμη και αν τροφοδοτούνται δεδομένα με υψηλότερο ρυθμό εισαγωγής, τα δεδομένα αυτά δεν μπορούν να χρησιμοποιηθούν από τους εξειδικευμένους πυρήνες υλικού μέσω του συστήματος επαλήθευσης (DUT).



Σχήμα 2.13 Αξιολόγηση της απόδοσης της προτεινόμενης συν-προσομοίωσης για επιλεγμένους αλγόριθμους της βιβλιοθήκης εφαρμογών MiBench.

2.4.2. Αξιολόγηση του επεξεργαστή OpenRISC σε τρισδιάστατη τεχνολογία ολοκλήρωσης 3-D

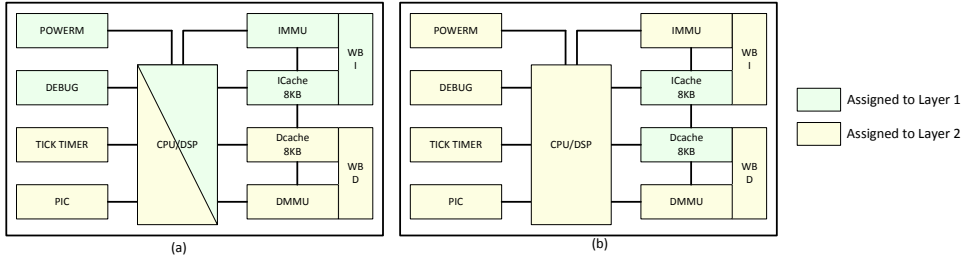
Το δεύτερο μέρος αυτής της ενότητας παρέχει μια σειρά από ποιοτικά αποτελέσματα σχετικά με την αξιολόγηση των ψηφιακών συστημάτων που υλοποιούνται σε φυσικό επίπεδο με τεχνολογία ολοκλήρωσης 3-D. Για την ανάλυση αυτή χρησιμοποιείται ο ανοικτού-κώδικα επεξεργαστής OpenRISC [21], ενώ ως τεχνολογία ολοκλήρωσης χρησιμοποιείται η TSMC 45nm με 9 επίπεδα μετάλλου, ενώ οι παράμετροι για τις κάθετες διασυνδέσεις μεταξύ των επιπέδων πυριτίου (Through Silicon Via, TSV) συνοψίζονται στον Πίνακα 2.3.

Πίνακας 2.3 Χαρακτηριστικά της χρησιμοποιούμενης τεχνολογίας TSV [23].

Διάμετρος	Ελάχιστο Pitch	Αντίσταση	Χωρητικότητα	Μήκος
1.2um	4um	0.35Ω	2.5fF	4-9um

Το Σχήμα 2.14 απεικονίζει δύο διαφορετικές προσεγγίσεις κατάτμησης του επεξεργαστή OpenRisc, αναφορικά προς το στάδιο σχεδιασμού της στοίβας 3-D. Διαφορετικά χρώματα στα διαγράμματα αυτά υποδηλώνουν μπλοκ που έχουν ανατεθεί σε διαφορετικά επίπεδα πυριτίου. Πιο συγκεκριμένα, το πρώτο από αυτά, το οποίο απεικονίζεται στο Σχήμα 2.14(α) αντιστοιχεί σε έναν διαχωρισμό με αλγόριθμο min-cut, ενώ το δεύτερο, το οποίο απεικονίζεται στο Σχήμα 2.14(β) αντιστοιχεί σε ένα διαχωρισμό προσανατολισμένο στη τεχνολογία υλοποίησης. Παρόλο που είναι δυνατή η υλοποίησης επιπλέον κατατμήσεων, για παράδειγμα με τη χρήση του αλγορίθμου *Tabu*, οι λύσεις που απεικονίζεται στο προαναφερθέν σχήμα αντιστοιχούν σε δύο αντιπροσωπευτικές λύσεις. Συγκεκριμένα, η προσέγγιση min-cut οδηγεί στην ελάχιστη χρησιμοποίηση συνδέσεων TSVs, και ως εκ τούτου αναφέρεται στη περίπτωση της μέγιστης αξιοποίησης της επιφάνειας πυριτίου αναφορικά προς την απόδοση τπυ παραγωγικού σταδίου της λιθογραφίας (yield improvement). Κινούμενη σε άλλη κατεύθυνση, η δεύτερη προσέγγιση είναι εφαρμόσιμη σε σχεδιασμούς που περιλαμβάνουν ετερογενείς τεχνολογίες φυσικής υλοποίησης αναφορικά προς τις απαιτήσεις και τους περιορισμούς σχεδίασης, όπως για παράδειγμα οι More-than-Moore τεχνολογίες, οι οποίες υπόσχονται μεγάλη ολοκλήρωση σχεδίασης (π.χ. ψηφιακή σχεδίαση, αναλογική σχεδίαση RF, μπαταρία, αισθητήρες, κ.λπ.).

Η σύνθεση του επεξεργαστή OpenRisc πραγματοποιήθηκε με το εργαλείο *Cadence Encounter RTL Compiler*, θέτοντας ως στόχο περιόδου ρολογιού τα 10ns (συχνότητα 100MHz). Το μέγεθος του συνθέσιμου σχεδιασμού αποτελείται από 13,847 λογικά κελιά (standard cells), 14541 καλώδια (nets), και 374 θύρες E/E. Ο Πίνακας 2.4 δίνει πρόσθετες τεχνικές λεπτομέρειες σχετικά με την φυσική υλοποίηση του επεξεργαστή OpenRisc. Με βάση αυτά τα αποτελέσματα συμπεραίνετε ότι η μετάβαση σε τεχνολογία ολοκλήρωσης 3-D μπορεί να βελτιώσει τη μέγιστη συχνότητα λειτουργίας 1,19x και 1,26x, αντίστοιχα, για τις δύο προσεγγίσεις, σε σύγκριση με τη συμβατική υλοποίηση. Οι σχεδιασμοί σε φυσικό επίπεδο μάσκας (layouts) που ανακτώνται από το εργαλείο *Cadence SoC Encounter* για τις δύο αυτές αρχιτεκτονικές προσεγγίσεις απεικονίζονται στο Σχήμα 2.15.



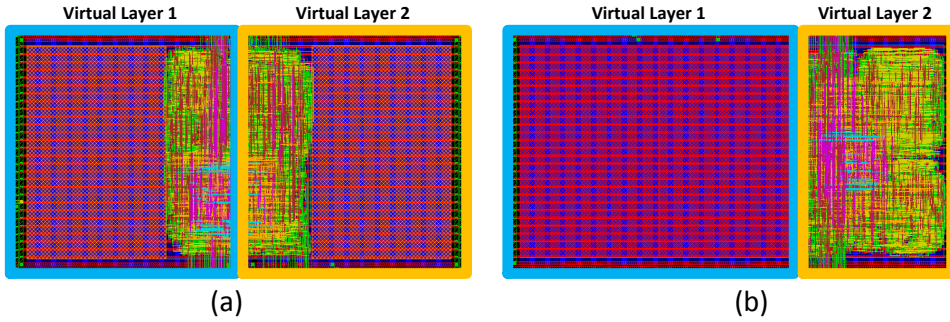
Σχήμα 2.14 Κατάτμηση του επεξεργαστή OpenRISC για εναλλακτικές απαιτήσεις σχεδίασης: (i) κατάτμηση με αλγόριθμο min-cut and (b) κατάτμηση προσανατολιμένη στη τεχνολογίας υλοποίησης.

Πίνακας 2.4 Αποτίμηση υλοποίησης σε φυσικό επίπεδο μάσκας πυριτίου του επεξεργαστή OpenRISC, με τεχνολογία 3-D.

Χαρακτηριστικά	2-D Σύστημα	3-D Σύστημα	
		Λύση Σχ. 2.14(α)	Λύση Σχ. 2.14(β)
Μέγιστη Συχνότητα (MHz)	95	113	120
Μήκος Καλωδίου (μm)	318,750	330,986	323,247
Ημιπερίμετρος (μm)	237,791	247,387	234,411
Αριθμός TSVs	0	272	408
Επιφάνεια για TSVs (μm^2)	0.00	1395.36	2093.04
Αναλογία διαστάσεων	1.00	2.00	2.00
Επιφάνεια κάθε επιπέδου	149,344 μm^2	149,368 μm^2	149,369 μm^2

2.4.3. Εφαρμογή του πλαισίου Plug&Chip σε ένα ετερογενές ενσωματωμένο σύστημα - Το έργο SPARTAN

Αυτή η υποενότητα περιγράφει τον τρόπο με τον οποίο εφαρμόστηκε το προτεινόμενο πλαίσιο σε ένα έργο του Ευρωπαϊκού Οργανισμού Διαστήματος (ESA), το έργο SPARTAN [24], το οποίο εστιάζει στην υλοποίηση αλγορίθμων όρασης υπολογιστών σε επαναδιαμορφούμενες πλατφόρμες. Σε αντίθεση με παρόμοιες προσεγγίσεις, που αντιμετωπίζουν αυτό το πρόβλημα μόνο σε επίπεδο λογισμικού, στόχος του συγκεκριμένου έργου είναι η ανάπτυξη πυρήνων όρασης υπολογιστών (Computer Vision, CV) σε γλώσσα VHDL, συμβατή με τις προδιαγραφές της ESA [25], ώστε να καταστεί δυνατή η επεξεργασία εικόνων πολύ μεγάλης ανάλυσης, σε πραγματικό χρόνο. Το έργο βασίζεται στην αρχιτεκτονική ενός ετερογενούς συστήματος, αποτελούμενο από έναν ενσωματωμένο επεξεργαστή και ένα FPGA. Εφόσον η διαθέσιμη διαστημική τεχνολογία είναι περιορισμένης επεξεργαστικής ισχύος σε σχέση με τη γήινη, ο ενσωματωμένος επεξεργαστής είναι μικρής απόδοσης και αντίστοιχα η συσκευή FPGA διαθέτει λιγότερους λογικούς πόρους σε σχέση με τον αριθμό πόρων των σύγχρονων FPGAs. Συγκεκριμένα, η ESA σκιαγραφεί, βάση εκτιμήσεων, τα χαρακτηριστικά ενός τέτοιου ετερογενούς συστήματος με τεχνολογία αντίστοιχη της διαθέσιμης διαστημικής τεχνολογίας για το έτος 2018, οπότε και αναμένεται η πραγματοποίηση του προγράμματος Exomars, στο οποίο ανήκει και το έργο SPARTAN, ως 150 MIPS για τον ενσωματωμένο επεξεργαστή και μια συσκευή Virtex-6 για την FPGA. Το έργο αναπτύσσεται βάση μιας μεθοδολογίας συν-σχεδιασμού έτσι ώστε ο

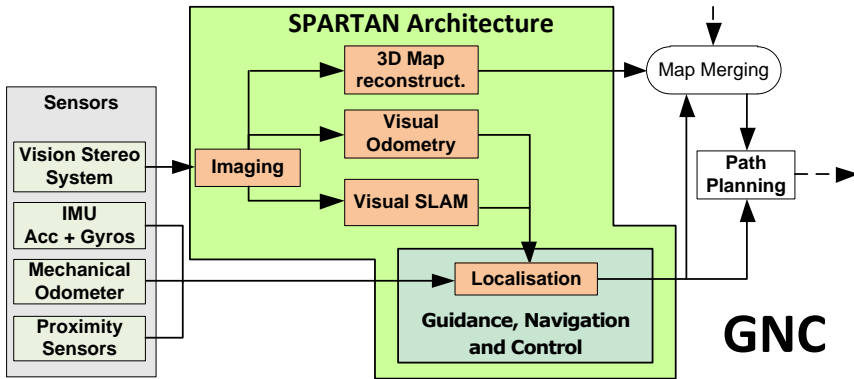


Σχήμα 2.15 Φυσική υλοποίηση του επεξεργαστή OpenRISC σε τεχνολογία 3-D TSMC 45nm, με τη χρήση του εργαλείου Cadence SoC Encounter, για τη κατάτμηση των Σχημάτων 2.14(α) και 2.14(β).

επεξεργαστής να παρέχει όλες τις βασικές υπηρεσίες, σε επίπεδο λειτουργικού συστήματος, οι οποίες απαιτούνται για τη σωστή πλοήγηση του ρομποτικού οχήματος (rover), όπως η παροχή μηχανισμών αφαιρετικότητας του υλικού, ο έλεγχος των συσκευών σε χαμηλό επίπεδο, η ανταλλαγή μηνυμάτων μεταξύ των διεργασιών, και η διαχείριση των πακέτων λογισμικού, ενώ η συσκευή FPGA ενεργεί ως επιταχυντής για τους υπολογιστικά απαιτητικούς αλγορίθμους μηχανικής όρασης. Πρόσθετες πληροφορίες για την αρχιτεκτονική του συνολικού συστήματος SPARTAN αποτυπώνονται στη δημοσίευση [26].

Η επιτυχής ολοκλήρωση του έργου βασίζεται στην επίτευξη συγκεκριμένων προδιαγραφών, όπου πρωτεύον στόχος είναι η διασφάλιση υψηλής απόδοσης με ταυτόχρονη όμως διατήρηση της συστημικής αποτελεσματικότητας, αναφορικά με τη κατανάλωση ενέργειας τη δέσμευση μνήμης κτλ. Ξεκινώντας από μία ολοκληρωμένη λύση λογισμικού σε γλώσσα C/C++, σε πρώτη φάση χαρακτηρίζονται οι αλγόριθμοι βάση των επεξεργαστικών τους απαιτήσεων ώστε να αναγνωριστούν τα τμήματα κώδικα που χρίζουν επιτάχυνσης στην επαναδιαμορφούμενη πλατφόρμα. Καθώς η συνολική απόδοση του συστήματος SPARTAN, η αρχιτεκτονική του οποίου απεικονίζεται στο Σχήμα 2.16, εξαρτάται σε μεγάλο βαθμό από τα δεδομένα εισόδου (π.χ. πληροφοριακό περιεχόμενο της εικόνας), το σύστημα παρουσιάζει αυξημένο δυναμισμό και συνεπώς το κόστος εκτέλεσής του (αριθμός κύκλων επεξεργαστή, αποτύπωμα μνήμης, ενέργεια κ.α.) και η ποιότητα των μετρήσεων (π.χ. ακρίβεια αλγορίθμου, μέσο σφάλμα συστήματος κ.α.) εξαρτώνται από το περιβάλλον και τις συνθήκες εκτέλεσης του αλγορίθμου. Ως εκ τούτου, κατά τη διάρκεια ανάπτυξης του έργου, υφίστατο η πρόκληση της εκτίμησης της δυναμικής συμπεριφοράς κατά το χρόνο σχεδίασης.

Προκειμένου να εξασφαλιστεί η διαχείριση της δυναμικής φύσης του συστήματος SPARTAN, υιοθετήθηκε η έννοια των σεναρίων συστήματος, η οποία παρουσιάζεται στην εργασία [27]. Σύμφωνα με την προσέγγιση αυτή, οι μετρικές απόδοσης του συστήματος που εμφανίζουν παρόμοιο σχεδιαστικό κόστος αναφορικά με μία πολυδιάστατη προσέγγιση βελτιστοποίησης, όπως για παράδειγμα η ελαχιστοποίηση των πόρων υλικού, η καθυστέρηση, και η κατανάλωση ενέργειας, ομαδοποιούνται κατά τέτοιο τρόπο ώστε το σύστημα να ρυθμιστεί



Σχήμα 2.16 Αρχιτεκτονική σε επίπεδο συστήματος του έργου SPARTAN.

προκειμένου να αξιοποιείται αυτή η ομοιότητα του κόστους. Η έννοια της χρήσης σεναρίων σε ενσωματωμένα συστήματα ξεκινά από τις διαφορετικές καταστάσεις χρόνου εκτέλεση (Run-Time Situations, RTSs), στις οποίες ένα σύστημα μπορεί να βρεθεί, όταν χρησιμοποιεί μια συγκεκριμένη πλατφόρμα εκτέλεσης. Αναφορικά προς το έργο SPARTAN, οι παράμετροι των αλγορίθμων μηχανικής όρασης, καθώς και εκείνες που επηρεάζουν την αρχιτεκτονική συν-σχεδίασης, οδηγούν σε διαφορετικές RTSs, οι οποίες επηρεάζουν ιδιαίτερα το κόστος μιας μετρικής και συνεπώς τη ποιότητα της σχεδίασης (π.χ. πόροι FPGA, συχνότητα της CPU, κλπ). Σε γενικές γραμμές, οποιοσδήποτε συνδυασμός του χώρου N-μετρικών μπορεί να επιλυθεί με αυτή την προσέγγιση. Ωστόσο, ο αριθμός των μετρικών σε συνδυασμό με όλες τις πιθανές τιμές των παραμέτρων των RTS συνήθως οδηγεί σε εκθετική αριθμό RTSs. Ως εκ τούτου, η διαδικασία σχεδιασμού περιπλέκεται ενώ το κόστος επαλήθευσης και χαρακτηρισμού και μελέτης όλων αυτών των RTSs είναι υπερβολική ψηλή αναφορικά με το χρόνο σχεδίασης. Μια προσέγγιση στο πρόβλημα αυτό βασίζεται σε κατάλληλη ομαδοποίηση αυτών των παραμέτρων από τις N-διαστάσεις μετρικών σε σεναρία συστήματος [27]. Παρόλο που υπάρχουν υφιστάμενοι μηχανισμοί για την πρόβλεψη των βέλτιστων σεναρίων, κατά το χρόνο εκτέλεσης, καθώς και για τη μεταγωγή σεναρίου [27], εφαρμόζονται κυρίως σε ομογενείς πλατφόρμες. Σε αντίθεση, το SPARTAN έργο αναφέρεται στο συν-σχεδιασμό αρχιτεκτονικής CPU-FPGA. Συνεπώς, η ιδέα της χρησιμοποίησης σεναρίων πρέπει να προσαρμοστεί στην εν λόγω αρχιτεκτονική.

Μετά από μία διαδικασία λεπτομερούς χαρακτηρισμού του πηγαίου κώδικα (profiling), διαπιστώθηκε ότι οι πιο υπολογιστικά απαιτητικοί πυρήνες μηχανικής όρασης είναι ο αλγόριθμος “Disparity” και ο αλγόριθμος SURF [28]. Ο πρώτος χρησιμοποιείται στη διαδικασία τρισδιάστατης ανακατασκευής του χάρτη της περιοχής στην οποία κινείται το ρομποτικό όχημα (3D Map Reconstruction), ενώ ο δεύτερος χρησιμοποιείται στο στάδια της εικονικής οδομετρίας του συ-

στήματος. Αυτοί οι αλγόριθμοι επηρεάζουν ιδιαίτερα την απόδοση του τελικού συστήματος. Επίσης, δεδομένου ότι υπάρχει εξάρτηση μεταξύ αυτών των αλγορίθμων με τις διάφορες λειτουργίες του SPARTAN συστήματος, κατά τη διάρκεια της φάσης ανάπτυξης, είναι σημαντικό να πραγματοποιείται επαναληπτικός έλεγχος και διερεύνηση των επιπτώσεων των αλγοριθμικών παραμέτρων σε σχέση με τις προδιαγραφές του τελικού συστήματος.

Γενικά, για ένα έργο λογισμικού, η διαδικασία δοκιμής και επαλήθευσης ορθής λειτουργίας είναι συνήθως ελέγξιμη και προβλεπόμενη, λόγω της ενιαίας και ομοιογενής δομής του μέσου υλοποίησης, δηλαδή του λειτουργικού συστήματος, του προγραμματιστικού περιβάλλοντος ανάπτυξης (API) και των κοινόχρηστων βιβλιοθηκών λογισμικού. Επίσης σε ένα τέτοιο περιβάλλον είναι δυνατόν να χρησιμοποιηθούν κοινά εργαλεία αποσφαλμάτωσης και ελέγχου ποιότητας, καθολικά σε όλο το υπό-ανάπτυξη σύστημα. Αντίθετα, σε ένα ετερογενές σύστημα, όπως είναι η αρχιτεκτονική του έργου SPARTAN, δεν δύναται να χρησιμοποιηθούν κοινώς εφαρμόσιμες λύσεις ανάπτυξης και ελέγχου, καθώς τα διαφορετικά επίπεδα υλικού και λογισμικού συνοδεύονται από συγκεκριμένα εργαλεία ανάπτυξης και αποσφαλμάτωσης. Συνεπώς στο έργο SPARTAN χρησιμοποιήθηκε το πλαίσιο Plug&Chip εφόσον μέσω αυτού, είναι δυνατή η ποσοτικοποίηση της επίδρασης των RTS παραμέτρων στους αλγορίθμους μηχανικής όρασης, τόσο σε πρώιμα στάδια του σχεδιασμού, καθώς και σε όλη τη φάση της σχεδίασης του τελικού συστήματος.

Προς απόδειξη της ευελιξίας που παρέχει το προτεινόμενο πλαίσιο, το Σχήμα 2.17 απεικονίζει την αξιολόγηση του συστήματος SPARTAN χρησιμοποιώντας ανάλυση σχεδιαστικών συμβιβασμών, βασισμένη σε σενάρια λειτουργίας. Συγκεκριμένα, κατά τα πρώιμα στάδια ανάπτυξης του υλικού σε γλώσσα VHDL, είχαν επιτυχώς αναπτυχθεί δύο πυρήνες υλικού του συστήματος. Ωστόσο λόγω του ότι οι υπόλοιποι πυρήνες δεν είχαν αναπτυχθεί πλήρως, το σύστημα δεν μπορούσε να ελεγχθεί ως προς τη συνολική απόδοση. Ωστόσο, το πλαίσιο Plug&Chip βοήθησε στην αξιολόγηση 16 RTS παραμέτρων και επέτρεψε τη δοκιμή των δύο υλοποιημένων πυρήνων σε πραγματικό FPGA μέσω στιγμιότυπου ελέγχου (testbench), όπου το υπόλοιπο σύστημα προσομοιωνόταν με μία εικονική μηχανή ενώ η επικοινωνία της εικονικής μηχανής με τους πυρήνες υλικού διασφαλιζόταν από τις βιβλιοθήκες Host2VP και VP2HW. Με αυτόν τον τρόπο κατέστη δυνατή η εκτίμηση σφάλματος θέσης του ρομποτικού συστήματος κατά την πλοήγηση, χαρτογράφηση και εξερεύνηση μιας άγνωστης περιοχής, όπου η μόνη πληροφορία για την περιοχή παρέχεται από δύο κάμερες ανάλυσης 1120×1120 και 520×380 τοποθετημένες στο εμπρόσθιο τμήμα του ρομποτικού οχήματος, σε ένα μέτρο και σαράντα πόντους αντίστοιχα σε απόσταση από το έδαφος. Η πληροφορία των εικόνων εισέρχεται στους αλγορίθμους ρομποτικής όρασης, οι οποίοι έχουν ως έξοδο την εκτίμηση κίνησης του ρομποτικού οχήματος. Οι δύο αλγόριθμοι που αναπτύχθηκαν σε VHDL είναι υπεύθυνοι για την εξαγωγή των χαρακτηριστικών εικόνας, επί των οποίων λειτουργούν οι αλγόριθμοι εκτίμησης κίνησης και χαρτογράφησης.

Οι πυρήνες VHDL αναπτύχθηκαν παραμετρικά, ώστε να επιτρέπουν την όσο το δυνατό λεπτομερέστερη και ακριβείς ανάλυση των χαρακτηριστικών εικό-

Σχήμα 2.17 Αξιολόγηση του συστήματος SPARTAN χρησιμοποιώντας ανάλυση σχεδιαστικών συμβιβασμών βασισμένη σε σενάρια λειτουργίας.

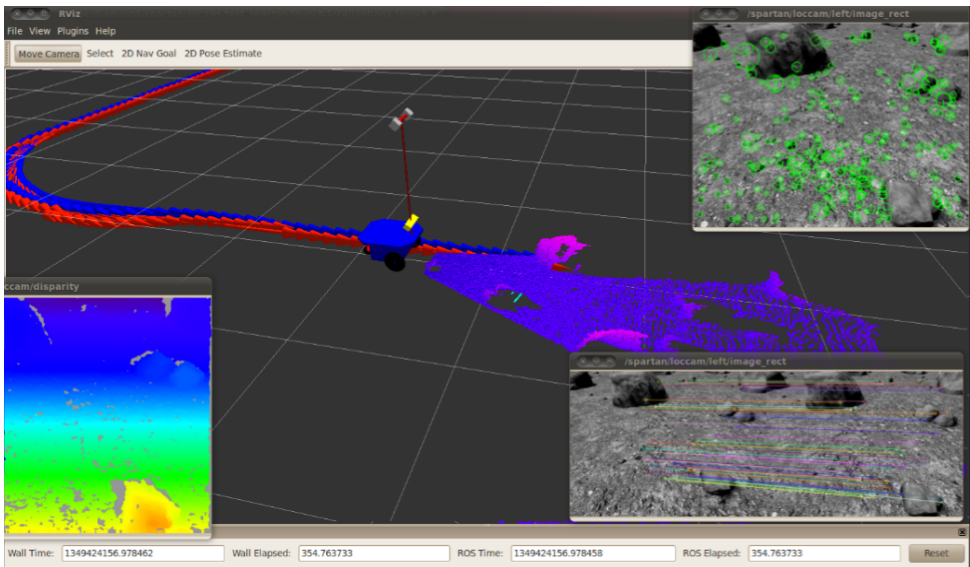
νας καθώς και το πλήθος αυτών σε βάρος του χρόνου επεξεργασίας. Συνολικά διακριτοποιήθηκαν τέσσερα σενάρια λειτουργίας, αναφορικά προς τις διαθέσιμες εικόνες εισόδου, καθώς το σύστημα έπρεπε να δοκιμασθεί σε διαφορετικά περιβάλλοντα λειτουργίας. Περισσότερες πληροφορίες για αυτά τα σενάρια ει-κονοροών μπορούν να βρεθούν στη δημοσίευση [26]. Το Σχήμα 2.17 απεικονίζει το σφάλμα της εκτίμησης κίνησης, των τεσσάρων σεναρίων σε σχέση με τον αριθμό των εξαγόμενων χαρακτηριστικών και το μέσο χρόνο επεξεργασίας για την εξαγωγή τους. Από αυτή την ανάλυση είναι επιτεύξιμη η παραγωγή μίας Pareto επιφάνειας των σημείων αξιολόγησης του συστήματος από τον τρισδιά-στατο χώρο εξερεύνησης. Αυτή η έρευνα επιτρέπει τη δημιουργία των βέλτιστων ρυθμίσεων των παραμέτρων RTS ανάλογα με τις προδιαγραφές του έργου. Για παράδειγμα, στο έργο SPARTAN υπάρχει ο περιορισμός για μόλις ένα δευτερό-λεπτο επεξεργασίας για την εκτίμηση κίνησης. Ως εκ τούτου, αναφορικά προς τα σενάρια 1 και 2, το σύστημα μπορεί να ρυθμιστεί ώστε να παράγει μέχρι 200 χαρακτηριστικά εικόνες κατά μέσο όρο και να λειτουργεί σε λιγότερο από ένα δευτερόλεπτο, παρέχοντας ταυτόχρονα υψηλή ακρίβεια στην εκτίμηση της κίνησης. Ωστόσο είναι εμφανές ότι τα σενάρια 3 και 4 αποτυγχάνουν σε αυτή τη προσέγγιση. Σε μια τέτοια περίπτωση, η ομάδα σχεδιασμού θα πρέπει να τροποποιήσει κατάλληλα τόσο τις υλοποιημένες μονάδες υλικού, ή ακόμα και

τις υπό-ανάπτυξη μονάδες, ώστε το σύστημα να πληρεί τις προδιαγραφές λειτουργίας για όλα τα σενάρια. Σημειώνεται, ότι αυτό το στάδιο ελέγχου δεν θα μπορούσε να επιτευχθεί σε πρόωρα στάδια λειτουργίας, με τις συμβατικές μεθόδους σχεδίασης, χωρίς την ύπαρξη της εικονικής μηχανής η οποία προσφέρει αφαιρετικότητα στο υπό-ανάπτυξη σύστημα, καθώς και τις βιβλιοθήκες Host2VP και VP2HW οι οποίες επιτρέπουν την συν-προσομοίωση της εικονικής μηχανής με το FPGA.

Επίσης, όπως προαναφέρθηκε στην υποενότητα 2.4.1, η διαμόρφωση του διαύλου επικοινωνίας HotTalk επηρεάζει σημαντικά την απόδοση της συν-προσομοίωσης. Προς ποσοτικοποίησης αυτής της συσχέτισης διαμόρφωσης-επίδοσης, το Σχήμα 2.18 απεικονίζει την επίδραση του μεγέθους της μνήμης ανατροφοδότησης στην ταχύτητα συν-προσομοίωσης, καθώς και το ποσοστό αξιοποίησης του συστήματος επαλήθευσης DUT. Όπως φαίνεται η αύξηση μνήμης μέχρι τα 30KB αντιστοιχεί σε μια σχεδόν γραμμική συσχέτιση μεταξύ του αριθμού των κύκλων συν-προσομοίωσης και του ποσοστού χρησιμοποίησης της μονάδας DUT. Σημειώνεται ότι η χρήση της μονάδας DUT είναι πολύ χαμηλή, δεδομένου ότι η αργή σύνδεση του διαύλου HotTalk, μέσω σύνδεσης Ethernet 100Mbps, καθυστερεί τη συνεχή αλυσιδωτή λειτουργία του συστήματος SPARTAN. Ωστόσο, από μια τέτοια ανάλυση, η ομάδα σχεδίασης υπολογίζει λεπτομερώς την εναέρια επικοινωνία του διαύλου HotTalk και έτσι μπορεί να εκτιμήσει αναλυτικά την απόδοση του τελικού συστήματος.

Σχήμα 2.18 Αξιολόγηση του συστήματος SPARTAN αναφορικά προς την αποτελεσματικότητα της συν-προσομοίωσης.

Το Σχήμα 2.19 απεικονίζει ένα στιγμιότυπο αξιολόγησης του συστήματος SPARTAN στο περιβάλλον ROS, με εφαρμογή του πλαισίου Plug&Chip. Όπως φαίνεται, το σύστημα δύναται να προσομοιωθεί σε ρεαλιστικά σενάρια λειτουργίας, κατά τα πρώιμα στάδια σχεδιασμού, όπου δεν είναι ολοκληρωμένος ο προγραμματισμός των αλγορίθμων στην τελική τους μορφή, σύμφωνα με τις αποφάσεις διαχωρισμού υλικού/λογισμικού, κατά τη φάση του συν-σχεδιασμού. Στο συγκεκριμένο στιγμιότυπο οι τρεις αλγόριθμοι μηχανικής όρασης, που απεικονίζονται στα αναδεδόμενα παράθυρα του σχήματος, εκτελούνται σε μία συσκευή FPGA, μέσω του πλαισίου Plug&Chip, ενώ το σύνολο των υπόλοιπων αλγορίθμων εκτίμησης κίνησης του ρομποτικού συστήματος (μπλε και κόκκινη γραμμή πορείας), καθώς και το λογισμικό εικονικής εκτέλεσης που φαίνεται στο Σχήμα 2.19 εκτελούνται σε έναν υπολογιστή υποδοχής (host), μέσω του λειτουργικού συστήματος ROS και της βιβλιοθήκης OpenCV.



Σχήμα 2.19 Στιγμιότυπο αξιολόγησης του συστήματος SPARTAN στο περιβάλλον ROS, με εφαρμογή του πλαισίου Plug&Chip.

Παράλληλα, το Σχήμα 2.20 απεικονίζει ένα στιγμιότυπο αξιολόγησης ενός συστήματος ελέγχου κινητήρα αυτοκινήτου (Engine Control Unit - ECU). Σκοπός του συστήματος είναι ο έλεγχος διαδρομής κίνησης (track control) με τη βοήθεια αλγορίθμων μηχανικής όρασης⁷. Σε αυτό το στιγμιότυπο ο αλγόριθμος μηχανικής όρασης SURF εκτελείται σε μία συσκευή FPGA, μέσω του πλαισίου Plug&Chip, ενώ το υπόλοιπο σύστημα εκτελείται σε ένα περιβάλλον εικονικής εξομοίωσης για αγώνες ταχύτητας αυτοκινήτων [30]. Στόχευση της εργασίας αυ-

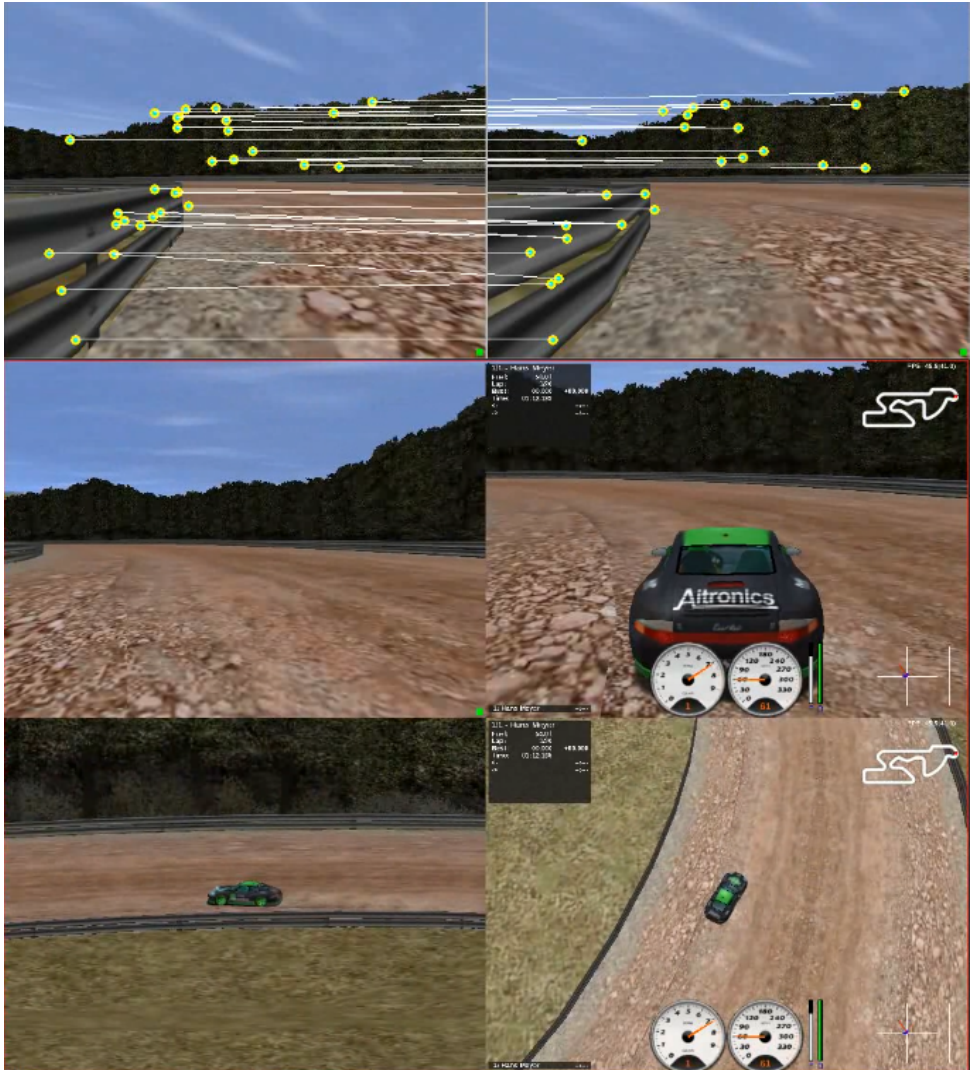
⁷Το σύστημα σχεδιάστηκε στα πλαίσια ενός διαγωνισμού εικονικής προτυποποίησης για ερευνητικές εργασίες σε επίπεδο διατριβής, της εταιρείας Cadence, για τον κλάδο της αυτοκινητοβιομηχανίας [29] και απέσπασε το 2^ο βραβείο.

τής είναι ο εντοπισμός σημείων ενδιαφέροντος σε διαδοχικές εικόνες της εμπρός θέσης του οχήματος προκειμένου να βελτιστοποιηθεί η πλοήγηση. Στη συγκεκριμένη εφαρμογή το πλαίσιο Plug&Chip επέτρεψε τη δοκιμή πολλαπλών σεναρίων χρήσης της υπό-ανάπτυξης ECU ακόμα και με την έλλειψη πραγματικού φυσικού μοντέλου αυτοκινήτου και προδιαγραφών περιβάλλοντος. Αυτή η πρόωρη αξιολόγηση του αλγορίθμου SURF βοήθησε στη βελτιστοποίησή του σε πλήθος σεναρίων λειτουργίας. Περισσότερες πληροφορίες για τον αλγόριθμο SURF μπορούν να βρεθούν στην εργασία [31].

2.5. Γενικά Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάζεται ένα νέο πλαίσιο για την υποστήριξη ταχείας, καθώς και σταδιακής προτυποποίησης, ετερογενών ενσωματωμένων συστημάτων, τόσο σε επαναδιαμορφούμενες αρχιτεκτονικές, όσο και σε εξειδικευμένες λύσεις με συμβατική τεχνολογία ολοκλήρωσης 2-D, TSMC 45nm, καθώς και σε τρισδιάστατη ολοκλήρωση 3-D. Μεταξύ άλλων, το πλαίσιο Plug&Chip παρέχει στις ομάδες σχεδιασμού τις συνδέσεις μεταξύ των διαφόρων σχεδιαστικών μονάδων του υπό-ανάπτυξη συστήματος, όπως είναι το “εξαρτώμενο από το υλικό λογισμικό”, το λογισμικό ελέγχου ροής, καθώς και οι εξειδικευμένες μονάδες υλικού. Αυτή η διαδικασία εννοιολογικού και αρχιτεκτονικού διαχωρισμού επιτρέπει τη σταδιακή ανάπτυξη ενός σύνθετου συστήματος, όπως είναι τα ετερογενή συστήματα. Επίσης, μέσα από το προτεινόμενο πλαίσιο καθίσταται δυνατή η εκκίνηση της ανάπτυξης, η δοκιμή και η επικύρωση του ενσωματωμένου λογισμικού, πολύ νωρίτερα από ό,τι ήταν δυνατό στο παρελθόν. Πειραματικά αποτελέσματα από διάφορα σενάρια ελέγχου, που εκτείνονται από το σχεδιασμό εξειδικευμένων πυρήνων υλικού, έως το συν-σχεδιασμό υλικού/λογισμικού ετερογενών ενσωματωμένων συστημάτων, αποδεικνύουν την αποτελεσματικότητα του προτεινόμενου πλαισίου, καθώς παρέχει ανταγωνιστικά αποτελεσματικότερες προσομοιώσεις, αναφορικά προς τις υφιστάμενες προσεγγίσεις.

2



Σχήμα 2.20 Στιγμιότυπο αξιολόγησης ενός συστήματος ελέγχου κινητήρα αυτοκινήτου (ECU) σε περιβάλλον Linux, με εφαρμογή του πλαισίου Plug&Chip.

Βιβλιογραφία

- [1] V. F. Pavlidis and E. G. Friedman, *Three-dimensional Integrated Circuit Design* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2009).
- [2] ITRS, *International technology roadmap for semiconductors*, (2012).
- [3] Synopsys, *Hybrid prototyping delivers the best of both virtual and fpga prototyping to soc hardware and software teams*, .
- [4] OVP, *Open virtual platforms (ovp)*, online: www.ovpworld.org, (2013).
- [5] R3LOGIC, *R3logic inc.* (2013).
- [6] CADENCE, *Cadence 3d-ic solution*, (2013).
- [7] T. Borgstrom, E. Haritan, R. Wilson, D. Abada, R. Chandra, C. Cruse, A. Dauman, O. Mielo, and A. Nohl, *System prototypes: Virtual, hardware or hybrid?* in *Design Automation Conference, 2009. DAC '09. 46th ACM/IEEE* (2009) pp. 1–3.
- [8] SCE-MI, *Standard co-emulation modeling interface (sce-mi) reference manual*, (2011), http://www.accellera.org/downloads/standards/sce-mi/SCE_MI_v21-110112-final.pdf.
- [9] Altera, *Altera virtual target*, (2013), <http://www.altera.com/devices/processor/arm/cortex-a9/virtual-target/proc-a9-virtual-target.html>.
- [10] QEMU, *1st international qemu users forum*. (2011), <http://adt.cs.upb.de/quf/>.
- [11] T.-C. Yeh, Z.-Y. Lin, and M.-C. Chiang, *A novel technique for making qemu an instruction set simulator for co-simulation with systemc*, in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2011*, Vol. I (2011) pp. 288 – 291.
- [12] J. Jovic, S. Yakoushkin, L. Murillo, J. Eusse, R. Leupers, and G. Ascheid, *Hybrid simulation for extensible processor cores*, in *Design, Automation Test in Europe Conference Exhibition (DATE), 2012* (2012) pp. 288–291.
- [13] HAPS, *Synopsys high-performance asic prototyping systems*, (2013), <http://www.synopsys.com/Systems/FPGABasedPrototyping/Pages/HAPS.aspx>.
- [14] Xilinx, *Xilinx hybrid co-simulation flow quick reference*, (2013), http://www.xilinx.com/tools/feature/14_1_isim_hw_cosim_qrg.pdf.
- [15] HTG, *Hitech global virtex-6 fpga x8 pci express gen 2 kit*, .
- [16] OpenCores, *Ethernet mac 10/100 mbps*, (2013).
- [17] IP-XACT, *Ip-xact technical committee*, (2013).

- [18] N. Selvakkumaran and G. Karypis, *Multiobjective hypergraph-partitioning algorithms for cut and maximum subdomain-degree minimization*, *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* **25**, 504 (2006).
- [19] K. Siozios and D. Soudris, *A tabu-based partitioning and layer assignment algorithm for 3-d fpgas*, *Embedded Systems Letters, IEEE* **3**, 97 (2011).
- [20] I. Savidis, S. M. Alam, A. Jain, S. Pozder, R. E. Jones, and R. Chatterjee, *Electrical modeling and characterization of through-silicon vias (tsvs) for 3-d integrated circuits*, *Microelectronics Journal* **41**, 9 (2010).
- [21] OpenRISC, *Openrisc 1000 project main page*, (2012), http://opencores.org/or1k/Main_Page.
- [22] M. R. Guthaus, J. S. Ringenberg, D. Ernst, T. M. Austin, T. Mudge, and R. B. Brown, *Mibench: A free, commercially representative embedded benchmark suite*, in *Proceedings of the Workload Characterization, 2001. WWC-4. 2001 IEEE International Workshop*, WWC '01 (IEEE Computer Society, Washington, DC, USA, 2001) pp. 3–14.
- [23] S. Gupta, M. Hilbert, S. Hong, and R. Patti, *Techniques for producing 3d ics with high-density interconnect*, in *Proceedings of the 21st International VLSI Multilevel Interconnection Conference* (2004).
- [24] SPARTAN, (2013), <http://proteas.microlab.ntua.gr/spartan>.
- [25] ESA-Compatible-VHDL., <http://www.eda.org/rassp/vhdl/guidelines/guidelines.html>, .
- [26] I. Kostavelis, L. Nalpantidis, E. Boukas, M. A. Rodrigalvarez, I. Stamoulias, G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, and A. Gasteratos, *Spartan: Developing a vision system for future autonomous space exploration robots*, *Journal of Field Robotics* **31**, 107 (2014).
- [27] S. V. Gheorghita, M. Palkovic, J. Hamers, A. Vandecappelle, S. Mamagkakis, T. Basten, L. Eeckhout, H. Corporaal, F. Catthoor, F. Vandeputte, and K. D. Bosschere, *System-scenario-based design of dynamic embedded systems*, *ACM Trans. Des. Autom. Electron. Syst.* **14**, 3:1 (2009).
- [28] D. Diamantopoulos, K. Siozios, G. Lentaris, D. Soudris, and M. Rodrigalvarez, *Spartan project: On profiling computer vision algorithms for rover navigation*, in *Adaptive Hardware and Systems (AHS), 2012 NASA/ESA Conference on* (2012) pp. 174–181.
- [29] Cadence, *Cadence thesis contest for automotive embedded systems, cdnlive! emea 2013*, .
- [30] SD, *Speed dreams: An open motorsport simulator*, .

- [31] G. Lentaris, I. Stamoulias, D. Diamantopoulos, K. Siozios, and D. Soudris, *An fpga implementation of the surf algorithm for the exomars programme*, Workshop on Reconfigurable Computing .

3

Πολυεπίπεδη Σύνθεση Ετερογενών Αρχιτεκτονικών Ειδικού Σκοπού

Στο παρόν κεφάλαιο παρουσιάζονται τεχνικές πολυεπίπεδης σύνθεσης για ετερογενείς αρχιτεκτονικές εξειδικευμένης αποστολής. Αναλύονται οι ιδιαιτερότητες της αρχιτεκτονικής και τα βασικά σύγχρονα προβλήματα του σχεδιασμού των ετερογενών συστημάτων εξειδικευμένης αποστολής. Προτείνονται μεθοδολογικά πλαίσια και εργαλεία σχεδιασμού, τα οποία βελτιστοποιούν αντικρουόμενες μετρικές σχεδίασης που απαιτεί η σύγχρονη αγορά ημιαγωγών, όπως η κατανάλωση ενέργειας, η επιφάνεια πυριτίου, η ανοχή θερμοκρασίας και η αξιοπιστία. Παράλληλα, προτείνονται ροές σχεδίασης που βελτιστοποιούν τη διαδικασία συν-σχεδιασμού υλικού/λογισμικού για ετερογενείς πλατφόρμες. Το κεφάλαιο αναπτύσσεται με τη μελέτη τριών ετερογενών πλατφορμών, μίας αρχιτεκτονικής SoC για τηλεπικοινωνίες (SDR), ενός συστήματος ρομποτικής όρασης για διαστημικές εφαρμογές και μίας αρχιτεκτονικής SoC σχεδιασμένη με τη τρισδιάστατη τεχνολογία ολοκλήρωσης. Το κεφάλαιο 3 συνδέεται με τη γενικότερη μεθοδολογία της διατριβής που παρουσιάζεται στην Ενότητα 1.3, ως προς τη συνεισφορά του στις κατηγορίες της “Εξειδικευμένες Πλατφόρμες - Εργαλεία EDA για συστήματα 2D/2.5D/3D τεχνολογίας ολοκλήρωσης”, “Σύνθεσης του Μονοπατιού Επιτάχυνσης” και “Συστημικής Αρχιτεκτονικής Διερεύνησης”.

3.1. Εισαγωγή στις Αρχιτεκτονικές Ειδικού Σκοπού

Η ανάγκη για εξειδίκευση αποτελεί θεμέλιο λίθο των σύγχρονων υπολογιστικών συστημάτων, καθώς η τάση της αγοράς και οι ανάγκες των χρηστών επιβάλλουν τη σχεδίαση συστημάτων με επαρκή επεξεργαστική ισχύ, φορητότητα και προσαρμοστικότητα στα εναλλακτικά πρότυπα λειτουργίας και περιβάλλοντα χρήσης. Κατά το παρελθόν αυτά τα χαρακτηριστικά μπορούσαν να προσεγγι-

στούν ταυτόχρονα μόνο με τη χρήση ολοκληρωμένων κυκλωμάτων ειδικού σκοπού (Application Specific Integrated Circuits, ASICs). Ωστόσο το μεγάλο κόστος κατασκευής, ο επιμήκης χρόνος σχεδίασης και η έλλειψη δυνατότητας προσαρμογής αυτών των σχεδιαστικών κυκλωμάτων τα καθιστά μη-κλιμακούμενα στις σύγχρονες σχεδιαστικές προκλήσεις. Αντίθετα οι επαναδιαμορφούμενες πλατφόρμες αποτελούν μία ελκυστική αρχιτεκτονική που ανταποκρίνεται στα προαναφερόμενα λειτουργικά χαρακτηριστικά. Βέβαια το υψηλό κόστος αγοράς ανά μονάδα πλατφόρμας και η υψηλή κατανάλωση στατικής ενέργειας λόγω της αδρομερούς υφής αυτών των αρχιτεκτονικών, τις καθιστά δύσχρηστες για ένα περιβάλλον εφαρμογών, όπου το μικρό κόστος της μαζικής παραγωγής υπερτερεί έναντι των υπολοίπων χαρακτηριστικών. Στο παρόν Κεφάλαιο προτείνονται μεθοδολογικά πλαίσια καθώς και καινοτόμα εργαλεία σχεδίασης, τα οποία βελτιστοποιούν τη διαδικασία σχεδίασης τόσο επαναδιαμορφούμενων όσο και εξειδικευμένων συστημάτων.

Δια τα πρώτα, προτείνεται μία μεθοδολογία συν-σχεδιασμού υλικού/λογισμικού για ετερογενή επαναδιαμορφούμενες πλατφόρμες, η οποία αξιολογεί το διαμερισμό των διεργασιών του συστήματος στους κατανεμημένους υπολογιστικούς πόρους, βάση ενός καινοτομικού σταδίου αξιολόγησης των εφαρμογών σε πολυ-κριτηριακό επίπεδο. Το πλαίσιο χρησιμοποιείται στη σχεδίαση ενός ετερογενούς συστήματος PC-FPGA, προοριζόμενου για αυτόνομη πλοήγηση διαστημικών ρομποτικών οχημάτων. Το προτεινόμενο πλαίσιο αναλύεται στην Ενότητα 3.4.

Δια τα δεύτερα, αναπτύσσεται ένα μεθοδολογικό πλαίσιο καθώς ένα σύνολο εργαλείων σχεδίασης, το οποίο προσφέρει αρχιτεκτονικές λύσεις που συμβιβάζουν πλήθος αντικρουόμενων σχεδιαστικών παραμέτρων για εξειδικευμένα ολοκληρωμένα κυκλώματα ASIC. Κίνητρο έρευνας για τη συγκεκριμένη εργασία αποτελεί η συνεχής κλιμάκωση της τεχνολογίας ολοκλήρωσης, η οποία τείνει να τοποθετήσει αυξημένους υπολογιστικούς πόρους σε μικρότερες διαστάσεις, με αποτέλεσμα την συγκέντρωση μεγάλης πυκνότητας ενέργειας ανά μονάδα επιφάνειας πυριτίου. Μεταξύ των μετρικών βελτιστοποίησης, συγκαταλέγονται η επιφάνεια πυριτίου, το θερμικό προφίλ, η κατανάλωση ενέργειας και η αξιοπιστία του κυκλώματος. Το προτεινόμενο πλαίσιο, το οποίο αναλύεται στην Ενότητα 3.2, αξιολογείται μέσω ενός συστήματος SoC για εφαρμογές επικοινωνίας SDR.

Επίσης αναφορικά προς τη σχεδίαση εξειδικευμένων συστημάτων, προτείνεται μία μεθοδολογία για ταχεία εκτίμηση σχεδιαστικών λύσεων από την υιοθέτηση της τρισδιάστατης τεχνολογίας ολοκλήρωσης πυριτίου (3-D). Σύμφωνα με τους τεχνολογικούς οδικούς χάρτες των οργανισμών ITRS και HiPEAC, η συγκεκριμένη τεχνολογία αναμένεται να κυριαρχήσει στην αγορά τα επόμενα χρόνια, υποστηρίζοντας μεγάλη πυκνότητα λογικής ολοκλήρωσης ανά μονάδα επιφάνειας και σύνδεση ετερογενών τεχνολογιών πυριτίου στο τρισδιάστατο χώρο. Η προτεινόμενη τεχνική μελετάει το πρόβλημα κατάτμησης της εφαρμογής και ανάθεσης των αρχιτεκτονικών πυρήνων στα επίπεδα πυριτίου του τρισδιάστατου chip, ενώ επιτυγχάνει τη σχεδίαση αυτών, με εμπορικά εργαλεία σχεδίασης 2-D, δίχως έκπτωση των σχεδιαστικών επιλογών. Το προτεινόμενο πλαίσιο ανα-

λύεται στην Ενότητα 3.5.

3.2. Αρχιτεκτονικές SDR με βελτιστοποιημένο προφίλ θερμότητας και αξιοπιστίας

3.2.1. Εισαγωγή

Τα τελευταία χρόνια η επικοινωνία αποτελεί μια από τις κύριες εφαρμογές της τεχνολογίας υπολογιστών. Αρχιτεκτονικές που διευκολύνουν την επικοινωνία, όπως τα κινητά τηλέφωνα και τα ασύρματα δίκτυα έχουν γίνει οι κύριοι συντελεστές στην εξέλιξη των μικροεπεξεργαστών και των υπολογιστικών συστημάτων. Με την εξέλιξη των ασύρματων κινητών επικοινωνιών, το πρόβλημα έχει στραφεί στα διαδικτυακά πρωτόκολλα και στην επεξεργασία σημάτων που απαιτούνται για τη κλιμάκωση του απαραίτητου εύρους ζώνης των σύγχρονων εφαρμογών. Χαρακτηριστικό παράδειγμα αυτής της κλιμάκωσης αποτελεί ο συνεχής αυξανόμενος αριθμός ασύρματων πρωτοκόλλων επικοινωνιών για διάφορους τύπους δικτύων, π.χ. 2G, 3G, GPRS, 802.11a/b/g/n/ac, κλπ.

Η τεχνολογία Software Defined Radio (SDR) αναπτύχθηκε για να υποστηρίξει την επαναδιατάξη, την επεκτασιμότητα και τη διαλειτουργικότητα μεταξύ διαφορετικών ασύρματων δικτυακών συσκευών [1]. Η τεχνολογία SDR αποτελείται από υλικό και λογισμικό το οποίο μπορεί να αναδιαρθρωθεί δυναμικά προκειμένου να επιτρέψει επικοινωνία σε ένα μεγάλο εύρος μεταβαλλόμενων πρωτοκόλλων επικοινωνίας και φυσικών καναλιών επικοινωνιακών ζεύξεων. Η πρόσφατη τεχνολογία SDR επιτρέπει στους αρχιτέκτονες συστημάτων να σχεδιάζουν ασύρματες συσκευές πολλαπλών καταστάσεων λειτουργίας και εύρους ζώνης, καθώς και εξοπλισμό δικτύου, ο οποίος μπορεί εύκολα να βελτιωθεί, να επεκταθεί και να αναβαθμιστεί μέσω ενημερώσεων λογισμικού και αναδιαμορφώσεων υλικού.

Καθώς τα συστήματα SDR χαρακτηρίζονται από υψηλή απόδοση, μικρές απαιτήσεις ενέργειας και μικρό χρόνο διάθεσης στην αγορά (time-to-market), έχουν ήδη δημοσιευθεί εργασίες οι οποίες προτείνουν τη χρήση αρχιτεκτονικών System-on-Chip (SoC) [2, 3]. Η ικανοποίηση των θερμικών περιορισμών και η μείωση των hotspots θερμοκρασίας σε αυτές τις πλατφόρμες είναι κρίσιμα ζητήματα κατά το σχεδιασμό αξιόπιστων συστημάτων. Επιπλέον, η θερμοκρασία του chip έχει σημαντικό αντίκτυπο στην απόδοση, στην αξιοπιστία, στη κατανάλωση ενέργειας καθώς και στο κόστος ψύξης και συσκευασίας. Η διαδικασία σχεδιασμού που λαμβάνει υπόψη τη θερμότητα είναι δύσκολη, ενώ ο σχεδιασμός ενός ολοκληρωμένου κυκλώματος το οποίο θα είναι λειτουργικό στο χειρότερο σενάριο αναφορικά προς τη κατανάλωση ενέργειας μπορεί να είναι απαγορευτικά ακριβό.

Προηγούμενες μελέτες έχουν δείξει ότι η θερμοκρασία ενός ολοκληρωμένου κυκλώματος είναι στενά συνδεδεμένη με θέματα αξιοπιστίας [4]. Για παράδειγμα, η αναπτυσσόμενη θερμότητα σε ένα ολοκληρωμένο κύκλωμα κατά το χρόνο εκτέλεσης μπορεί να μοντελοποιηθεί με τη σχέση Coffin-Manson, η οποία συνδέει σε εκθετικό βαθμό τον αριθμό των κύκλων εκτέλεσης έως την κατάρ-

ρευση της λειτουργίας του κυκλώματος με το μέγεθος της θερμότητας [5]. Οι υπάρχουσες προσεγγίσεις διαχειρίζονται το πρόβλημα της θερμότητας με τεχνικές μείωσης ενέργειας.

Μια τυπική και ευρέως διαδεδομένη υλοποίηση αυτής της λύσης είναι η τεχνική Dynamic Voltage and Frequency Scaling (DVFS) [6]. Δεδομένης της σχετικά τετράγωνης σχέσης μεταξύ της παροχής τάσης ηλεκτρικού ρεύματος και της κατανάλωσης ενέργειας, οι τεχνικές DVFS είναι ικανές να μειώσουν σημαντικά τη κατανάλωση ενέργειας, επιβάλλοντας ωστόσο, μικρότερη συχνότητα λειτουργίας. Επιπλέον, αυτές οι τεχνικές δε μπορούν να εγερθούν ότι τα hotspots και οι διαβαθμίσεις θερμοκρασίας στην επιφάνεια πυριτίου του chip θα μειωθούν, εφόσον εφαρμόζονται κατά το χρόνο εκτέλεσης ως αντίδραση στη θερμική εικόνα του chip.

Μία επιπλέον τεχνική που χρησιμοποιείται στην μείωση των hotspots θερμοκρασίας, ειδικά για πολυπύρηνες αρχιτεκτονικές, βασίζεται σε μεθόδους εξισορρόπησης των εργασιών εκτέλεσης [7]. Παρόλο που αυτές οι τεχνικές έχουν μελετηθεί για γενικής χρήσης παράλληλους υπολογιστές, στοχεύουν κατά κύριο λόγο στην αποφυγή συμφορήσεων των διεργασιών εκτέλεσης παρά σε θέματα θερμότητας.

Μια παρόμοια προσέγγιση παρουσιάζεται στην εργασία [8], όπου τα νήματα εκτέλεσης ενός συστήματος ανταλλάσσονται μεταξύ θερμικά ζεστών και ψυχρών πυρήνων, προκειμένου να διατηρηθεί συνολική θερμική ισορροπία στην επιφάνεια του chip. Ωστόσο, η προσέγγιση αυτή υποθέτει ότι τα νήματα μεταφέρονται αυτούσια ανάμεσα σε διαφορετικούς πυρήνες, ενώ δεν υπάρχει δυνατότητα μεταφοράς μέρους της εργασίας των νημάτων μεταξύ πυρήνων. Συνεπώς, στη περίπτωση όπου υπάρχει ανισομερής καταμερισμός του υπολογιστικού φόρτου σε νήματα εκτέλεσης, υπάρχει μεγάλη πιθανότητα να αυξηθεί ραγδαία η θερμοκρασία ενός πυρήνα στον οποίο έχει ανατεθεί ένα νήμα με μεγάλο χρόνο εκτέλεσης, εφόσον δεν δύναται να διακοπεί η λειτουργία εκτέλεσης σε αυτόν τον πυρήνα, προκειμένου να επανεκκινηθεί σε έναν άλλον χαμηλότερης θερμοκρασίας. Προς λύση αυτού του περιορισμού, υπάρχουν δημοσιευμένες εργασίες, οι οποίες προτείνουν διαχείριση θερμότητας μέσω βελτιστοποιήσεων του μεταγλωττιστή [9, 10]. Ο κύριος περιορισμός αυτής της προσέγγισης είναι η δυσκολία υπολογισμού, με επαρκή ακρίβεια, των θερμοκρασιακών μεταβολών που συμβαίνουν λόγω των εκδόσεων του πηγαιού κώδικα. Επιπλέον, αυτή η τεχνική είναι εφαρμόσιμη μόνο σε μικρό ποσοστό πλατφορμών SoCs, εφόσον οι απαιτούμενες αρχιτεκτονικές λεπτομέρειες αποτελούν προϊόν πνευματικής ιδιοκτησίας και είναι κλειστού κώδικα. Προκειμένου να ξεπεραστούν τέτοιοι περιορισμοί, έχουν χρησιμοποιηθεί υπολογιστικές τεχνικές σε επίπεδο εντολών που αντιστοιχίζουν εντολές κώδικα μηχανής με θερμικά κατώφλια, βάση εμπειρικών μετρήσεων.

Ένα κοινό μειονέκτημα όλων αυτών των τεχνικών που συζητήθηκαν παραπάνω είναι ότι δεν ενσωματώνουν κάποιο μηχανισμό για διαχείριση του θερμικού προφίλ των υπολογιστικών πυρήνων εξετάζοντας το ιστορικό θερμοκρασίας κατά το χρόνο εκτέλεσης. Αυτό το χαρακτηριστικό παρέχει χρήσιμες πληροφορίες για τη μελλοντική συμπεριφορά του συστήματος και μπορεί να εκμεταλλευθεί προκειμένου να κατευθύνει τη διαδικασία ανάθεσης εργασιών στα

υπολογιστικά στοιχεία του συστήματος. Επιπρόσθετα, οι υφιστάμενες προσεγγίσεις επικεντρώνονται σε τεχνικές, οι οποίες αξιολογούν το θερμικό προφίλ των εφαρμογών σε πλατφόρμες SoC, βασισμένες σε αποτελέσματα προσομοίωσης.

Αυτές οι προσεγγίσεις υποθέτουν ότι η πλατφόρμα ανάπτυξης αγνοεί πιθανές βελτιστοποιήσεις σε επίπεδο αρχιτεκτονικής [11]. Στα πλαίσια αυτής της έρευνας προτείνεται μια νέα μεθοδολογία καθώς και το υποστηρικτικό πλαίσιο λογισμικού, για σχεδιασμό σε επίπεδο αρχιτεκτονικής καθώς και φυσικού επιπέδου, σύμφωνα με θερμικούς περιορισμούς. Συγκεκριμένα, η βασική ιδέα που εισηγείται η παρούσα εργασία εκμεταλλεύεται την επιλεκτική αντιγραφή των μονάδων υλικού που παρουσιάζουν αυξημένη πυκνότητα ενέργειας. Έτσι, χρονοπρογραμματίζοντας κατάλληλα τις τρέχουσες διεργασίες στα διαθέσιμα αντίγραφα των μονάδων υλικού, καθίσταται εφικτή η αποδοτική μείωση του θερμικού προφίλ του ολοκληρωμένου κυκλώματος.

Η προτεινόμενη προσέγγιση στοχεύει στη βελτιστοποίηση της θερμοκρασίας, ενώ μπορεί να θεωρηθεί ως μια προληπτική στρατηγική που ελαχιστοποιεί τη καταπόνηση του ολοκληρωμένου κυκλώματος από τη θερμότητα, κατά το χρόνο εκτέλεσης. Το εισαγόμενο πλαίσιο δεν επιβάλλει καμία τροποποίηση στην αρχιτεκτονική ή/και στον μεταγλωττιστή, ενώ μπορεί να εφαρμοστεί παράλληλα με κάποια άλλη μεθοδολογία βελτιστοποίησης του θερμικού προφίλ, όπως αυτές που αναφέρθηκαν στις προηγούμενες παραγράφους.

Συγκεκριμένα, σε αυτή την εργασία εξετάζουμε την ανάπτυξη ενός αυτοματοποιημένου σχεδιαστικού πλαισίου εξερεύνησης του χώρου λύσεων, το οποίο αξιολογεί ένα μεγάλο αριθμό αρχιτεκτονικών λύσεων. Κάθε λύση διερευνά επιλεκτική αντιγραφή μονάδων υλικού. Η αυτόματη εξερεύνηση του χώρου λύσεων, μέσω κατάλληλου λογισμικού, επιτρέπει το χαρακτηρισμό υψηλής ποιότητας Pareto λύσεων, σε αντίθεση με πολλές παρόμοιες υφιστάμενες τεχνικές βελτιστοποίησης, οι οποίες χαρακτηρίζουν μόνο μια αρχιτεκτονική λύση [8, 9]. Συνεπώς, παρέχονται προτάσεις πολυ-κριτηριακής βελτιστοποίησης, ανάμεσα στην εκλυόμενη θερμότητα και το κόστος χρόνου, επιφάνειας πυριτίου και ενέργειας. Επιπλέον, το προτεινόμενο ερευνητικό πλαίσιο παρέχει αξιοσημείωτη επιτάχυνση στη διαδικασία εξερεύνησης του χώρου λύσεων.

Προηγούμενες εργασίες έχουν εισηγηθεί τη χρήση παραλληλισμού προκειμένου να επιτευχθεί μείωση ενέργειας, η οποία με τη σειρά της οδηγεί σε μείωση της θερμοκρασίας [12]. Συγκεκριμένα, προτείνεται μία παράλληλη υλοποίηση ενός σχεδιασμού, στην οποία τα παράλληλα νήματα εκτέλεσης επεξεργάζονται τμηματικά δείγματα της εισόδου. Ένας πολυπλέκτης συνδυάζει τις επί-μέρους εξόδους προκειμένου να δημιουργήσει ένα μοναδικό μονοπάτι δεδομένων.

Η κύρια διαφορά της προτεινόμενης εργασίας, συγκρινόμενη με αυτή τη προσέγγιση, είναι ότι η πρώτη δεν υποθέτει την παράλληλη εκτέλεση εργασιών των αντιγράφων μονάδων υλικού. Συγκεκριμένα, στη προτεινόμενη μεθοδολογία, μία μόνο από τα διαθέσιμες μονάδες υλικού είναι ενεργή σε κάθε χρονική στιγμή. Η επιλογή αυτής της ενεργής μονάδας βασίζεται στη θερμική της κατάσταση, όπως περιγράφεται στις επόμενες ενότητες. Οι συνεισφορές αυτής της ερευνητικής εργασίας μπορούν να συνοψιστούν ως ακολούθως:

- Εισηγείται βελτιστοποίηση της θερμοκρασίας λειτουργίας, μέσω επιλεκτι-

κής αντιγραφής συγκεκριμένων αρχιτεκτονικών μονάδων υλικού.

- Εισάγεται μια νέα μεθοδολογία που στοχεύει: (i) στην ελάττωση των κρίσιμων σημείων θερμοκρασίας (thermal hotspots) σε πλατφόρμες SoCs για αρχιτεκτονικές SDR και (ii) μείωση των θερμοκρασιακών διαβαθμίσεων στην επιφάνεια πυριτίου του ολοκληρωμένου κυκλώματος.
- Η προτεινόμενη μεθοδολογία παρέχει έναν αριθμό από αρχιτεκτονικές λύσεις Pareto, καθεμιά από τις οποίες βελτιστοποιεί διαφορετικούς σχεδιαστικούς περιορισμούς/κριτήρια.
- Προτείνεται μία νέα σχεδιαστική μεθοδολογία που μπορεί να εφαρμοσθεί παράλληλα με τις υπάρχουσες προσεγγίσεις που προτείνονται στη σχετική βιβλιογραφία [9] [7] [6], [8], [10], [13], [14], [15], [16].
- Παρέχεται αυτοματοποιημένη ροή εκτέλεσης εργαλείων CAD, μέσω της ανάπτυξης εξειδικευμένου λογισμικού, το οποίο είναι δημόσια διαθέσιμο για την ερευνητική κοινότητα σε ελεύθερο αποθετήριο λογισμικού [17].
- Η προτεινόμενη μεθοδολογία εφαρμόζεται σε μία πλατφόρμα SoC για ρεαλιστική και πλήρη αρχιτεκτονική SDR, αποτελούμενη από έναν RISC επεξεργαστή, τον LEON3 [18].

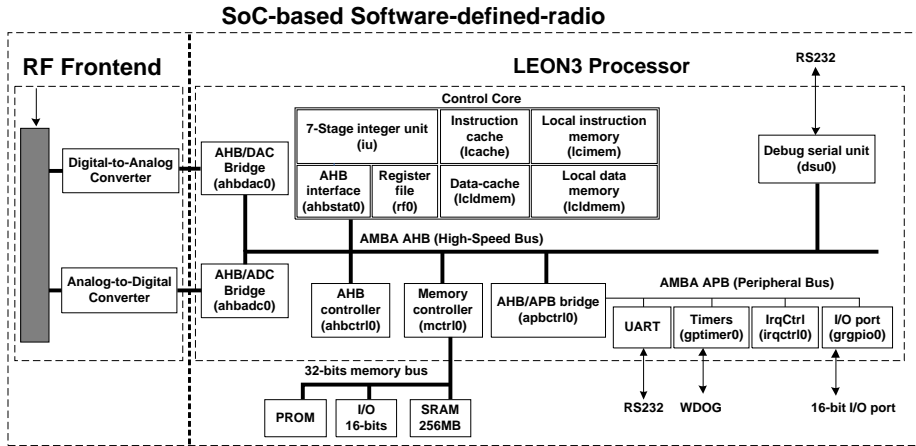
Τα πειραματικά αποτελέσματα αποδεικνύουν την αποδοτικότητα της προτεινόμενης μεθοδολογίας, δείχνοντας ότι η επιλεγμένη αρχιτεκτονική οδηγεί σε μείωση θερμοκρασίας κατά 8% (από 380 Kelvin σε 363 Kelvin), με ελεγχόμενη αύξηση της επιφάνειας πυριτίου έως 15%. Η επιτευχθείσα μείωση της θερμοκρασίας καθώς και η ελάττωση των περιοχών της επιφάνειας πυριτίου που παρουσιάζουν υψηλή θερμοκρασία, σε συγκεκριμένο φόρτο εργασιών, κατά το χρόνο εκτέλεσης, συμβάλουν αφενός στη μείωση του κόστους φύξης, καθώς και στη αύξηση της αξιοπιστίας του συστήματος. Αναφορικά προς τη τελευταία, στη παρούσα εργασία σημειώθηκε αξιοσημείωτη βελτίωση του χρόνου γήρανσης κατά 14%

Η υπόλοιπη εργασία σε αυτή την ενότητα οργανώνεται ως εξής: στην υπόενοτητα 3.2.2 περιγράφεται η βασική αρχιτεκτονική SDR, ενώ η υποενοτητα 3.2.3 περιγράφει το κίνητρο καθώς και τις παρατηρήσεις οι οποίες οδήγησαν στην εισήγηση της επιλεκτικής εισαγωγή αντιγράφων μονάδων υλικού. Στην ενότητα 3.2.4 περιγράφονται οι μικρο-αρχιτεκτονικές βελτιστοποιήσεις που προτείνονται προκειμένου να εφαρμοστεί η επιλεγμένη αντιγραφή μονάδων υλικού. Η προτεινόμενη μεθοδολογία αναλύεται λεπτομερώς στην υποενοτητα 3.2.5, ενώ τα αποτελέσματα αξιολόγησης παρουσιάζεται στην υποενοτητα 3.2.6. Τέλος, η υποενοτητα 3.2.7 παρέχει τα βασικά συμπεράσματα της παρούσας εργασίας.

3.2.2. Περιγραφή αρχιτεκτονικού προτύπου SDR

Τα τελευταία χρόνια έχουν προταθεί και αναπτυχθεί ποικίλες εναλλακτικές αρχιτεκτονικές SDR, τόσο από τον ερευνητικό χώρο, όσο και από το βιομηχανικό.

Το Σχήμα 5.3 απεικονίζει μια τυπική αρχιτεκτονική SDR, η οποία χρησιμοποιείται στα πλαίσια της παρούσας μελέτης. Το σύστημα αποτελείται από δύο διακριτά τμήματα, το αναλογικό και το ψηφιακό. Το αναλογικό (RF frontend) είναι υπεύθυνο για τη μετατροπή του σήματος μεταξύ του αναλογικού πεδίου και μιας ενδιάμεσης συχνότητας βασικής ζώνης, ενώ οι μονάδες A/D και D/A μετατρέπουν το σήμα μεταξύ αναλογικού-ψηφιακού και ψηφιακού-αναλογικού πεδίου, αντίστοιχα. Στη παρούσα ανάλυση η επεξεργασία βασικής ζώνης (baseband) εκτελείται μέσω λογισμικού σε ένα σύστημα SoC (System-on-Chip), βασισμένο στον ενσωματωμένο επεξεργαστή LEON3 [18].



Σχήμα 3.1 Το προτεινόμενο αρχιτεκτονικό πρότυπο πλατφόρμας SDR.

Ο επεξεργαστής LEON3 αποτελείται από τη μονάδα ακεραίων (integer unit), τη σκιώδη μνήμη (cache subsystem), τη μονάδα διαχείρισης μνήμης (memory management system) και τη διεπαφή διαύλου AMBA. Η μονάδα εντολών (instruction unit) είναι πλήρως συμβατή με την αρχιτεκτονική SPARC V8, ενώ προσφέρει σωλήνωση επτά επιπέδων. Η μονάδα ακεραίων διαθέτει προγραμματιζόμενες ξεχωριστές σκιώδεις μνήμες δεδομένων και εντολών (αρχιτεκτονική Harvard). Στη παρούσα εργασία χρησιμοποιήθηκε σκιώδης μνήμη μεγέθους 1KByte. Επιπλέον, η μονάδα ακεραίων περιέχει ένα προγραμματιζόμενο αρχείο καταχωρητών. Η σκιώδης μνήμη πρώτου επιπέδου (L1 cache) διαχειρίζεται από έναν ειδικό ελεγκτή, μέσω διεπαφής διαύλου AMBA AHB. Η επικοινωνία με τις περιφερειακές μονάδες του LEON3 πραγματοποιείται μέσω δύο ελεγκτών, τον ελεγκτή διαύλου AHB (Advanced High-performance Bus) και τον ελεγκτή διαύλου APB (Advanced Peripheral Bus). Ο πρώτος χρησιμοποιείται σε υψηλής ταχύτητας επικοινωνίες (π.χ. μονάδα ακεραίων, ελεγκτής μνήμης, κτλ.), ενώ ο δεύτερος παρέχει επικοινωνία σε μονάδες χαμηλού ρυθμού δεδομένων (π.χ. UARTs, μονάδες E/E, κτλ.). Τέλος, ο επεξεργαστής LEON3 περιέχει προγραμματιζόμενες τοπικές μνήμες για δεδομένα και εντολές που μπορούν να χρησιμοποιηθούν ως μνήμες scratchpad. Στη παρούσα εργασία επιλέχθηκαν αντίστοιχες μνήμες με-

γέθους 2KByte.

3.2.3. Κίνητρο έρευνας για θερμικά-ασφαλείς αρχιτεκτονικές SDR

Στην υποενοότητα αυτή χαρακτηρίζονται οι ιδιότητες των μονάδων υλικού, οι οποίες είναι ικανές να αναπτύξουν αυξημένη θερμοκρασία κατά το χρόνο εκτέλεσης και συνεπώς να δημιουργήσουν πρόβλημα λειτουργίας και αξιοπιστίας στο ολοκληρωμένο κύκλωμα. Το πρόβλημα της αύξησης θερμοκρασίας γίνεται ακόμα μεγαλύτερο στις πρόσφατες αρχιτεκτονικές κλιμάκωσης, όπως π.χ. πολυπύρρηνοι σχεδιασμοί SoC, όπου προστίθενται πολλαπλές μονάδες υλικού στην ίδια επιφάνεια πυριτίου, με αποτέλεσμα την ελαχιστοποίηση της επιφάνειας απαγωγής θερμότητας. Ως εκ τούτου, μια από τις προκλήσεις που αντιμετωπίζουν σήμερα οι σχεδιαστές υλικού, είναι η ταυτοποίηση και ο χαρακτηρισμός των μονάδων υλικού, οι οποίες επηρεάζουν σε μεγαλύτερο βαθμό το προφίλ θερμότητας.

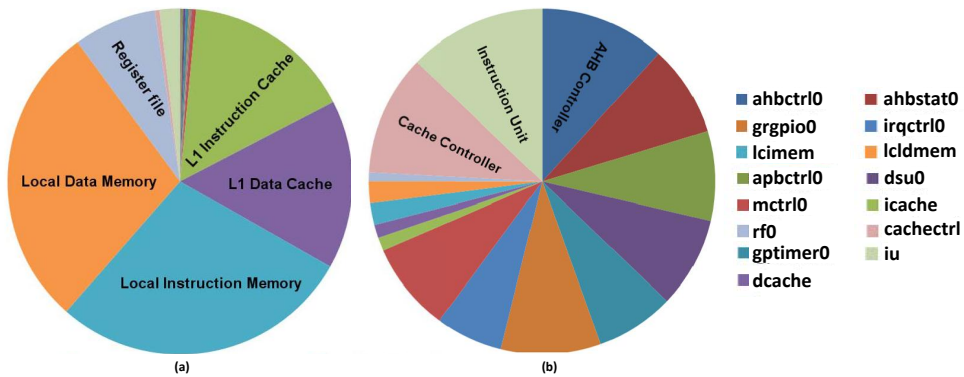
Στα πλαίσια αυτής της εργασίας, προτείνεται μία διαδικασία χαρακτηρισμού που αξιολογεί τόσο τη κατανάλωση ενέργειας, όσο και τη πυκνότητα ενέργειας των μονάδων υλικού μιας αρχιτεκτονικής. Σε πρώτη φάση γίνεται χαρακτηρισμός της κατανάλωσης ενέργειας. Το Σχήμα 5.1(a) απεικονίζει τη κατανάλωση ενέργειας για τις μονάδες του επεξεργαστή LEON3, κατά τη διάρκεια μιας τυπικής εκτέλεσης του συστήματος SDR.

Επιλέχθηκε ο συγκεκριμένος ενσωματωμένος επεξεργαστής καθώς χρησιμοποιείται ευρέως σε πλήθος προϊόντων εμπορικών ή/και ερευνητικών. Παράλληλα παρέχεται ως ελεύθερο λογισμικό και λογισμικό ανοικτού κώδικα σε γλώσσα VHDL, οπότε και αποτελεί ιδανικό σύστημα μελέτης για τη παρούσα εργασία, καθώς μπορεί να αναλυθεί η συμπεριφορά σε λεπτομερές επίπεδο μικροαρχιτεκτονικής. Ωστόσο, η μεθοδολογία που προτείνεται στη παρούσα εργασία μπορεί να εφαρμοσθεί σε οποιαδήποτε άλλη αρχιτεκτονική για την οποία υπάρχει επαρκής περιγραφή RTL σε επίπεδο μονάδων υλικού.

Σχετικά με τον επεξεργαστή LEON3, οι ενεργειακά περισσότερο απαιτητικές μονάδες είναι οι τοπικές μνήμες δεδομένων/εντολών, οι κρυφές μνήμες πρώτου επιπέδου δεδομένων/εντολών (L1 cache), καθώς και τα αρχεία καταχωρητών. Συγκεκριμένα, η μέση κατανάλωση ενέργειας σε αυτές τις μονάδες, σε σύγκριση με την ολική κατανάλωση ενέργειας, είναι 57%, 31% και 8%, αντίστοιχα.

Παρόλο που το Σχήμα 5.1(a) απεικονίζει μια πρώτη μετρική για τις μονάδες με αυξημένη κατανάλωση ενέργειας, ωστόσο δεν είναι αρκετή προκειμένου να εξαχθούν ορθά συμπεράσματα για την επιρροή τους στο συνολικό θερμικό προφίλ του ολοκληρωμένου κυκλώματος. Αυτό συμβαίνει διότι η μετρική της ενέργειας δε λαμβάνει υπ' όψιν την επιφάνεια πυριτίου της εκάστοτε μονάδας, η οποία αποτελεί το μέσο διάδοσης και κατ' επέκταση απαγωγής της θερμότητας. Συνεπώς είναι απαραίτητος ο χαρακτηρισμός μιας αντιπροσωπευτικής μετρικής για την αξιολόγηση της συνεισφοράς κάθε μονάδας στη συνολική θερμοκρασία του ολοκληρωμένου κυκλώματος.

Μία υποψήφια μετρική για το σκοπό αυτό είναι η πυκνότητα ενέργειας, η οποία ορίζεται από την αναλογία της κατανάλωσης ενέργειας για κάθε μονάδα υλικού σε σχέση με την επιφάνεια πυριτίου που καταλαμβάνεται από τη μονάδα

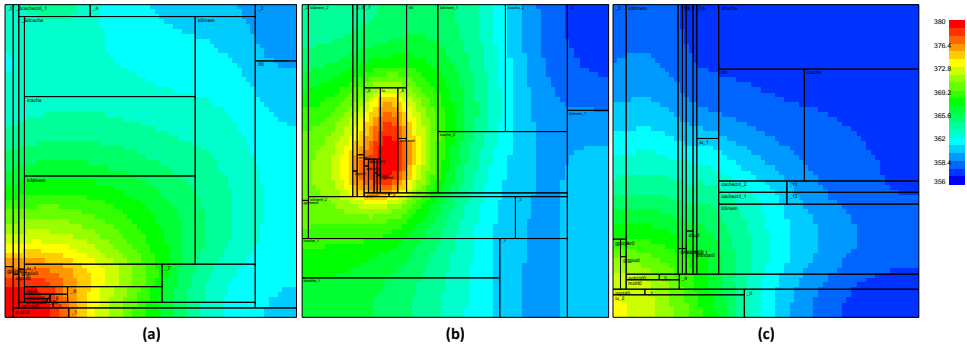


Σχήμα 3.2 Διάγραμμα κατανομής για (a) κατανάλωση ενέργειας και (b) πυκνότητα ενέργειας του επεξεργαστή LEON3.

αυτή. Το Σχήμα 5.1(b) απεικονίζει την αντίστοιχη πυκνότητα ενέργειας για τον επεξεργαστή LEON3. Είναι εμφανές ότι οι μονάδες με αυξημένη πυκνότητα ενέργειας δεν συμπίπτουν με τις μονάδες με αυξημένη κατανάλωση ενέργειας. Συγκεκριμένα, η μετρική της πυκνότητας ενέργειας χαρακτηρίζει ως υποψήφιες μονάδες για υψηλή συνεισφορά θερμότητας τον ελεγκτή AHB, τη μονάδα εκτέλεσης εντολών και τον ελεγκτή κρυφής μνήμης. Αυτές οι μονάδες συνεισφέρουν στη συνολική πυκνότητα ενέργειας περίπου κατά 12%, 13% και 11%, αντίστοιχα, ενώ οι υπόλοιπες πέντε μονάδες που χαρακτηρίζονται από υψηλή κατανάλωση ενέργειας αντιστοιχούν στο 5% της συνολικής πυκνότητας ενέργειας. Αυτό το αντικρουόμενο φαινόμενο συμβαίνει κυρίως διότι οι μονάδες με αυξημένη κατανάλωση ενέργειας καταλαμβάνουν επίσης αξιοσημείωτη επιφάνεια πυριτίου (περίπου το 91% της συνολικής επιφάνειας της αρχιτεκτονικής), η οποία με τη σειρά της οδηγεί σε σχεδόν αμελητέες τιμές πυκνότητας ενέργειας για τις μονάδες αυτές.

Ακολούθως δείχνεται ότι η μετρική της πυκνότητας ενέργειας είναι περισσότερο σημαντικό από την αντίστοιχη της κατανάλωσης ενέργειας. Τα Σχήματα 3.3(b) και 3.3(c) απεικονίζουν το θερμικό προφίλ όπως αναλύεται από το εργαλείο εξομοίωσης διάδοσης θερμοκρασίας Hotspot [19] για έναν επεξεργαστή LEON3 επεξεργαστή που εκτελεί εφαρμογές SDR (π.χ. ψηφιακά φίλτρα, κωδικοποίηση/αποκωδικοποίηση, κλπ). Σε αυτή τη μελέτη έχουν προστεθεί διπλά αντίγραφα για τις τρεις πιο κρίσιμες μονάδες υλικού, όσων αφορά την κατανάλωση ενέργειας και τη πυκνότητα ενέργειας αντίστοιχα, σύμφωνα με τη προηγούμενη ανάλυση. Η αντιγραφή των μονάδων υλικού πραγματοποιήθηκε μέσω της προτεινόμενης μεθοδολογίας που παρουσιάζεται σε αυτή την εργασία. Οι χωροθετήσεις (floorplans) πραγματοποιήθηκαν με το εργαλείο HotFloorplan [20].

Ως σημείο αναφοράς για τη μελέτη αυτή χρησιμοποιήθηκε το θερμικό προφίλ ενός SoC βασισμένο στον επεξεργαστή LEON3 για εφαρμογές SDR. Στη βασική αρχιτεκτονική αναφοράς δεν υφίστανται αντίγραφα μονάδων υλικού. Ο θερμι-



Σχήμα 3.3 Θερμικό προφίλ του επεξεργαστή LEON3: (a) δίχως τη χρήση μονάδων αντιγραφής, (b) με τη χρήση μονάδων αντιγραφής 2×local data/instruction memories, 2×L1 data/instruction caches, 2×register file, και (c) με τη χρήση μονάδων αντιγραφής 2×instruction unit, 2×cache controller, 2×AHB controller.

κός χάρτης, ο οποίος απεικονίζεται στο Σχήμα 3.3(a), παρουσιάζει ένα hotspot θερμοκρασίας στην επιφάνεια πυριτίου όπου έχουν τοποθετηθεί οι μονάδες με αυξημένη πυκνότητα ενέργειας (ελεγκτής διαύλου AHB, μονάδα εντολών και ελεγκτής κρυφής μνήμης). Αυτό το hotspot χαρακτηρίζεται από αυξημένη τιμή θερμοκρασίας περίπου έως 7%, σε σύγκριση με τη μέση θερμοκρασία on-chip.

Η επιλογή αντιγραφής των μονάδων υλικού που παρουσιάζουν αυξημένη κατανάλωση ενέργειας οδηγεί σε θερμικά κρίσιμες χωροθετήσεις, όπως φαίνεται και στο Σχήμα 3.3(b). Αυτό συμβαίνει διότι οι μονάδες υλικού που επιλέγονται για αντιγραφή χαρακτηρίζονται από χαμηλή πυκνότητα ενέργειας και συνεπώς δεν έχουν αξιοσημείωτη συνεισφορά στο θερμικό προφίλ. Ωστόσο η προσέγγιση αυτή επιδεικνύει ελαφρώς μειωμένες τιμές μέγιστης θερμοκρασίας, συγκρινόμενη με την αρχιτεκτονική αναφοράς, η οποία δεν περιέχει αντίγραφα μονάδων υλικού, εφόσον με αυτή τη προσέγγιση αυξάνεται η επιφάνεια πυριτίου του ολοκληρωμένου κυκλώματος (τα αντίγραφα μονάδων υλικού καταλαμβάνουν περίπου το 91% της συνολικής επιφάνειας πυριτίου του chip). Παράλληλα, οι μονάδες υλικού με αυξημένα ποσοστά πυκνότητας ενέργειας εξακολουθούν να συνεισφέρουν υψηλή θερμοκρασία στο θερμικό προφίλ, όπως απεικονίζεται στο Σχήμα 3.3(b).

Σημειώνεται ότι η διακύμανση θερμοκρασίας είναι σταθερή για όλους τους θερμικούς χάρτες που απεικονίζονται στα Σχήματα 3.3(a), 3.3(b) και 3.3(c), έτσι ώστε να φαίνεται εμφανώς ότι μόνο η αντιγραφή των μονάδων υλικού με αυξημένη πυκνότητα ενέργειας είναι ικανή να ελαττώσει τις υψηλές θερμοκρασίες στο θερμικό προφίλ του ολοκληρωμένου κυκλώματος.

3.2.4. Επέμβαση σε επίπεδο μικρο-αρχιτεκτονικής

Προκειμένου να υποστηριχθεί η προτεινόμενη επιλεκτική αντιγραφή μονάδων υλικού, πρέπει να τροποποιηθεί κατάλληλα η μικρο-αρχιτεκτονική του συστήματος. Γενικά, αυτή η ερευνητική εργασία επικεντρώνεται κυρίως στη μεθοδολογία που αναπτύσσεται προκειμένου να υποστηριχθούν αρχιτεκτονικές επιλεκτικές

αντιγραφής μονάδων υλικού. Σε αυτή την υποενότητα παρουσιάζονται οι μικρο-αρχιτεκτονικές τροποποιήσεις οι οποίες επιτρέπουν το σχεδιασμό των αρχιτεκτονικών επεξεργαστή με επιλεκτικά αντιγραμμένες μονάδες υλικού.

Κατά γενικό κανόνα, το μονοπάτι δεδομένων και ελέγχου της αρχικής αρχιτεκτονικής του επεξεργαστή τροποποιείται σε δύο κατευθύνσεις: (i) διασφάλιση αμοιβαίας αποκλειστικότητας επεξεργασίας μεταξύ των αντιγράφων μονάδων υλικού, και (ii) επίτρεψη διαχείρισης της λειτουργίας των αντιγράφων μονάδων υλικού, κατά το χρόνο εκτέλεσης, ανάλογα με τη θερμική κατάσταση του επεξεργαστή. Η παρούσα εργασία εστιάζει την ανάλυση της μικρο-αρχιτεκτονικής RISC του ενσωματωμένου επεξεργαστή LEON3 [18]. Το αρχιτεκτονικό πρότυπο ενός συστήματος SDR το οποίο βασίζεται στον επεξεργαστή LEON3 απεικονίζονται στο Σχήμα 5.3.

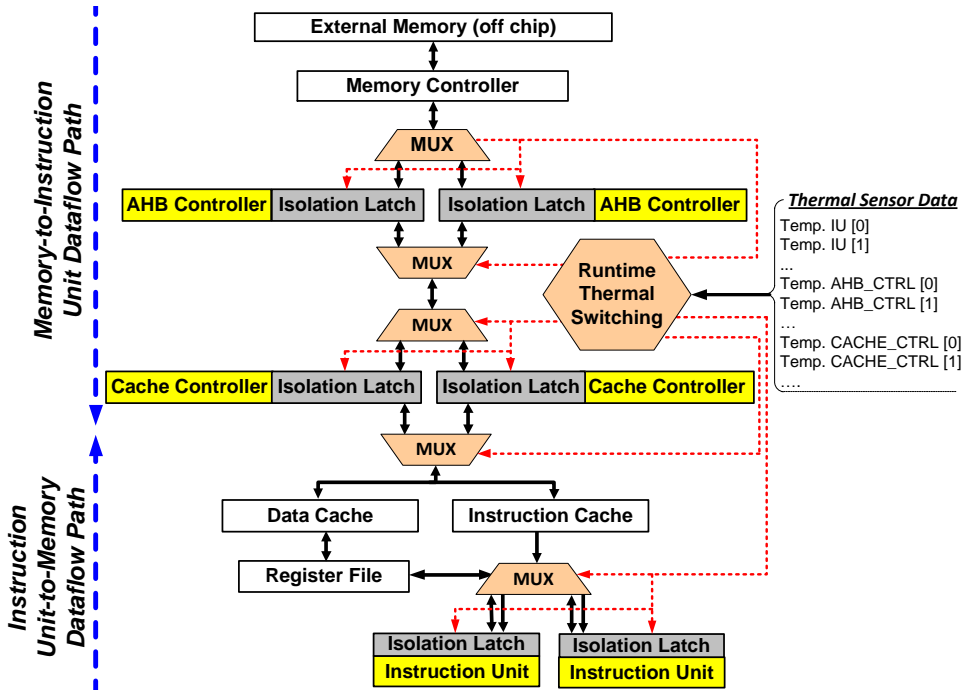
Η προτεινόμενη μεθοδολογία στοχεύει στην κατά το δυνατόν μη-παρεμβατική υλοποίηση των μηχανισμών σχεδίασης των αντιγράφων μονάδων υλικού, ώστε να αποφευχθούν μεγάλες επιβαρύνσεις στο χρόνο επεξεργασίας και στην επιφάνεια πυριτίου. Για το σκοπό αυτό, εφαρμόζεται επιλεκτική αντιγραφή κατά μία αδρομερή προσέγγιση, σύμφωνα με την οποία η επιλογή ιεραρχίας για την εφαρμογή αντιγραφής ορίζεται στο επίπεδο λειτουργικότητας των μονάδων υλικού. Επιπλέον, αποφεύγετε η αντιγραφή των μονάδων μνήμης, δηλαδή του αρχείου καταχωρητών, οι κρυφές μνήμης δεδομένων/εντολών κλπ., καθώς αυτή η επιλογή αντιγραφής απαιτεί κατάλληλους μηχανισμούς ελέγχου για την εξασφάλιση συνεκτικότητας δεδομένων (data coherency) μεταξύ των αντιγράφων μονάδων υλικού.

Παρά το γεγονός ότι, η επιλεκτική αντιγραφή μονάδων υλικού σε λεπτομερές επίπεδο ανάλυσης είναι μια έγκυρη επιλογή, η παρούσα εργασία αποδεικνύει ότι η αδρομερής αντιγραφή μπορεί να επιτύχει σημαντική μείωση της θερμοκρασίας και μείωση των hotspot θερμοκρασίας, με μικρή επιβάρυνση στην απόδοση του σχεδιασμού. Η προτεινόμενη προσέγγιση εξασφαλίζει επίσης βελτίωση σε φαινόμενα γήρανσης του ολοκληρωμένου κυκλώματος. Επίσης το προτεινόμενο περιβάλλον δεν είναι περιοριστικό για συγκεκριμένα αρχιτεκτονικά πρότυπα. Όπως φαίνεται στο Σχήμα 5.1(b), εκτός από το αρχείο καταχωρητών που αποκλείεται από την αντιγραφή, το μεγαλύτερο ποσοστό της πυκνότητας ενέργειας αντιστοιχεί στις μονάδες που επιλέγονται για αντιγραφή, όπως π.χ. η μονάδα εντολών (IU), ο ελεγκτής κρυφής μνήμης (CACHE_CTRL) και ο ελεγκτής του διαύλου AHB (AHB_CTRL).

Σε συνάφεια με την προηγούμενη ανάλυση, προτείνεται η υιοθέτηση της μικρο-αρχιτεκτονικής επέκτασης, η οποία απεικονίζεται στο Σχήμα 3.4. Στο συγκεκριμένο παράδειγμα, έχει επιλεχθεί αντιγραφή δύο μονάδων υλικού. Ωστόσο, ο αριθμός αντιγραφής είναι παραμετρικός στη προτεινόμενη μεθοδολογία και η ακριβής τιμή του καθορίζεται από τους περιορισμούς σχεδίασης της εκάστοτε υλοποίησης, π.χ. μέγιστη επιφάνεια πυριτίου. Σε κάθε μονάδα που αντιγράφεται προστίθενται μανδαλωτές απομόνωσης λειτουργιών προκειμένου να εκμηδενίσουν τη κατανάλωση δυναμικής ενέργειας κατά τα χρονικά πλαίσια εκτέλεσης στα οποία οι μονάδες αντιγραφής δεν είναι λειτουργικές. Η ιδέα των μανδαλωτών απομόνωσης λειτουργιών βασίζεται στην εργασία [21]. Επίσης, σε ορ-

θογώνια προσέγγιση, μπορούν να εφαρμοσθούν κλασικές τεχνικές μείωσης της δυναμικής ενέργειας, όπως π.χ. η τεχνική power-gating.

Στο μονοπάτι δεδομένων κάθε μονάδας αντιγραφής προστίθενται ζεύγη πολυπλεξίας και απο-πολυπλεξίας. Ειδικότερα, οι εισόδοι του εκάστοτε αντιγράφου οδηγούνται από έναν αποπλέκτη, ο οποίος συνδέει τα δεδομένα εισόδου στο ενεργές αντίγραφο για δεδομένη χρονική στιγμή. Αντίστοιχα, τα σήματα εξόδου από κάθε αντίγραφο καταλήγουν σε έναν πολυπλέκτη προκειμένου να διαδοθούν στο επόμενο λογικό επίπεδο του μονοπατιού δεδομένων. Διακρίνονται δύο μονοπάτια ροής δεδομένων στον ενσωματωμένο επεξεργαστή, το μονοπάτι μνήμης-σε-μονάδα εντολών (memory-to-instruction unit) και το μονοπάτι μονάδας εντολών-σε-μνήμη (instruction unit-to-memory). Το Σχήμα 3.4 απεικονίζει το συνδυασμό αυτών των δύο μονοπατιών ροής δεδομένων, διασχίζοντας την αρχιτεκτονική σε προσέγγιση top-down (memory-to-instruction unit) ή bottom-up (instruction unit-to-memory).



Σχήμα 3.4 Προτεινόμενες μικρο-αρχιτεκτονικές τροποποιήσεις.

Η αρχιτεκτονική του επεξεργαστή LEON3 ενισχύεται με έναν ελεγκτή θερμοκρασίας πραγματικού χρόνου, ο οποίος είναι υπεύθυνος για τη διανομή του φόρτου εργασίας στα διαθέσιμα αντίγραφα μονάδων υλικού κατά το χρόνο εκτέλεσης. Η κατανομή του φόρτου εργασίας πραγματοποιείται μέσω της κατάλληλης προώθησης σημάτων ελέγχου στους πολυπλέκτες και τους τελεστές λογικής απομόνωσης. Ο ελεγκτής θερμοκρασίας λειτουργεί ανεξάρτητα από την λογική

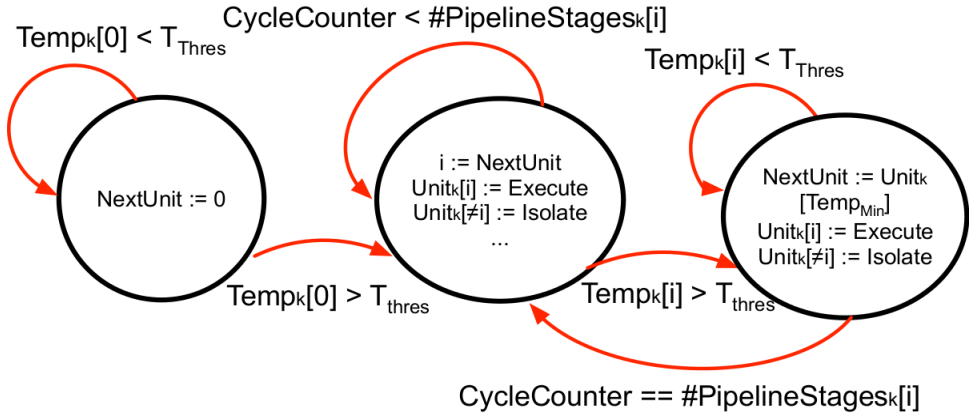
ελέγχου της υπόλοιπης αρχιτεκτονικής του επεξεργαστή LEON3. Συγκεκριμένα, Ο ελεγκτής αποφασίζει ποιο αντίγραφο μονάδας υλικού θα είναι ενεργό σε δεδομένη χρονική στιγμή, σύμφωνα με το θερμικό προφίλ του επεξεργαστή. Βασική προϋπόθεση αυτής της προσέγγισης, αποτελεί το γεγονός ότι στην εκάστοτε αρχιτεκτονική μελέτης, υφίστανται διαθέσιμα δεδομένα θερμοκρασίας, κατά το χρόνο εκτέλεσης, π.χ. μέσω θερμικών αισθητήρων.

Ο ελεγκτής είναι σχεδιασμένος ώστε να ενεργεί προληπτικά. Πρακτικά, ο ελεγκτής αλλάζει κατάσταση κάθε φορά που αναγνωρίζεται ένα ανώτερο όριο θερμοκρασίας, T_{thres} . Ακολουθείτε η προσέγγιση της εργασίας [22], κατά την οποία το όριο θερμοκρασίας για την αλλαγή κατάστασης του ελεγκτή ρυθμίζεται σε χαμηλότερη τιμή θερμοκρασίας από εκείνη που εκτιμάται ότι θα φθάσει κατά το χρόνο εκτέλεσης το ολοκληρωμένο κύκλωμα. Με αυτό τον τρόπο διασφαλίζεται ορθή λειτουργία κατά τη διάρκεια εκτέλεσης.

Ο ελεγκτής θερμοκρασίας αντιδρά στις αναγνώσεις θερμοκρασίας από τους αισθητήρες, οι οποίοι βρίσκονται τοποθετημένοι σε διάφορα σημεία του ολοκληρωμένου κυκλώματος, καλύπτοντας το μεγαλύτερο μέρος της επιφάνειάς του. Η θερμοκρασία $Temp_k[i]$ αντιστοιχεί στη μετρική θερμοκρασίας που διαβάστηκε από έναν αισθητήρα, για το αντίγραφο της μονάδας i , όπου $i \in \{0, Max_{replicas}\}$ και η μονάδα αυτή θα είναι ενεργή στον επόμενο κύκλο ρολογιού, σύμφωνα με τον τύπο της μονάδας k , $k \in \{U, CACHE_CTRL, AHB_CTRL\}$. Η παράμετρος $Unit_k[i]$ είναι ένας πίνακας καταχωρητών εύρους 1-bit. Κάθε ένας από αυτούς τους καταχωρητές μανδαλώνει το σήμα ενεργοποίησης/απενεργοποίησης - *Execute/Isolate* για το αντίγραφο i της μονάδας τύπου k .

Οι μεταβάσεις κατάστασης του ελεγκτή θερμοκρασίας για ένα συγκεκριμένο τύπο μονάδας υλικού, π.χ. τη μονάδα εντολών, απεικονίζονται στο Σχήμα 5.7. Αντίστοιχη λογική ελέγχου προστίθεται σε κάθε τύπο αντιγράφου μονάδας υλικού k , καθένas εκ των οποίων διαχειρίζεται μοναδικά και ανεξάρτητα από τους ελεγκτές άλλων τύπων. Με αυτό το τρόπο διασφαλίζεται μια αρθρωτή αρχιτεκτονική από μονοπάτια ελέγχου, όπως αυτά που απεικονίζονται στο Σχήμα 5.7, ενώ κάθε ένα από αυτά αντιστοιχεί σε συγκεκριμένο τύπο μονάδων υλικού. Εάν η θερμοκρασία ανίχνευσης είναι μικρότερη από το ανώτερο όριο θερμοκρασίας, $Temp_k[i] < T_{thres}$, ο ελεγκτής παραμένει στην ίδια κατάσταση, ενώ κάθε φορά που η θερμοκρασία ανίχνευσης του ενεργού αντιγράφου $Unit_k[i]$ ξεπερνά το μέγιστο όριο θερμοκρασίας, $Temp_k[i] < T_{thres}$, ο ελεγκτής εισέρχεται σε μία ενδιάμεση κατάσταση. Σε αυτήν, ο ελεγκτής διατηρεί την προηγούμενη ρύθμισή του για έναν αριθμό από κύκλους που αντιστοιχούν στην εκτέλεση των βαθμίδων σωλήνωσης, $\#PipelineStages_k[i]$, της συγκεκριμένης μονάδας $Unit_k[i]$. Μέσω αυτού του μηχανισμού, ο ελεγκτής περιμένει τη πλήρη εκτέλεση της σωλήνωσης της μονάδας υλικού, προκειμένου να αποφευχθούν εσφαλμένες αναγνώσεις/εγγραφές δεδομένων (data hazards). Επιπλέον, κατά τη διάρκεια αυτού του ενδιάμεσου σταδίου, αναγνωρίζεται η μονάδα υλικού με τη χαμηλότερη θερμοκρασία ανάμεσα στα διαθέσιμα αντίγραφα του ίδιου τύπου, χρησιμοποιώντας δεδομένα από τους αισθητήρες, $NextUnit_k := Unit_k[Temp_{Min}]$. Όταν η σωλήνωση ολοκληρωθεί, ο ελεγκτής προωθεί σήματα ενεργοποίησης στους κατάλληλους πολυπλέκτες και τελεστές απομόνωσης, ενώ παράλληλα επαναπρογραμματίζει

το μονοπάτι εντολών και δεδομένων της βασικής αρχιτεκτονικής, έτσι ώστε να διέλθει από το νέο επιλεγμένο αντίγραφο υλικού.

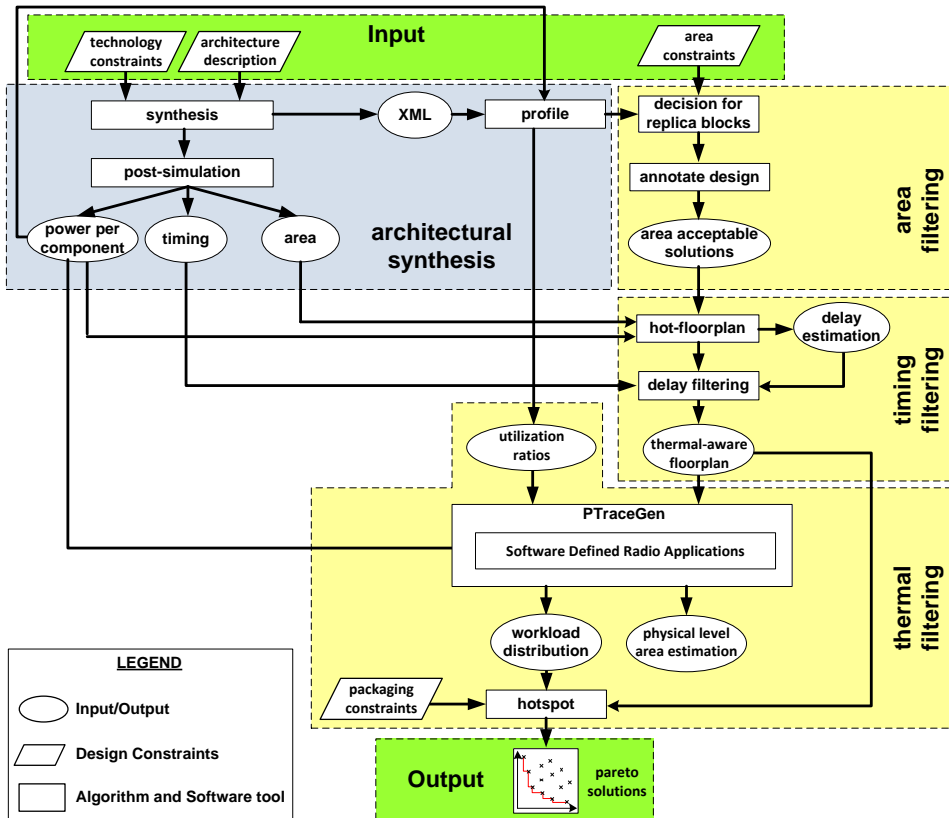


Σχήμα 3.5 Ο ελεγκτής θερμοκρασίας πραγματικού χρόνου για κάθε αντίγραφο μονάδας υλικού.

3.2.5. Προτεινόμενη μεθοδολογία

Αυτή η υποενοήτητα περιγράφει αναλυτικά την προτεινόμενη μεθοδολογία για τη μείωση των hotspots θερμοκρασίας μέσω επιλεκτικής αντιγραφής για τις μονάδες υλικού μίας αρχιτεκτονικής που παρουσιάζουν αυξημένη πυκνότητα ενέργειας. Σημειώνεται ότι αυτή η μεθοδολογία δεν στοχεύει στην επανασχεδιασμό ολόκληρης της μικρο-αρχιτεκτονικής, αλλά επικεντρώνεται μόνο σε συγκεκριμένες μονάδες της. Οι στόχοι αυτής της μεθοδολογίας είναι: (i) η παροχή μιας προληπτικής και δυναμικής διαχείρισης θερμοκρασίας μέσω τροποποίησης της μικρο-αρχιτεκτονικής, και (ii) η υποστήριξη ταχείας διερεύνησης/αξιολόγησης εναλλακτικών αρχιτεκτονικών επιλογών κατά τη διάρκεια της θερμικής καταπόνησης του ολοκληρωμένου κυκλώματος. Σημειώνεται ότι οι αρχιτεκτονικές διαμορφώσεις που χαρακτηρίζονται από την προτεινόμενη μεθοδολογία είναι ανεξάρτητες από τη ροή σχεδίασης και δεν επηρεάζουν τα υφιστάμενα τυπικά εργαλεία σχεδίασης, ενώ οι τελικοί χρήστες (π.χ. προγραμματιστές) δεν χρειάζεται να ασχοληθούν με θέματα χρονοπρογραμματισμού των εργασιών στα διαθέσιμα αντίγραφα, καθώς αυτή η διαδικασία εξασφαλίζεται αυτόματα σε επίπεδο υλικού.

Η προτεινόμενη μεθοδολογία απεικονίζεται στο Σχήμα 3.6. Είσοδοι της μεθοδολογίας αυτής αποτελούν η περιγραφή της αρχιτεκτονικής σε γλώσσα περιγραφής υλικού VHDL/Verilog, οι τεχνολογικοί περιορισμοί σχετικά με την επιλεγμένη τεχνολογία CMOS, καθώς και οι συνθήκες λειτουργίας. Στις τελευταίες εντάσσεται η μέγιστη ανεκτή επιφάνεια πυριτίου που μπορεί να δεσμευτεί για τα αντίγραφα μονάδων υλικού. Παρόλο που η μεθοδολογία εστιάζει σε διαδικασίες ανάπτυξης κατά το χρόνο σχεδίασης, οι επικείμενες ενότητες αξιολογούν τις διακυμάνσεις θερμοκρασίας κατά το χρόνο εκτέλεσης των εφαρμογών.



Σχήμα 3.6 Η προτεινόμενη μεθοδολογία για διαχείριση της θερμοκρασίας μέσω αντιγράφων μονάδων υλικού.

3.2.5.1. Σύνθεση αρχιτεκτονικής

Οι είσοδοι της προτεινόμενης μεθοδολογία είναι η περιγραφή της αρχιτεκτονικής σε γλώσσα περιγραφής υλικού VHDL/Verilog, καθώς και η επιλεγείσα τεχνολογία CMOS. Αρχικά, ο σχεδιασμός συντίθεται με το εργαλείο Synopsys Design Compiler [23] και στη συνέχεια πραγματοποιείται εξομοίωση με το εργαλείο Cadence Incisive Simulator [24], προκειμένου να εξαχθούν σχεδιαστικές μετρήσεις σχετικά με την αρχιτεκτονική. Μεταξύ άλλων, αυτές οι μετρήσεις περιλαμβάνουν την επιφάνεια πυριτίου που καταλαμβάνεται από κάθε λογική μονάδα του σχεδιασμού, καθώς και την κατανάλωση ενέργειας με βάση τη δραστηριότητα της εκάστοτε μονάδας κατά το χρόνο εκτέλεσης. Η κατανάλωση ενέργειας εκτιμάται με το εργαλείο Synopsys PrimeTime PX [25]. Παράλληλα με αυτή τη διαδικασία χαρακτηρισμού, χρησιμοποιείται το μοντέλο Elmore [26], προκειμένου να εξαχθεί μία πρώτη εκτίμηση σχετικά με τις καθυστερήσεις συνδέσεων και τη μέγιστη συχνότητα λειτουργίας του κυκλώματος. Σημειώνεται ότι για την εκτίμηση τόσο της καθυστέρησης, όσο και της κατανάλωσης ενέργειας, χρησιμοποιούνται σενάρια χειρότερης περίπτωσης σχετικά με το φόρτος εργασίας, προκειμένου να διασφαλιστεί ότι η ανάλυση της θερμικής καταπόνησης δεν θα παραβιαστεί κατά το χρόνο εκτέλεσης.

Η έξοδος από το στάδιο της σύνθεσης κωδικοποιείται σε περιγραφή XML προκειμένου να διαχειριστεί κατάλληλα από τα προτεινόμενα εργαλεία της μεθοδολογίας. Το επίπεδο λεπτομέρειας της περιγραφής του συστήματος σε αυτή τη περιγραφή XML είναι παραμετρικό, δεδομένου ότι η λεπτομερής ανάλυση αντιστοιχεί σε ακριβή θερμική ανάλυση, αλλά επιφέρει αυξημένο υπολογιστικό κόστος. Αντίθετα, η αδρομερής ανάλυση επιτυγχάνεται με μικρότερο υπολογιστικό κόστος, αλλά είναι λιγότερο αποτελεσματική στην αντιμετώπιση των hotspots. Επίσης η λεπτομερής ανάλυση επιφέρει μεγαλύτερη αύξηση επιπρόσθετου υλικού για την υλοποίηση των απαραίτητων πολυπλεκτών. Στη συνέχεια πραγματοποιείται εξομοίωση και χαρακτηρισμός του συστήματος έτσι ώστε να υπολογισθεί η πυκνότητα ενέργειας των μονάδων που περιγράφονται στο προαναφερθέν αρχείο XML.

3.2.5.2. Επιλογή βάση επιφάνειας πυριτίου

Σε αυτό το στάδιο της μεθοδολογίας, χαρακτηρίζονται οι μονάδες υλικού της αρχιτεκτονικής, οι οποίες επιλέγονται για αντιγραφή. Συγκεκριμένα εκτιμάται η πυκνότητα ισχύος για τις υποψήφιας προς αντιγραφή μονάδες. Σημειώνεται ότι ο συνολικός αριθμός των μπλοκ αντιγραφής περιορίζεται από το μέγιστη επιφάνεια πυριτίου, η οποία και αποτελεί προδιαγραφή της εκάστοτε αρχιτεκτονικής καθώς και επιλογή του σχεδιαστή. Οι μονάδες υλικού του σχεδιασμού κατατάσσονται κατά φθίνουσα σειρά με βάση τη μετρική πυκνότητας ισχύος. Στη συνέχεια αξιολογούνται οι αρχιτεκτονικές που περιέχουν όλα τους πιθανούς σχεδιασμούς αντιγράφων υλικού. Βασικός περιορισμός σε αυτή τη διαδικασία αποτελεί η μέγιστη επιφάνεια πυριτίου.

Ακολούθως, ορίζεται μία επιπλέον παράμετρος προκειμένου να χαρακτηρίσει το μέγιστο αριθμό επιτρεπτών αντιγράφων μονάδων υλικού. Όσο περισσότερα αντίγραφα προστίθενται, τόσο καλύτερη θερμική διαχείριση μπορεί να επιτευ-

χθεί. Ωστόσο η αύξηση αντιγράφων, επίσης επιφέρει ανάλογη αύξηση επιφάνειας πυριτίου καθώς και δυσκολία επίτευξης υψηλής συχνότητας λειτουργίας και σταθερής καθυστέρησης ανάμεσα στα μονοπάτια δεδομένων των αντιγράφων. Στη παρούσα έρευνα, ύστερα από ένα πρώτο στάδιο χαρακτηρισμού της δέσμευσης επιφάνειας πυριτίου για τον επεξεργαστή LEON3, διαπιστώθηκε ότι ο βέλτιστος μέγιστος αριθμός αντιγραφής αντιστοιχεί σε πέντε μονάδες. Αυτή η επιλογή βασίστηκε στο γεγονός ότι οι αρχιτεκτονικές με περισσότερες από πέντε μονάδες αντιγραφής δεν οδηγούν σε περαιτέρω μείωση της θερμοκρασίας. Ωστόσο, ανάλογα με τις ιδιαιτερότητες της εκάστοτε μονάδας αντιγραφής, ο αριθμός αυτός ενδέχεται να είναι μικρότερος, προκειμένου να μην επηρεάζεται η απόδοση του συνολικού κυκλώματος, π.χ. μείωση της μέγιστης συχνότητας λειτουργίας λόγω εκτεταμένης τοποθέτησης των αντιγραμμένων μονάδων.

Υστερα από τον καθορισμό των μονάδων αντιγραφής, καθώς και του αριθμού αντιγραφής αυτών, ο σχεδιασμός εισέρχεται σε ένα εργαλείο, το οποίο εκτελεί αυτόματα την διαδικασία της αντιγραφής. Σε αυτό το στάδιο προστίθενται επίσης οι πολυπλέκτες/αποπλέκτες των σημάτων, καθώς και οι απομονωτές των αντιγράφων σε περιγραφή XML. Επιπλέον, κατά τη διάρκεια αυτής της διαδικασίας επαληθεύεται η συνδεσιμότητα μεταξύ των αντιγράφων μονάδων υλικού και προστίθενται οι ελεγκτές θερμοκρασίας πραγματικού χρόνου. Σε αντίθεση με παρόμοιες προσεγγίσεις στη σχετική βιβλιογραφία [9] [7] [6] [8] [10], η προτεινόμενη προσέγγιση δεν μεταβάλλει τη λειτουργικότητα της αρχιτεκτονικής.

Το αποτέλεσμα από αυτό το βήμα είναι η συγκέντρωση ενός συνόλου από υποψήφιες αρχιτεκτονικές που πληρούν τους περιορισμούς επιφάνειας πυριτίου. Αυτό το κριτήριο κατακερματίζει το χώρο λύσεων, αφαιρώντας αυτές που οδηγούν σε μη αποδεκτή επιφάνεια πυριτίου, λόγω μεγάλου αριθμού αντιγράφων υλικού. Επίσης, θέτοντας ένα άνω όριο στην μέγιστη επιφάνεια πυριτίου που μπορεί να προστεθεί σε ένα σχεδιασμό, είναι δυνατή η διερεύνηση και η αξιολόγηση διαφορετικών αρχιτεκτονικών λύσεων. Στη παρούσα εργασία τίθεται ένα όριο της τάξης του 35%, διότι διαφορετικά η μεθοδολογία μας οδηγεί σε μη ανεκτή - σχεδιαστικά και οικονομικά - επιφάνεια πυριτίου για τον επεξεργαστή LEON3.

Το κριτήριο που χρησιμοποιείται σε αυτή τη προσέγγιση επιτρέπει στις μονάδες με αυξημένη πυκνότητα ενέργειας να αντιγραφούν περισσότερες φορές σε σύγκριση με τις μονάδες με αντίστοιχες μικρότερες τιμές. Η φυσική σημασία αυτού του αποτελέσματος αντιστοιχεί στο γεγονός ότι η επιφάνεια πυριτίου που καταλαμβάνεται από μονάδες υψηλής πυκνότητας ενέργειας είναι συνήθως περιορισμένη, και ως εκ τούτου, περισσότερες σε πλήθος μονάδες μπορούν να αντιγραφούν. Δεδομένου ότι μόνο ένα από τα αντίγραφα είναι ενεργό ανά πάσα στιγμή, όπως περιγράφηκε στην υποενότητα 3.2.4, η κατάλληλη αντιγραφή μονάδων υλικού με αυξημένη πυκνότητα ενέργειας μπορεί να οδηγήσει σε (i) ελαχιστοποίηση της μέγιστης και μέσης θερμοκρασίας και (ii) ομοιόμορφη διανομή της θερμοκρασίας στην επιφάνεια του ολοκληρωμένου κυκλώματος.

3.2.5.3. Επιλογή βάση χρονικής καθυστέρησης

Το επόμενο στάδιο περιλαμβάνει την αξιολόγηση της αποτελεσματικότητας των εναλλακτικών αρχιτεκτονικών αναφορικά προς τη χρονική καθυστέρηση των

σημάτων ελέγχου και δεδομένων και κατ'επέκταση τη μέγιστη επιτευχθείσα συχνότητα λειτουργίας. Οι λύσεις που έχουν συσταθεί στο προηγούμενο στάδιο, χωροθετούνται με τη χρήση του εργαλείου χωροθέτησης *Hotfloorplan*, το οποίο βελτιστοποιεί τη θερμοκρασία [20].

Ο στόχος της βελτιστοποίησης κατά την διάρκεια αυτής της διαδικασίας είναι η μείωση της θερμοκρασίας σε σχέση με τον περιορισμό χρονισμού του κυκλώματος. Η ανακούφιση της θερμικής καταπόνησης επιτυγχάνεται με τη διάδοση όσο το δυνατόν περισσότερο των μονάδων υλικού που συμβάλλουν περισσότερο στην αύξηση της on-chip θερμοκρασίας, δηλαδή των μονάδων με αυξημένη πυκνότητα ενέργειας. Ομοίως, με την ελαχιστοποίηση της περιμέτρου του πλαισίου που περιβάλλει όλα τα αντίγραφα μονάδων υλικού που συνδέονται με κάποιο δίαυλο επικοινωνίας, είναι δυνατόν να βελτιωθεί η καθυστέρηση αυτού του διαύλου. Ως εκ τούτου, οι μονάδες που συνδέονται μέσω κοινών διαύλων πρέπει να χωροθετηθούν σε σχετικά κοντινές θέσεις στην επιφάνεια του ολοκληρωμένου κυκλώματος.

Στη παρούσα εργασία, όλες οι μονάδες έχουν μεταβλητή αναλογία διαστάσεων ("soft" blocks) και επομένως προσφέρουν μεγάλη ευελιξία χωροθέτησης στο εργαλείο *Hotfloorplan*. Ωστόσο το μέγεθος της επιφάνειάς τους είναι σταθερό.

Στη συνέχεια οι προκύπτουσες λύσεις αξιολογούνται ως προς τη χρονική καθυστέρηση, σε σύγκριση με την εκτίμηση καθυστέρησης που ανακτάται κατά τη προσομοίωση μετά το στάδιο της σύνθεσης. Σε αυτό το στάδιο χρησιμοποιείται το μοντέλο καθυστέρησης Elmore [26]. Δεδομένου ότι ο κύριος στόχος της παρούσας εργασίας είναι η διερεύνηση λύσεων με βελτιστοποιημένη θερμοκρασία, είναι αναμενόμενη μία μείωση της απόδοσης του συστήματος αναφορικά προς τη συχνότητα επεξεργασίας. Για τη παρούσα μελέτη, η επιβάρυνση χρονισμού τίθεται σε ποσοστό έως 14%, προκειμένου να αποφευχθούν μεγάλες επιβαρύνσεις στις επί μέρους καθυστέρησης των μονάδων του κυκλώματος. Η έξοδος αυτού του βήματος είναι αρχιτεκτονικές λύσεις που ανταποκρίνονται στους περιορισμούς χρονισμού και καθυστέρησης.

3.2.5.4. Επιλογή βάση θερμικού προφίλ

Τέλος, στο τελευταίο στάδιο της προτεινόμενης μεθοδολογίας, οι διαφορετικές αρχιτεκτονικές λύσεις αξιολογούνται ως προς τη θερμική καταπόνηση. Το βήμα αυτό αυτοματοποιείται μέσω ενός νέου εργαλείου, που ονομάζεται *PTraceGen*, το οποίο παράγει κατάλληλα ίχνη κατανάλωσης ενέργειας στην υπό-εξέταση αρχιτεκτονική, σύμφωνα με τη στατιστική συμπεριφορά της εκάστοτε εφαρμογής. Συγκεκριμένα, το εργαλείο *PTraceGen* παρέχει λεπτομερείς πληροφορίες σχετικά με τις μονάδες της αρχιτεκτονικής και οι οποίες περιγράφουν i) το ποσοστό του χρόνου κατά τον οποίο μια μονάδα είναι ενεργή, ii) το ποσοστό του χρόνου κατά τον οποίο μια μονάδα είναι ανενεργή, και iii) τη δραστηριότητα των σημάτων και των λογικών πόρων που ανήκουν σε αυτές τις μονάδες. Οι δύο πρώτες μετρικές (ποσοστά ενεργού/ανενεργού χρόνου) αναφέρονται στον αριθμό των κύκλων εκτέλεσης για μία συγκεκριμένη χρονική διάρκεια εκτέλεσης, ενώ η τρίτη μετρική αντιστοιχεί στον αριθμό των μεταβάσεων μεταξύ λειτουρ-

γικών καταστάσεων (ενεργή/ανενεργή) κατά τη διάρκεια του χρόνου εκτέλεσης του σχεδιασμού.

Συγκεκριμένα, ο στατιστικός χαρακτηρισμός πραγματοποιείται με τον υπολογισμό των στατιστικών μέσων τιμών των στοιχειωδών πράξεων που εκτελούνται από τον επεξεργαστή, δηλαδή ο αριθμός και το είδος των εντολών ALU, οι προσβάσεις μνήμης, οι επιτυχίες/αποτυχίες κρυφής μνήμης κ.λπ. Δημιουργούνται ίχνη κατανάλωσης ενέργειας για όλα τα αντίγραφα μονάδων υλικού. Τα ίχνη αυτά περιγράφουν τη δραστηριότητα για όλες τις μονάδες της αρχιτεκτονικής κατά τη διάρκεια συγκεκριμένου χρονικού διαστήματος. Χρησιμοποιώντας αυτά τα στατιστικά στοιχεία, εξάγεται η αναλογία χρησιμοποίησης ανά μονάδα υλικού σε κάθε χρονική στιγμή. Αυτές οι αναλογίες συσχετίζονται με ακριβείς μετρήσεις κατανάλωσης ενέργειας που έχουν αποτυπωθεί μετά στην εξομοίωση, μετά το στάδιο της σύνθεσης. Στη συνέχεια, το εργαλείο αυτού του σταδίου εξάγει τις λύσεις Pareto σε σχέση με διάφορους σχεδιαστικούς στόχους και μετρικές αντίστοιχα, όπως η πυκνότητα ισχύος του ολοκληρωμένου κυκλώματος, η χρονική καθυστέρηση, η επιφάνεια πυριτίου, η μέγιστη θερμοκρασία, η θερμική διακύμανση, κλπ.

Η έξοδος του εργαλείου *PTraceGen* είναι ένα σύνολο από κατανομές εργασιών των προαναφερθέντων εφαρμογών, καθώς και από ακριβή εκτιμήσεων επιφάνειας πυριτίου. Συγκεκριμένα, η κατανομή των εργασιών επιτυγχάνεται με τον κατάλληλο χρονοπρογραμματισμό αυτών, βάση της πληροφορίας κατανάλωσης ενέργειας κάθε μονάδας υλικού, όπως αυτή έχει προκύψει κατά τη προσομοίωσης της αρχιτεκτονικής μετά το στάδιο της σύνθεσης. Η προσέγγιση αυτή εγγυάται ότι οι μονάδες υλικού καταναλώνουν ενέργεια μόνο κατά τις χρονικές περιόδους όπου είναι ενεργοποιημένες.

Οι χρονικές περίοδοι κατά τις οποίες ενεργοποιούνται και απενεργοποιούνται τα αντίγραφα μονάδων υλικού καθορίζονται από την προσομοίωση της εφαρμογής. Συγκεκριμένα, χαρακτηρίζονται με ακρίβεια οι χρονοθυρίδες λειτουργίας/αδράνειας για κάθε μονάδα. Αυτή η πληροφορία τροφοδοτείται ως είσοδος στο εργαλείο *PTraceGen* προκειμένου να αναδιανείμει το συνολικό φόρτο εργασίας που απαιτεί μία εφαρμογή στα διάφορα αντίγραφα μονάδων υλικού. Αυτή η ανάθεση επιτυγχάνεται με τον καθορισμό σε όλες τα αντίγραφα μονάδων υλικού χρονοθυρίδων λειτουργίας και αδράνειας.

Η κατανομή των εργασιών στα διαθέσιμα αντίγραφα μονάδων υλικού αποτελεί είσοδο του εργαλείου *hotspot* [19] προκειμένου να υπολογιστεί το θερμικό προφίλ της αρχιτεκτονικής SDR. Προκειμένου να εξασφαλιστεί αυξημένη ακρίβεια, χρησιμοποιείται αρχική κατάσταση θερμοκρασίας της κάθε μονάδας υλικού, έτσι ώστε το θερμικό προφίλ να εξομοιώνεται σε πραγματικές συνθήκες μόνιμης λειτουργίας και όχι κατά την εκκίνηση εκτέλεσης της εφαρμογής, όπου η θερμοκρασία του ολοκληρωμένου κυκλώματος είναι χαμηλή. Οι βασικές παράμετροι του μοντέλου θερμικής εξομοίωσης συνοψίζονται στον Πίνακα 3.1.

Με βάση το θερμικό προφίλ, όπως αυτό προκύπτει από την εξομοίωση με το εργαλείο *hotspot*, καθίσταται δυνατή η αξιολόγηση της αρχιτεκτονικής σε διαφορετικά κριτήρια σχεδιασμού σε σχέση με την on-chip θερμοκρασία. Στα πλαίσια της παρούσας εργασίας, οι λύσεις που δεν πληρούν τους επιλεγμένους θερμι-

Πίνακας 3.1 Παράμετροι του μοντέλου θερμικής εξομοίωσης της αρχιτεκτονικής LEON3 για εφαρμογές SDR.

Παράμετροι	Τιμή
Χρονικό Διάστημα Δειγματοληψίας	20 ms
Πάχος Ψηφίδας Πυριτίου	0.15 mm
Επιφάνεια Πυριτίου (χωρίς αντίγραφα)	0.426213 mm^2
Επιφάνεια Κρυφής Μνήμης (L1+Local I+D)	0.370561 mm^2
Αεροθερμική Αντίσταση	0.1 K/W
Αεροθερμική Χωρητικότητα	140.4 J/K

κούς περιορισμούς (μέγιστη θερμοκρασία και θερμική διακύμανση) αποκλείονται από την εξερεύνηση του χώρου λύσεων. Όσων αφορά την υλοποίηση του ολοκληρωμένου σε τελικό chip, χρησιμοποιούνται τυπικά κριτήρια για ενσωματωμένους επεξεργαστές, όπως αυτά που αναφέρονται στην εργασία [27]. Με αυτό το τρόπο εξασφαλίζεται ότι η προτεινόμενη μεθοδολογία εξετάζει αρχιτεκτονικές λύσεις, οι οποίες ικανοποιούν τους περιορισμούς επιφάνειας πυριτίου, καθυστέρησης και θερμοκρασίας.

3.2.6. Πειραματικά αποτελέσματα

Αυτή η υποενότητα αναφέρει λεπτομερώς το περιβάλλον επαλήθευσης του προτεινόμενου μεθοδολογικού πλαισίου, καθώς και τις βασικές μετρήσεις που το επαληθεύουν.

3.2.6.1. Περιγραφή πειραματικού περιβάλλοντος

Σε αυτή την εργασία μελετώνται βασικές εφαρμογές της ψηφιακής βασικής ζώνης μετάδοσης. Συγκεκριμένα, οι εφαρμογές μελέτης, οι οποίες αναλύονται λεπτομερώς στην εργασία [28], συνοψίζονται ως εξής:

- *Adaptive differential pulse-code modulation*: Η εφαρμογή ADPCM είναι ένα είδος διαφορικής παλμοκωδικής διαμόρφωσης (DPCM), η οποία μεταβάλλει το μέγεθος του βήματος κβάντισης, προκειμένου να επιτρέψει επιπλέον μείωση του εύρους ζώνης για δεδομένη αναλογία σήματος προς θόρυβο (SNR).
- *Cyclic redundancy check*: Η εφαρμογή CRC είναι ένας κώδικας ανίχνευσης σφαλμάτων, ο οποίος έχει σχεδιαστεί προκειμένου να ανιχνεύει αλλαγές λογικών καταστάσεων σε ακολουθίες ψηφιακών δεδομένων και χρησιμοποιείται ως επί το πλείστον σε ψηφιακά δίκτυα επικοινωνιών.
- *Fast Fourier transform*: Η εφαρμογή FFT/iFFT είναι ένας αποδοτικός αλγόριθμος υπολογισμού του διακριτού μετασχηματισμού Fourier (DFT) και του αντίστροφου αυτού, αντίστοιχα.

- **GSM 06.10:** Η εφαρμογή GSM 06.10 είναι ένα πρωτόκολλο κωδικοποίησης ψηφιακής ομιλίας, το οποίο χρησιμοποιείται στη GSM ψηφιακή επικοινωνία κινητών τηλεφώνων.

Οι παραπάνω τηλεπικοινωνιακές εφαρμογές εμφανίζουν αυξημένες απαιτήσεις εύρους ζώνης. Όσον αφορά την αρχιτεκτονική, οι μονάδες υλικού με τις υψηλότερες τιμές πυκνότητα κατανάλωσης ενέργειας, κατά τη διάρκεια εκτέλεσης αυτών των εφαρμογών είναι η μονάδα εκτέλεσης εντολών, ο ελεγκτής κρυφής μνήμης, ο ελεγκτής κύριας μνήμης, η μονάδα αποσφαλμάτωσης (DSU) και οι ελεγκτές διαύλων AMBA AHB/APB.

Η υποενοότητα αυτή παρέχει μια σειρά από πειραματικά αποτελέσματα που προκύπτουν από την προτεινόμενη μεθοδολογία και αποδεικνύουν την αποτελεσματικότητα αυτής στη θερμική καταπόνηση. Ως πλατφόρμα αναφοράς για τις μετρήσεις χρησιμοποιείται ένα σύστημα SDR βασισμένο στον επεξεργαστή LEON3 [18]. Σε αυτή τη πλατφόρμα εφαρμόζεται η μεθοδολογία που περιγράφεται στην υποενοότητα 3.2.5, δημιουργώντας εναλλακτικές σχεδιαστικές λύσεις με αντίγραφα μονάδων υλικού στην μικρο-αρχιτεκτονική προκειμένου να μειώσει τη μέγιστη on-chip θερμοκρασία. Σημειώνεται ότι η λειτουργικότητα του επεξεργαστή LEON3 δεν επηρεάζεται από τη προσθήκη των αντιγράφων μονάδων υλικού.

Η πλειοψηφία των φαινομένων γήρανσης συνδέεται άμεσα με τις on-chip θερμοκρασίες. Οι υψηλές μέγιστες θερμοκρασίες κατά το χρόνο εκτέλεσης εφαρμογών μπορούν να οδηγήσουν, μεταξύ άλλων, σε αυξημένα ποσοστά αποτυχίας της λειτουργίας του ολοκληρωμένου κυκλώματος. Συνεπώς, η προτεινόμενη μεθοδολογία θέτει ως πρώτο κριτήριο την μελέτη της κατανομής της θερμοκρασίας στην επιφάνεια πυριτίου.

3.2.6.2. Επιπτώσεις της επιλεκτικής αντιγραφής στη θερμοκρασία

Το Σχήμα 3.7 απεικονίζει τη μεταβολή της μέγιστης θερμοκρασίας (σε βαθμούς Kelvin) αναφορικά προς εναλλακτικές αρχιτεκτονικές υλοποιήσεις. Οι οριζόντιοι άξονες αυτού του σχήματος αναπαριστούν τη συχνότητα λειτουργίας και την επιφάνεια πυριτίου της αρχιτεκτονικής. Οι άξονες αυτοί είναι βαθμονομημένοι σε κανονικοποιημένη μορφή ως προς τις αντίστοιχες μέγιστες τιμές που βρέθηκαν μεταξύ όλων των υποψηφίων αρχιτεκτονικών λύσεων. Ο κατακόρυφος άξονας αντιστοιχεί στις τιμές της on-chip θερμοκρασίας. Με βάση το Σχήμα 3.7, συμπεραίνεται ότι η θερμοκρασία μεταβάλλεται σημαντικά μεταξύ αρχιτεκτονικών με διαφορετική επιλογή των αντιγράφων μονάδων υλικού. Συγκεκριμένα, για την αρχιτεκτονική LEON3, αναφέρονται διακυμάνσεις της θερμοκρασίας στο εύρος τιμών 354-382 βαθμών Kelvin.

Σύμφωνα με τις μετρήσεις που απεικονίζονται στο σχήμα αυτό, είναι εμφανές ότι η αύξηση της επιφάνειας πυριτίου οδηγεί σε μία σχεδόν μονοτονική μείωση της θερμοκρασίας. Ωστόσο, η μείωση της θερμοκρασίας δεν είναι σταθερή καθώς η μεμονωμένη αντιγραφή των μονάδων υλικού δεν εγγυάται την εξομάλυνση της θερμικής καταπόνησης. Αυτή η διαπίστωση έχει ήδη δειχθεί στο Σχήμα 3.3(c)). Ως εκ τούτου, εκτός από τον αριθμό των προστιθέμενων αντιγράφων μονάδων υλικού, οι ιδιότητές τους επίσης, όπως η πυκνότητα κατανάλωσης ενέργειας,

Σχήμα 3.7 Διακυμάνσεις θερμοκρασίας για διαφορετικές αρχιτεκτονικές υλοποιήσεις της πλατφόρμας LEON3 SDR.

η επιφάνεια πυριτίου και η δραστηριότητα εκτέλεσης, είναι κρίσιμες κατά τη σχεδίαση μιας θερμικά βελτιστοποιημένης αρχιτεκτονικής.

Εκτός από την επιφάνεια πυριτίου, η μέγιστη συχνότητα λειτουργίας επηρεάζει επίσης, σε μεγάλο βαθμό, την on-chip θερμοκρασία. Με βάση το Σχήμα 3.7, οι εναλλακτικές αρχιτεκτονικές οδηγούν σε διακυμάνσεις της απόδοσης έως 14% και οι οποίες εμφανίζονται κυρίως λόγω: (i) της εισαγωγής των επιπλέον αντιγράφων μονάδων υλικού, (ii) της διαφορετικής χωροθέτησης, και (iii) του αυξημένου μήκους καλωδίου για τη σύνδεση των εν λόγω μονάδων.

Ένα επιπλέον ενδιαφέρον συμπέρασμα προκύπτει από την ερμηνεία του Σχήματος 3.7. Παρόλο που οι αρχιτεκτονικές με υψηλότερες συχνότητες λειτουργίας συνήθως παρουσιάζουν υψηλές θερμοκρασίες, εντούτοις, οι αρχιτεκτονικές με ταυτόχρονη υψηλή συχνότητα λειτουργίας και αυξημένη επιφάνεια πυριτίου παρουσιάζουν χαμηλές θερμοκρασίες. Όσον αφορά την αρχιτεκτονική LEON3, η πρόσθετη επιφάνεια πυριτίου χρησιμοποιείται από μονάδες υλικού με τιμές

χαμηλής πυκνότητας κατανάλωσης ενέργειας. Ως εκ τούτου, με την εισαγωγή περισσότερα αντιγράφων μονάδων υλικού καθίσταται δυνατή η βελτίωση της θερμικής εξάπλωσης. Ωστόσο, για αυτή τη διαπίστωση δεν μπορεί να εξαχθεί ασφαλές οικουμενικό συμπέρασμα καθώς η μετρική της θερμοκρασίας φαίνεται να είναι πιο αγνωστικιστή¹ σχετικά με τη μέγιστη συχνότητα λειτουργίας, σε σύγκριση με τη μετρική της επιφάνειας πυριτίου.

Με βάση αυτό το σχήμα είναι δυνατόν να επιλεγεί μια αρχιτεκτονική που συμβιβάζει με τον καλύτερο τρόπο τους σχεδιαστικούς περιορισμούς. Για τη παρούσα μελέτη της SDR αρχιτεκτονικής, διαπιστώθηκε ότι η βέλτιστη λύση σχεδιασμού στο πλαίσιο των προαναφερθέντων κριτηρίων (επιφάνεια πυριτίου, μέγιστη θερμοκρασία, και καθυστέρηση) είναι αυτή που χρησιμοποιεί τέσσερα αντίγραφα ελεγκτών AHB, τρία αντίγραφα μονάδες επεξεργασίας ακεραίων και δύο αντίγραφα ελεγκτών κρυφής μνήμης. Αυτή η αρχιτεκτονική λύση, η οποία αναφέρεται ως “Selected architecture” στα επερχόμενα σχήματα, ανήκει στις λύσεις που χαρακτηρίζονται ταυτόχρονα ως έγκυρες κατά την επιλογή βάση επιφάνειας πυριτίου, χρονικής καθυστέρησης και θερμικού προφίλ. Η επιλογή αυτής της αρχιτεκτονικής για περαιτέρω αξιολόγηση γίνεται εφόσον ανήκει στις λύσεις Pareto αναφορικά προς τη βελτίωση της αξιοπιστίας, όπως συζητείται με περισσότερες λεπτομέρειες στην υποενότητα 3.2.6.3.

Συγκεκριμένα, η αύξηση της επιφάνειας πυριτίου και της καθυστέρησης για την επιλεγμένη SDR αρχιτεκτονική LEON3 είναι 15% και 7%, αντίστοιχα, σε σύγκριση με την αρχική υλοποίηση (δίχως τη χρήση μονάδων αντιγραφής). Παρόλο που αυτές οι αυξήσεις δεν είναι αμελητέες για υλοποιήσεις ASIC, αναφέρεται ότι παρουσιάζουν μείωση της μέγιστης τιμής θερμοκρασίας κατά 17 βαθμούς Kelvin (ή 8%), η οποία με τη σειρά της οδηγεί σε βελτίωση της αξιοπιστίας. Επιπλέον, η προτεινόμενη μεθοδολογία επιλεκτικής αντιγραφής των μονάδων υλικού με αυξημένη πυκνότητα κατανάλωσης ενέργειας μπορεί επίσης να εφαρμοστεί σε αρχιτεκτονικές πολλαπλών πυρήνων, όπου η υποβάθμιση της απόδοσης ενδέχεται να είναι περισσότερο ανεκτή. Εκτός από την επιλεγμένη αρχιτεκτονική, οποιοσδήποτε άλλη αρχιτεκτονική υλοποίηση μπορεί να επιλεγεί, χωρίς να επηρεάζονται η αποτελεσματικότητα της προτεινόμενης μεθοδολογίας, στη περίπτωση που εφαρμοσθούν διαφορετικοί περιορισμοί σχεδίασης.

Προκειμένου να επισημανθεί η σημασία της σωστής αναγνώρισης των μονάδων υλικού που πρέπει να επιλεγθούν για αντιγραφή, το Σχήμα 5.12 απεικονίζει την μεταβολή της θερμοκρασίας, σε βαθμούς Kelvin, σε σχέση με την πυκνότητα κατανάλωσης ενέργειας (W/cm^2) για κάθε αρχιτεκτονική λύση. Επιλέγεται η αξιολόγηση αυτού του κριτηρίου, καθώς η πυκνότητα κατανάλωσης ενέργειας για τις υπάρχουσες και τις επερχόμενες συσκευές αποτελεί ένα σημαντικό ζήτημα για τους σχεδιαστές, ειδικά με τη συνεχή κλιμάκωση της τεχνολογίας. Η

¹Η λέξη “αγνωστικιστή” χρησιμοποιείται για να χαρακτηρίσει το γεγονός ότι σε αυτό το σημείο της μελέτης, δεδομένων των υφιστάμενων αποτελεσμάτων, δεν μπορεί να υπάρξει απόλυτη αλήθεια για την συσχέτιση των μετρικών της θερμοκρασίας και της μέγιστης συχνότητα λειτουργίας. Αγνωστικισμός ονομάζεται η φιλοσοφική θεώρηση ότι η αλήθεια ορισμένων μεταφυσικών υποθέσεων, όπως οι θεολογικοί ισχυρισμοί που αφορούν την ύπαρξη του Θεού, των θεών ή θεοτήτων, είναι είτε προς το παρόν άγνωστη είτε εγγενώς απρόσιτη. Ο όρος αυτός, καθώς και ο σχετικός όρος αγνωστικιστής πλάστηκαν από τον Τόμας Χάξλεϊ (Thomas Henry Huxley) το έτος 1869.

ερευνητική κοινότητα έχει ήδη εντοπιστεί αυτό το πρόβλημα, ενώ με βάση τις προβλέψεις, αναμένεται ότι η πυκνότητα κατανάλωσης ενέργειας για τον τεχνολογικό κόμβο των 14nm θα είναι υψηλότερη από 100 W/cm^2 [29].

Σχήμα 3.8 Η πυκνότητα κατανάλωσης ενέργειας σε σχέση με τη μέγιστη θερμοκρασία.

Για λόγους καλύτερης απεικόνισης, οι αρχιτεκτονικές λύσεις κατηγοριοποιούνται σε τρεις κατηγορίες, βάση της πυκνότητας κατανάλωσης ενέργειας, ως εξής:

- αρχιτεκτονικές με επιφάνεια πυριτίου μικρότερη από το 33% της μέγιστης επιφάνειας μεταξύ όλων των λύσεων
- αρχιτεκτονικές με επιφάνεια πυριτίου μεταξύ 33% και 66% σε σχέση με τη μέγιστη επιφάνεια μεταξύ όλων των λύσεων
- αρχιτεκτονικές με επιφάνεια πυριτίου μεγαλύτερη από το 66% της μέγιστης επιφάνειας μεταξύ όλων των λύσεων

Σημειώνεται ότι αυτή η ταξινόμηση, σε σχέση με την επιφάνεια πυριτίου που καταλαμβάνεται από διαφορετικές αρχιτεκτονικές, εφαρμόζεται επίσης και στα επερχόμενα σχήματα 3.9 και 3.10), δεδομένου ότι μπορεί να προσφέρει ποιοτικές συγκρίσεις για τη σημασία, καθώς και την αποτελεσματικότητα, της προτεινόμενης μεθοδολογίας αναφορικά προς τη μείωση της θερμική καταπόνησης.

Καθώς οι διαφορετικές αρχιτεκτονικές λύσεις αποτελούνται από διαφορετικές διατάξεις αντιγράφων μονάδων υλικού, η πυκνότητα κατανάλωσης ενέργειας επίσης διαφέρει σε μεγάλο βαθμό. Όπως συμπεραίνεται από το Σχήμα 3.9, δεν υπάρχει άμεση συσχέτιση μεταξύ της δεσμευμένης επιφάνειας πυριτίου, της συνολικής πυκνότητας κατανάλωσης ενέργειας και της μέγιστης on-chip θερμοκρασίας. Συγκεκριμένα, όσον αφορά τις αρχιτεκτονικές λύσεις που απεικονίζονται στο Σχήμα 3.9 και οι οποίες αντιστοιχούν σε αυξημένη επιφάνεια πυριτίου (περισσότερο από 66% της μέγιστης επιφάνειας), διαπιστώνεται ότι παρουσιάζουν αυξημένη πυκνότητα κατανάλωσης ενέργειας, αλλά παρουσιάζουν χαμηλές τιμές on-chip θερμοκρασίας. Αντίθετα, οι αρχιτεκτονικές με μικρότερη επιφάνεια πυριτίου (λιγότερο από το 66 % της μέγιστης επιφάνειας) εμφανίζουν μειωμένη πυκνότητα κατανάλωσης ενέργειας και αυξημένες τιμές on-chip θερμοκρασίας. Αυτό το παράδοξο αποτέλεσμα δικαιολογείται από το γεγονός ότι στη παρούσα ανάλυση δεν έχει αναλυθεί το είδος των μονάδων υλικού που επιλέγονται για αντιγραφή. Συνεπώς, συμπεραίνεται ότι ο τύπος των μονάδων αντιγραφής έχει μεγάλη επίδραση στη θερμική συμπεριφορά του ολοκληρωμένου κυκλώματος.

Ως τελευταίο συμπέρασμα αυτής της ανάλυσης αναφέρεται η παρατήρηση ότι ακόμη και οι αρχιτεκτονικές με αυξημένη πυκνότητα κατανάλωσης ενέργειας μπορούν να επιτύχουν σημαντική μείωση της on-chip θερμοκρασίας. Το σημείο αυτό επιβεβαιώνει το επιχείρημα που αναφέρθηκε στην υποενότητα 3.2.5, ότι η μη βέλτιστη αντιγραφή των μονάδων υλικού οδηγεί σε παρόμοιο, ή ακόμα και υψηλότερο, θερμική προφίλ, σε σύγκριση με την αρχική υλοποίηση της αρχιτεκτονικής SDR.

Στο Σχήμα 3.9, σημειώνονται επίσης οι λύσεις που αντιστοιχούν στις προτεινόμενες αρχιτεκτονικές λύσεις, καθώς και οι αρχικές λύσεις, δίχως την προσθήκη αντιγράφων μονάδων υλικού. Όπως φαίνεται, οι επιλεγμένες λύσεις επιτυγχάνουν μείωση της μέγιστης τιμής θερμοκρασίας κατά περίπου 17 βαθμούς Kelvin, που αντιστοιχούν σε ποσοστό 8%.

Το Σχήμα 3.9 απεικονίζει τη συσχέτιση μεταξύ της επιφάνειας πυριτίου και της μέγιστης τιμής θερμοκρασίας του ολοκληρωμένου κυκλώματος. Για λόγους πληρότητας, οι εναλλακτικές αρχιτεκτονικές κατηγοριοποιούνται σε τρεις ομάδες ανάλογα με την επιφάνεια πυριτίου. Επίσης απεικονίζεται η κλίση της θερμοκρασίας για κάθε μία από αυτές τις ομάδες.

Με βάση το Σχήμα 3.9, η μέγιστη κλίση θερμοκρασίας παρουσιάζεται στις αρχιτεκτονικές με μικρή επιφάνεια πυριτίου, ενώ οι αρχιτεκτονικές με επιφάνεια πυριτίου μεγαλύτερη από το 66% της μέγιστης επιφάνειας πυριτίου, παρουσιάζουν σχεδόν γραμμική και σταθερή κλίση. Αν λάβουμε υπόψη και η μικρότερη καθυστέρηση εναέρια τίθενται από τις συσκευές αποτελείτο από λιγότερες ρεπλικά μπλοκ (όπως ήδη φαίνεται στο Σχήμα ref fig: Σχ.7), ένα επιπλέον φίλτράρισμα των παράγωγων αρχιτεκτονικής instantiations μπορεί να πραγματοποιηθεί. Οι λύσεις που αντιστοιχούν σε δεσμευμένη επιφάνειας πυριτίου υψηλότερη από το 66% της μέγιστης επιφάνειας θεωρούνται μη αποδεκτές, λόγω αυξημένης καθυστέρησης, και ως εκ τούτου απομακρύνονται από το χώρο λύσεων.

Επιπλέον, οι τιμές θερμοκρασίας για αρχιτεκτονικές με δέσμευση επιφά-

Σχήμα 3.9 Η επιφάνεια πυριτίου σε σχέση με τη μέγιστη θερμοκρασία.

νειας πυριτίου λιγότερο από το 33% της συνολικής επιφάνειας, είναι περίπου 3x υψηλότερες, σε σύγκριση με τις υπόλοιπες λύσεις. Οι διακυμάνσεις υψηλής θερμοκρασίας συνήθως οδηγούν σε αυξημένο κόστος κατά τα στάδια σχεδιασμού συσκευασίας και ψύξης και ως εκ τούτου δεν είναι επιθυμητές. Οι λύσεις αυτές επίσης απομακρύνονται από το χώρο λύσεων.

Το συμπέρασμα αυτό είναι ιδιαίτερα σημαντικό στη διαδικασία εύρεσης του ακριβές αριθμού αντιγραφής μονάδων υλικού. Συγκεκριμένα, με βάση Σχήμα 3.9, είναι σαφές ότι μόνο μερικά αντίγραφα των μονάδων με αυξημένη πυκνότητα κατανάλωσης ενέργειας πρέπει να χρησιμοποιηθούν ώστε να επιτευχθεί η επιθυμητή ισορροπία μεταξύ της μείωσης της θερμοκρασίας, της επιφάνειας πυριτίου και της χρονικής καθυστέρησης του ολοκληρωμένου κυκλώματος.

3.2.6.3. Μελέτη επίδρασης της θερμοκρασίας στην αξιοπιστία

Η αξιοπιστία ορίζεται ως η πιθανότητα για ένα σύστημα να εκτελεί την απαιτούμενη λειτουργία του σε καθορισμένες λειτουργικές συνθήκες για μια συγκεκριμένη χρονική περίοδο. Ωστόσο, καθώς η μετρική αυτή αποτελεί μια μαθηματική πρόβλεψη με κάποιο βαθμό εμπιστοσύνης, εξαρτάται σε μεγάλο βαθμό από έναν αριθμό παραμέτρων.

Το στάδιο επιτάχυνσης ελέγχου ορθής λειτουργίας (accelerated life testing) αποτελεί μία διαδικασία επαλήθευσης της λειτουργικότητας ενός προϊόντος ύστερα από τη φάση της υλοποίησης και πριν την εναπόθεση στην αγορά. Χρησιμοποιεί πολλαπλές μεθόδους δοκιμής που προσπαθούν να μειώσουν τη διάρκεια ζωής ενός προϊόντος, ή να υποβαθμίσουν την απόδοση αυτών. Στόχος αυτών των δοκιμών είναι η λήψη δεδομένων απόδοσης, τα οποία όταν αναλυθούν σωστά, παρέχουν λογικές εκτιμήσεις της μέγιστης διάρκειας ορθής λειτουργίας καθώς και εκτιμήσεις απόδοσης των συστημάτων υπό κανονικές συνθήκες λειτουργίας. Μερικές από τις μεθόδους αυτές αναπαράγουν τις κατάλληλες συνθήκες προκειμένου να εκδηλωθούν πρώιμες βλάβες υλικού, οι οποίες συνήθως εκδηλώνονται στα πρώτα χρόνια λειτουργίας των υπό ανάπτυξη συστημάτων, με αποτέλεσμα την ελαχιστοποίηση του ρίσκου παραγωγής μεγάλου αριθμού συστημάτων. Τόσο ο τύπος των μεθόδων ελέγχου όσο και η διάρκεια των ελέγχων αυτών, χρησιμοποιούνται για τον προσδιορισμό των συνθηκών κανονικής λειτουργίας. Όσον αφορά τους σχεδιασμούς SoC, συνήθως η πλειοψηφία αυτών των μηχανισμών γήρανσης είναι στενά συνδεδεμένη με τις θερμοκρασίες στην επιφάνεια πυριτίου του ολοκληρωμένου κυκλώματος.

$$A_f \propto \exp\left[\left(\frac{E_a}{k}\right) \times \left(\frac{1}{T_u} - \frac{1}{T_t}\right)\right] \quad (3.1)$$

Η επίδραση αυτών των παραγόντων ελέγχου μπορεί να οριστεί με μαθηματικό τρόπο. Ακολούθως, παρουσιάζεται μία μαθηματική προσέγγιση της γήρανσης ενός ολοκληρωμένου κυκλώματος σε σχέση με τη θερμική καταπόνηση στην επιφάνεια πυριτίου του. Πρόκειται για το μοντέλο Arrhenius που παρουσιάζεται στην Εξίσωση 3.1. Συγκεκριμένα, αυτό το μοντέλο συσχετίζει το χρονικό προσδόκιμο ορθής λειτουργίας ενός συστήματος SoC με την μέγιστη τιμή θερμοκρασίας, σε σύγκριση με τη θερμοκρασία λειτουργίας του, υπό κανονικές συνθήκες. Το χρονικό προσδόκιμο ορθής λειτουργία αναπαριστάται από μία παράμετρο επιτάχυνσης γήρανσης του ολοκληρωμένου κυκλώματος. Το Σχήμα 3.10 απεικονίζει πώς αυτή η παράμετρος διαφέρει για τις εναλλακτικές αρχιτεκτονικές που εξετάστηκαν στη παρούσα εργασία. Συγκεκριμένα, ο οριζόντιος άξονας στο σχήμα αυτό απεικονίζει τη μέση θερμοκρασία, ενώ ο κάθετος άξονας απεικονίζει την παράμετρο A_f , δηλαδή τον τρόπο με τον οποίο η on-chip θερμοκρασία επηρεάζει τη γήρανση του ολοκληρωμένου κυκλώματος.

όπου A_f είναι ο παράγοντας επιτάχυνσης, E_a είναι η ενέργεια ενεργοποίησης σε μονάδες electron-volts (αυτή η τιμή είναι 0.5 eV για τη περίπτωση σφαλμάτων σε πυρίτιο), k είναι η σταθερά Boltzmann, ενώ οι μεταβλητές T_u και T_t είναι οι θερμοκρασίες αναφοράς και λειτουργίας σε βαθμούς Kelvin, αντίστοιχα.

Με βάση τις τιμές που απεικονίζονται στο Σχήμα 3.10, συμπεραίνεται ότι η επιλεγείσα αρχιτεκτονική λύση επιτυγχάνει σχεδόν την ελάχιστη τιμή της παραμέτρου επιτάχυνσης γήρανσης A_f σε σχέση με όλες τις υποψήφιες λύσεις. Επιπλέον, η συμβατική προσέγγιση για το σχεδιασμό της αρχιτεκτονικής LEON3 παρουσιάζει περίπου 14% υψηλότερη τιμή για την παράμετρο αυτή. Αυτό συμβαίνει κυρίως λόγω της πρόσθετης θερμικής καταπόνησης που εισάγουν οι μονάδες της αρχιτεκτονικής με αυξημένη πυκνότητα κατανάλωσης ενέργειας. Ωστόσο, αναφέρεται ότι παρόλο που η βασική αρχιτεκτονική αναφοράς, η οποία δεν χρησιμοποιεί αντίγραφα μονάδων υλικού, δεν παρουσιάζει την μέγιστη επιτάχυνση γήρανσης, όπως θα ήταν αναμενόμενο. Συγκεκριμένα, υπάρχουν λύσεις που αντιστοιχούν σε παραμέτρους A_f με τιμές έως 20% μεγαλύτερες σε σχέση με την αρχιτεκτονική αναφοράς. Αυτό συμβαίνει καθώς δεν είναι δεδομένο ότι όλοι οι δυνατοί συνδυασμοί των μονάδων αντιγραφής υλικού οδηγούν σε λύσεις βελτιστοποιημένης θερμικής καταπόνησης. Ωστόσο το προτεινόμενο πλαίσιο καταφέρει να συγκεντρώσει τις υποψήφιες λύσεις που καλύπτουν το χώρο λύσεων μέσω συστηματικής ανάλυσης και καταφέρει να αποδώσει λύσεις με μικρότερο παράγοντα γήρανσης έως 14%, για τη συγκεκριμένη πλατφόρμα SDR.

Εκτός από την εξίσωση Arrhenius, οι λύσεις που προτείνει η παρούσα ανάλυση αξιολογούνται επίσης στο πλαίσιο της χρονικά εξαρτώμενης διηλεκτρική κατάρρευσης (TDDB) [30]. Το φαινόμενο κατάρρευσης του οξειδίου αποτελεί σοβαρό πρόβλημα αξιοπιστίας στη βιομηχανία ημιαγωγών, ειδικότερα λόγω της κλιμάκωσης της τεχνολογίας προς μικρότερες διαστάσεις. Τα ελαττώματα κατασκευής λόγω του φαινομένου TDDB οφείλονται κατά κύριο λόγο στην παγίδευση φορτίου στο οξειδίο πύλης, το οποίο δημιουργεί ένα ηλεκτρικό πεδίο

και το οποίο στη συνέχεια εκφορτίζεται διαμέσου του οξειδίου, με αποτέλεσμα την κατάρρευση μετά από κάποιο χρονικό διάστημα. Η μέση τιμή κατάρρευσης (MTTF) λόγω του φαινομένου TDDB περιγράφεται από την Εξίσωση 3.2.

$$MTTF = A_0 \times \exp(-\lambda \times E_{ox}) \times \exp\left(\frac{E_a}{k \times T}\right) \quad (3.2)$$

όπου η μεταβλητή λ είναι μία παράμετρος επιτάχυνσης πεδίου και εξαρτάται από την on-chip θερμοκρασία.

Σχήμα 3.11 Αξιολόγηση διαφορετικών αρχιτεκτονικών αναφορικά προς το πρόβλημα TDDB.

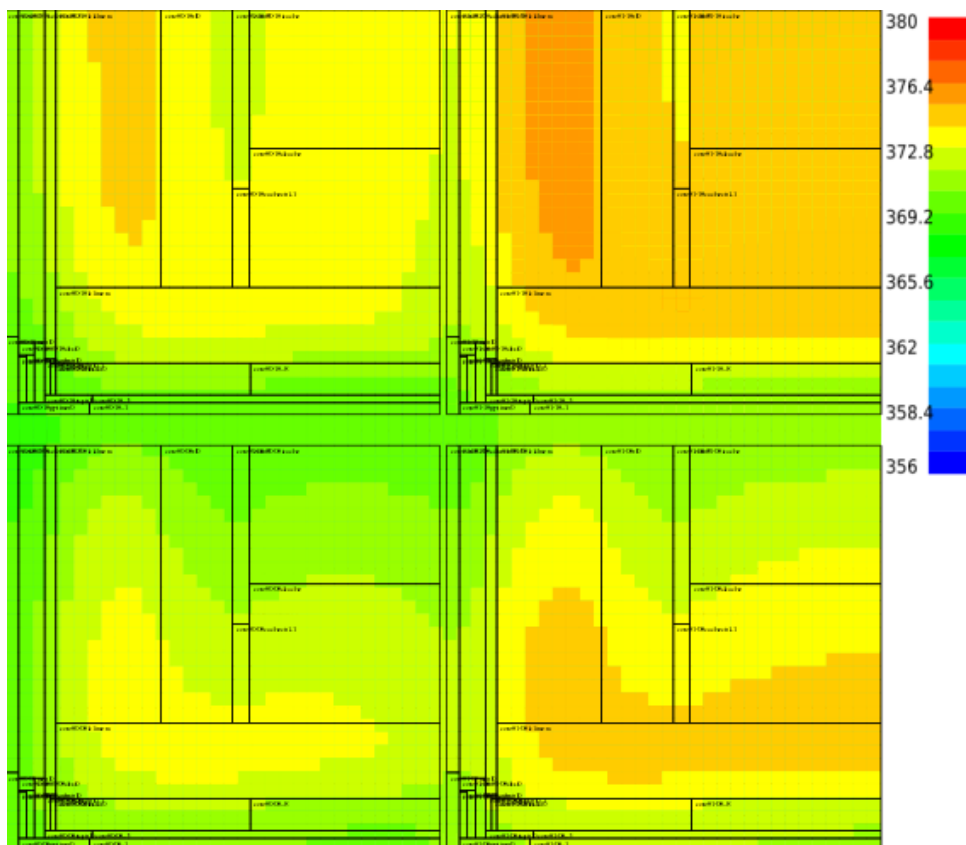
Το Σχήμα 3.11 απεικονίζει τη διακύμανση της παραμέτρου MTTF για διαφορετικές αρχιτεκτονικές, όπως αυτές προκύπτουν από τη παρούσα εργασία. Οι τιμές στον κατακόρυφο άξονα έχουν βαθμονομηθεί σε κανονικοποιημένη μορφή, σε σχέση με την τιμή MTTF της αρχιτεκτονική αναφοράς.

Με βάση το Σχήμα 3.11, συμπεραίνεται ότι η αύξηση της μέγιστης on-chip τιμής θερμοκρασίας, οδηγεί σε αύξηση του χρόνου κατάρρευσης MTTF. Αυτό το συμπέρασμα αποδίδεται στο στενό συσχετισμό του φαινομένου γήρανσης και της τιμής της θερμοκρασίας. Επιπλέον, η επιλεγμένη αρχιτεκτονική επιτυγχάνει βελτίωση της παραμέτρου MTTF, περίπου κατά 14% σε σύγκριση με την αρχιτεκτονική αναφοράς του επεξεργαστή LEON3. Σημειώνεται ότι σε περίπτωση

που αυξηθεί η επιφάνεια πυριτίου της αρχιτεκτονικής, μέσω της εισαγωγής περισσότερων αντιγράφων μονάδων υλικού, η παράμετρος MTTF ενδέχεται να αυξηθεί περαιτέρω. Αυτό οφείλεται στο γεγονός ότι οι αρχιτεκτονικές με περισσότερες μονάδες αντιγράφων υλικού συνήθως καταλαμβάνουν μεγαλύτερη επιφάνεια πυριτίου, η οποία με τη σειρά της βελτιώνει τη θερμική διασπορά. Ωστόσο, όπως έχει ήδη αναφερθεί, αυτή η βελτίωση στην παράμετρο MTTF (0,86 - 0,81) μπορεί να οδηγήσει σε αρχιτεκτονικές λύσεις με μη-αποδεκτή επιφάνεια πυριτίου.

3

3.2.6.4. Σχεδιασμός αρχιτεκτονικών SDR πολλαπλών επεξεργαστών ενιαίας ψηφιάδας



Σχήμα 3.12 Θερμική ανάλυση για ένα σύστημα CMP που αποτελείται από 2x2 επεξεργαστές LEON3.

Αυτό το εδάφιο περιγράφει τα αποτελέσματα που προέρχονται από την εφαρμογή της προτεινόμενης μεθοδολογίας για το σχεδιασμό αρχιτεκτονικών SDR πολλαπλών επεξεργαστών. Στη παρούσα εργασία επιλέχθηκαν τέσσερις επεξεργαστές LEON3. Ωστόσο ο αριθμός αυτός είναι παραμετρικός στο προτεινόμενο πλαίσιο και μπορεί να τροποποιηθεί κατάλληλα με βάση τις σχεδιαστικές

απαιτήσεις. Ως αρχιτεκτονική αναφοράς στην παρούσα μελέτη χρησιμοποιούμε τη βελτιστοποιημένη λύση από τη προηγούμενη ανάλυση, που είχε σημειωθεί ως “selected” στα προηγούμενα σχήματα. Για τα ακόλουθα σχήματα, η λύση αυτή συμβολίζεται ως “reference solution”.

Το Σχήμα 3.12 απεικονίζει το θερμικό προφίλ για την επιλεγμένη αρχιτεκτονική πολλαπλών επεξεργαστών, όπως ανακτήθηκε με το εργαλείο Hotspot. Με βάση αυτό το σχήμα συμπεραίνεται ότι μερικές μονάδες υλικού παρουσιάζουν αυξημένες τιμές θερμοκρασίας σε συγκεκριμένες θέσεις στην επιφάνεια του ολοκληρωμένου κυκλώματος. Σημειώνεται ότι αυτές οι θέσεις διαφέρουν από αυτές που αναφέρθηκαν προηγουμένως στο Σχήμα 3.3, λόγω (i) της θερμικής διάδοσης, και (ii) των διαφορετικών χωροθετήσεων του κάθε επεξεργαστή LEON3. Στη συνέχεια ποσοτικοποιείται η αποτελεσματικότητα της ανωτέρω αρχιτεκτονικής, όταν χρησιμοποιούνται πολλαπλά στιγμιότυπα της αρχιτεκτονικής πολλαπλών επεξεργαστών.

Σχήμα 3.13 Κανονικοποιημένη πυκνότητα κατανάλωσης ενέργειας σε σχέση με την επιφάνεια πυριτίου, για συστήματα πολλαπλών επεξεργαστών LEON3.

Το Σχήμα 3.13 δείχνει τη μεταβολή της πυκνότητας κατανάλωσης ενέργειας, σε σχέση με λύσεις υψηλότερης επιφάνειας πυριτίου. Με βάση αυτό το σχήμα συμπεραίνεται ότι η πυκνότητα κατανάλωσης ενέργειας δεν εξαρτάται σε μεγάλο βαθμό από την επιφάνεια πυριτίου. Ωστόσο αναφέρεται ότι κατά τη διάρκεια αυτής της ανάλυσης, η πρόσθετη επιφάνεια πυριτίου δεσμεύεται από την εισαγωγή κενών κελιών (filler cells) μεταξύ των μονάδων υλικού, παρά από τα

αυτούσια αντίγραφα μονάδων υλικού.

Σχήμα 3.14 Κανονικοποιημένη μέγιστη θερμοκρασία σε σχέση με την επιφάνεια πυριτίου, για συστήματα πολλαπλών επεξεργαστών LEON3.

Σε αντίθεση με το συμπέρασμα για την πυκνότητα κατανάλωσης ενέργειας, η επιφάνεια πυριτίου έχει μεγάλη επίδραση στη μέγιστη θερμοκρασία στην επιφάνεια του ολοκληρωμένου κυκλώματος. Η συσχέτιση αυτή απεικονίζεται στο Σχήμα 3.14. Συγκεκριμένα, με βάση αυτό το σχήμα συμπεραίνεται ότι η ελεγχόμενη αύξηση της επιφάνειας πυριτίου έως 20%, οδηγεί σε μείωση της μέγιστης θερμοκρασίας κατά 25% (0.85x).

3.2.7. Συμπεράσματα

Σε αυτή την εργασία προτάθηκε η υιοθέτηση τεχνικών επιλεκτικής αντιγραφής για τη βελτιστοποίηση της θερμικής διάδοσης στην επιφάνεια πυριτίου αρχιτεκτονικών SoC, για εφαρμογές SDR. Αναπτύχθηκε μια αυτοματοποιημένη μεθοδολογία, η οποία επιτρέπει τη θερμική αξιολόγηση σε επίπεδο μικροαρχιτεκτονικής.

Αποδείχθηκε ότι η χρήση επιλεκτικής αντιγραφής οδηγεί σε βελτιστοποιημένες αρχιτεκτονικές λύσεις με ελάχιστη θερμική καταπόνηση και χαμηλά επίπεδα απώλειας χρονικής καθυστέρησης και αύξησης επιφάνειας πυριτίου. Τα πειραματικά αποτελέσματα έδειξαν ότι είναι δυνατή η βελτιστοποίηση της διερεύνησης χώρου λύσεων, προκειμένου να αξιολογηθούν αρχιτεκτονικές λύσεις

που συμβιβάζουν ταυτόχρονα τις προαναφερθείσες μετρικές. Επιπλέον, δείχνεται ότι η προτεινόμενη προσέγγισή βελτιώνει κατά 14% τα φαινόμενα γήρανσης του ολοκληρωμένου κυκλώματος.

3.3. Ετερογενείς Αρχιτεκτονικές Δικτύων-σε-Ψηφίδα για Εφαρμογές Πολυμέσων

3.3.1. Εισαγωγή

Στην τεχνολογία ημιαγωγών, ιστορικά, η τεχνολογία επεξεργασίας δεδομένων απαιτούσε, μέχρι πρότινος, μεγαλύτερη έρευνα, καινοτομία και κόστος κατασκευής, σε σχέση με την τεχνολογία επικοινωνίας. Ωστόσο, με την συνεχή κλιμάκωση της τεχνολογίας ολοκλήρωσης, αυτή η κατάσταση αντιστράφηκε. Συγκεκριμένα, η τεχνολογία ολοκλήρωσης προσφέρει συνεχώς αυξημένη υπολογιστική ισχύ στην επεξεργασία δεδομένων. Αντίθετα, η σμίκρυνση των διαστάσεων και των τάσεων κατωφλίου δημιουργεί βασικούς φυσικούς περιορισμούς στην επικοινωνία δεδομένων, όπως στο χρόνο μετάδοσης των ηλεκτρικών σημάτων, στη κατανάλωση ενέργειας για επιμήκη καλώδια μεταφοράς, κλπ. Συγκριτικά με την επικοινωνία εκτός του ολοκληρωμένου κυκλώματος (off-chip), η επικοινωνία εντός του ολοκληρωμένου κυκλώματος (on-chip) είναι περισσότερο πλεονεκτική, αναφορικά προς τη ταχύτητα επικοινωνίας, το κόστος κατασκευής και την ανοχή σφαλμάτων. Αυτός είναι και ο βασικότερος λόγος που οι αρχιτεκτονικές ενιαίου συστήματος-σε-ψηφίδα single-chip SoCs έχουν καθιερωθεί ως λύσεις σε προβλήματα επικοινωνίας. Παρόλο που ο διαμοιραζόμενος δίαυλος στις αρχιτεκτονικές SoC αποτελεί μία απλή διεπαφή και επιτρέπει εύκολο προγραμματισμό διεργασιών των υποσυστημάτων, εντούτοις, τα καλώδια διασύνδεσης δεν κλιμακώνονται αναλογικά την κλιμάκωση των τρανζίστορ και συνεπώς η αναλογία μεταξύ κόστους επικοινωνίας και επεξεργασίας συνεχώς αυξάνεται. Αυτό το πρόβλημα γίνεται ακόμα πιο μεγάλο στα σύγχρονα πολυπύρνα συστήματα. Αντιθέτως, η αρχιτεκτονική Network-on-Chip (NoC) είναι μια σχετικά νέα προσέγγιση που επιτρέπει όχι μόνο πιο αποδοτικές διασυνδέσεις αλλά και πιο αποδοτικό σχεδιασμό και διαδικασίες επαλήθευσης για μοντέρνα συστήματα MPSoCs [31].

Λόγω του προβλήματος διασύνδεσης, η ερευνητική κοινότητα έχει επικεντρωθεί στη βελτιστοποίηση των αρχιτεκτονικών NoC, καθώς και στα εργαλεία σχεδιασμού τους [32] [33] [34] [35] [36]. Επιπλέον, πρόσφατα, έχουν γίνει προσπάθειες υλοποίησης προχωρημένων τεχνολογιών σχεδίασης υλικού, προκειμένου να επιτευχθεί επιπλέον βελτίωση της απόδοσης των πλατφορμών NoC. Μεταξύ άλλων, η τεχνολογία ολοκλήρωσης τριών διαστάσεων (3-D), η οποία επιτρέπει κάθετη τοποθέτηση πολλαπλών επιπέδων πυριτίου και διασύνδεση μεταξύ αυτών χρησιμοποιώντας Through-Silicon Vias (TSVs), εισάγει τοπικότητα στον άξονα-z, επιτρέποντας κατά μέσο όρο μικρότερες διασυνδέσεις μεταξύ των τμημάτων του συνολικού συστήματος [37].

Η υπάρχουσα προσέγγιση στο σχεδιασμό τρισδιάστατων NoCs αποτελείται αποκλειστικά από τρισδιάστατους δρομολογητές. Υποθέτοντας μια τρισδιάστατη τοπολογία πλέγματος, αυτοί οι δρομολογητές, εκτός από την άμεση σύνδεση στους τέσσερις γειτονικούς κόμβους που βρίσκονται στο ίδιο επίπεδο, παρέχουν επίσης συνδεσιμότητα σε κάθετα ευθυγραμμισμένους δρομολογητές (σε υψηλότερα και χαμηλότερα επίπεδα). Παρόλο που μια τέτοια επιλογή οδηγεί

σε ομοιόμορφη αξιοποίηση υλικού, εντούτοις, σπάνια μπορεί να υιοθετηθεί σαν αποδοτική λύση, καθώς δε λαμβάνει υπόψιν τις απαιτήσεις της εφαρμογής για τη μεταφορά των δεδομένων. Συγκεκριμένα, καθώς το NoC είναι συνήθως μια υποδομή επικοινωνίας προσανατολισμένη στις ανάγκες της εφαρμογής, θα πρέπει να γίνει προσεχτική συστημική ανάλυση προκειμένου να υλοποιηθεί μία βέλτιστη αρχιτεκτονική. Προς αυτή τη κατεύθυνση, στη παρούσα εργασία, προτείνεται η χρήση ετερογενών τρισδιάστατων NoCs, τα οποία αντιμετωπίζουν καλύτερα τους περιορισμούς επικοινωνίας που τίθενται από τις εκάστοτε εφαρμογές. Με τον όρο *ετερογενείς* αναφερόμαστε σε αρχιτεκτονικές που συνδυάζουν ένα σύνολο δισδιάστατων και τρισδιάστατων δρομολογητών σε ένα μόνο NoC, ενώ η χωρική τοποθέτηση αυτών των δρομολογητών στην αρχιτεκτονική ορίζεται από τις απαιτήσεις της εφαρμογής.

Μεγάλο μέρος των υφιστάμενων ερευνητικών εργασιών στο σχεδιασμό ετερογενών τρισδιάστατων NoCs [37] [32] [35] [36] βασίζονται σε αφηρημένα μοντέλα και συνεπώς δε μπορούν να εξαχθούν άμεσα συμπεράσματα απόδοσης για την υλοποίησή τους. Συγκεκριμένα, στις υφιστάμενες ερευνητικές εργασίες, η απόδοση του NoC καθώς και η κατανάλωση ενέργειας συνήθως προσεγγίζονται με βάση τον αριθμό των hops (συνδέσεις μεταξύ διπλανών δρομολογητών) που ένα πακέτο πρέπει να διασχίσει για να μεταφερθεί από την πηγή στους κόμβους προορισμού, αγνοώντας παραμέτρους που σχετίζονται με τη φυσική υλοποίηση. Επίσης, αυτές οι προσεγγίσεις δε λαμβάνουν υπόψιν περιορισμούς που τίθενται από την επιλεγμένη τρισδιάστατη τεχνολογία, οδηγώντας σε μη αποδεκτές αρχιτεκτονικές λύσεις, π.χ. χαρακτηρίζουν λύσεις με μη-αποδεκτό αριθμό από TSVs, αναφορικά προς τους περιορισμούς κατασκευής και κόστους. Στη παρούσα εργασία γίνεται χρήση ενός πλαισίου βασισμένο σε εμπορικά εργαλεία (C-to-Silicon and SoC Encounter) της εταιρείας Cadence [38]. Η εφαρμογή αυτού του πλαισίου σε εφαρμογές DSP απέδειξε ότι το προτεινόμενο ετερογενές τρισδιάστατο NoC υπερτερεί, σχεδιαστικά, σε σχέση με τα συμβατικά τρισδιάστατα και δισδιάστατα NoCs, καθώς επιτυγχάνει κατά μέσο όρο 25% υψηλότερη μέγιστη συχνότητα λειτουργίας και κατά 39% λιγότερη κατανάλωση ενέργειας.

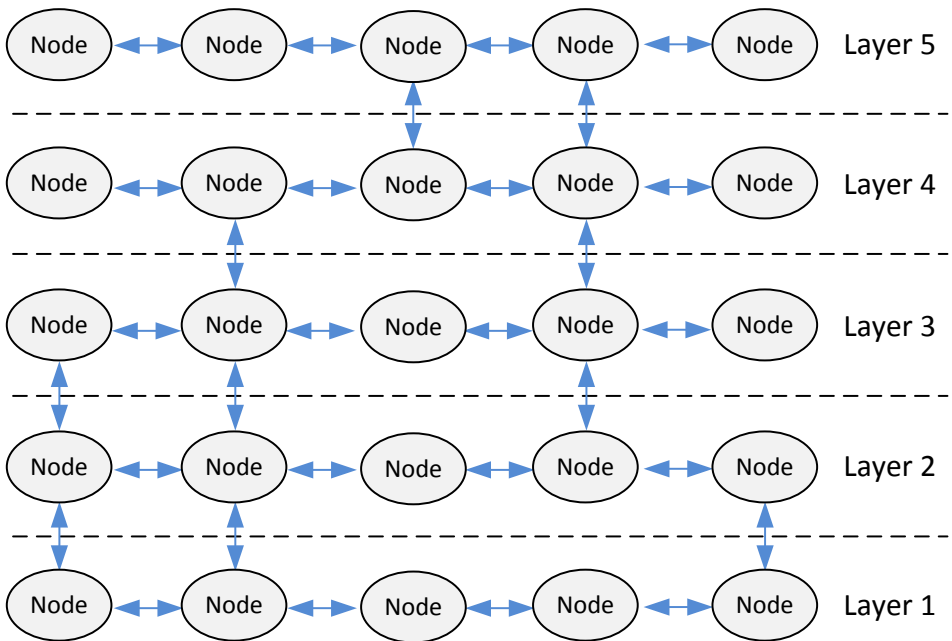
Η υπόλοιπη εργασία οργανώνεται ως ακολούθως: η υποενότητα 3.3.2 περιγράφει την έννοια της ετερογενούς τρισδιάστατης NoC αρχιτεκτονικής. Η προτεινόμενη μεθοδολογία για το σχεδιασμό μιας τέτοιας αρχιτεκτονικής καθώς και το πλαίσιο εργαλείων για την εκτέλεση γρήγορης αξιολόγησης αυτών των NoCs παρουσιάζεται στην υποενότητα 3.3.3. Η υποενότητα 3.3.4 παρουσιάζει έναν αριθμό ποιοτικών και ποσοτικών αποτελεσμάτων που αποδεικνύουν την αποτελεσματικότητα της προτεινόμενης λύσης. Τέλος, τα συμπεράσματα συνοψίζονται στην υποενότητα 3.3.5.

3.3.2. Η αρχιτεκτονική της προτεινόμενης προσέγγισης διασύνδεσης

Αυτή η υποενότητα εισηγείται την αρχιτεκτονική οργάνωση του προτεινόμενου σχήματος επικοινωνίας που αποτελείται από ένα συνδυασμό δισδιάστατων και τρισδιάστατων δρομολογητών. Συγκεκριμένα, ένας δισδιάστατος δρομολογητής μπορεί να χρησιμοποιηθεί όπου ένα εισερχόμενο κομμάτι δρομολόγησης συνδέεται με καλώδια στο ίδιο επίπεδο σχεδιασμού ($F_s = 3$). Η ευελιξία του δρο-

μολογητή, ορισμένη ως F_S , δίνει τον αριθμό των κατευθύνσεων στις οποίες κάθε εισερχόμενο καλώδιο μπορεί να συνδεθεί. Εναλλακτικά, αφού ένας τρισδιάστατος δρομολογητής επίσης υποστηρίζει συνδέσεις σε τρίτη διάσταση (υψηλότερο και χαμηλότερο επίπεδο), η τιμή της ευελιξίας του δρομολογητή ισούται με πέντε ($F_S = 5$).

Για να αποτυπωθούν οι διαφορές μεταξύ αυτών των δύο βασικών δρομολογητών, υποθέτεται ένα γράφημα εκτέλεσης της εφαρμογής, το οποίο απεικονίζεται στο Σχήμα 3.15, το οποίο αντιστοιχίζεται σε ένα τρισδιάστατο chip αποτελούμενο από πέντε επίπεδα πυριτίου. Τα βέλη στο σχήμα αυτό δείχνουν δεσμούς επικοινωνίας είτε στην οριζόντια είτε στη κάθετη κατεύθυνση μεταξύ διπλανών δρομολογητών. Όπως φαίνεται, δεν έχουν όλοι αυτοί οι δρομολογητές παρόμοιες απαιτήσεις για τη μεταφορά δεδομένων. Καθώς ο σχεδιασμός μίας ομογενούς αρχιτεκτονικής είναι πιο απλός και πιο οικονομικός, συγκρινόμενος με μια μη-κανονική (irregular) πλατφόρμα, στη παρούσα εργασία οι δρομολογητές ομαδοποιούνται σε δύο ομάδες: (i) οι δρομολογητές που υποστηρίζουν συνδεσιμότητα στην κάθετη κατεύθυνση (ii) οι δρομολογητές που παρέχουν δρομολόγηση πακέτου αποκλειστικά μέσα στο ίδιο επίπεδο. Έτσι αντικαθιστώντας τους δρομολογητές που ανήκουν στη δεύτερη ομάδα με τις ισοδύναμες δυσδιάστατες υλοποιήσεις, όπως φαίνεται στο Σχήμα 3.16(a), αναμένεται μια πιο αποδοτική υλοποίηση υλικού.



Σχήμα 3.15 Παράδειγμα γράφου επικοινωνίας εφαρμογής.

Η εργασία εισηγείται μια ακόμα βελτιστοποίηση. Συγκεκριμένα, παραμετροποιούνται οι δισδιάστατοι και τρισδιάστατοι δρομολογητές ανάλογα με τη χω-

ρική τοποθέτηση σε κάθε επίπεδο. Συγκεκριμένα, όσον αφορά τους τρισδιάστατους δρομολογητές που ανατίθενται στο χαμηλότερο επίπεδο (Επίπεδο 1) και υψηλότερο (Επίπεδο 5), δεν υπάρχει ζήτηση για παροχή συνδεσιμότητας στο χαμηλότερο και υψηλότερο επίπεδο αντίστοιχα, όπως φαίνεται στο Σχήμα 3.16(b). Ομοίως, οι δρομολογητές που ανατίθενται στην περιφέρεια κάθε επιπέδου, μπορούν να σχεδιαστούν με λιγότερες θύρες καθώς έχουν περιορισμένο αριθμό γειτόνων. Σημειώνεται ότι η αρχιτεκτονική λύση που παρουσιάζεται στην εργασία αυτή παρέχει τη μέγιστη δυνατή παραμετροποίηση για το σχεδιασμό μιας ετερογενούς τρισδιάστατης NoC αρχιτεκτονικής, προσανατολισμένης στην εφαρμογή.

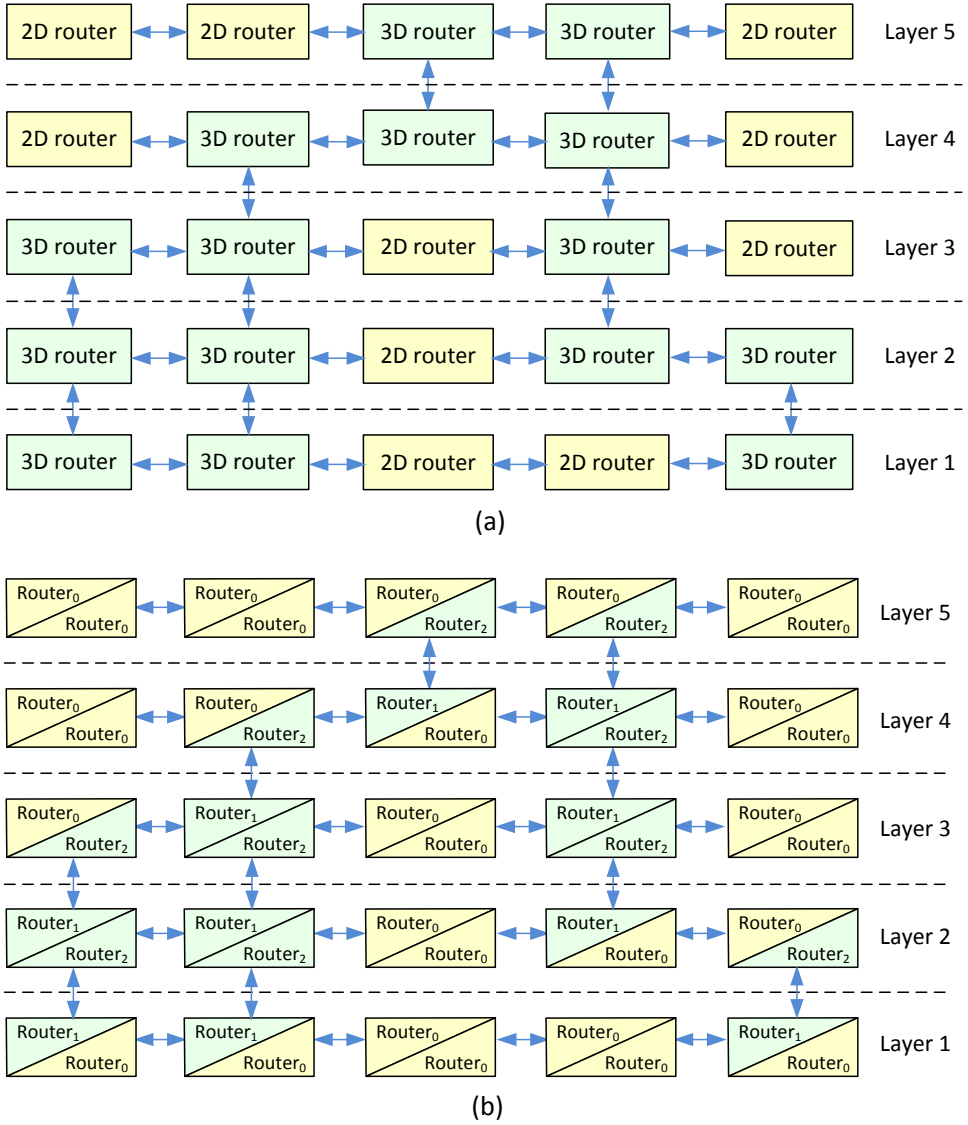
Βασίζόμενοι στην ανάλυση αυτή, εκτός από τον δυσδιάστατο δρομολογητή, ο οποίος αναφέρεται ως $Router_0$, υλοποιούνται τρία ακόμα είδη τρισδιάστατων δρομολογητών:

- $3-D Router_1$: Υποστηρίζει συνδεσιμότητα από τα κατώτερα επίπεδα στα ανώτερα. Καθώς ένα εισερχόμενο πακέτο μπορεί να δρομολογηθεί σε τέσσερις διαφορετικές κατευθύνσεις (σημειώνεται ότι οι θύρες εισόδου και εξόδου δεν μπορούν να συμπίπτουν προκειμένου να αποφεύγονται αδιέξοδα-deadlocks), ο βαθμός ευελιξίας του δρομολογητή είναι $F_S = 4$.
- $3-D Router_2$: Υποστηρίζει συνδεσιμότητα από τα ανώτερα επίπεδα στα κατώτερα. Ομοίως με τον προηγούμενο δρομολογητή, ο συγκεκριμένος δρομολογητής έχει βαθμός ευελιξίας $F_S = 4$.
- $3-D Router_3$: Υποστηρίζει συνδεσιμότητα από τα κατώτερα επίπεδα στα ανώτερα και αντίστροφα. Ο συγκεκριμένος δρομολογητής έχει βαθμός ευελιξίας ($F_S = 5$).

3.3.2.1. Σχεδιασμός 2-D και 3-D δρομολογητών

Το βασικό συστατικό της προτεινόμενης αρχιτεκτονικής είναι ο δρομολογητής NoC. Ο προτεινόμενος δρομολογητής σχεδιάστηκε σε γλώσσα προγραμματισμού SystemC ώστε να είναι παραμετροποιήσιμος κατά το χρόνο σχεδίασης, ενώ παράλληλα να μπορεί να συντεθεί σε υλικό με τη χρήση εργαλείων υψηλού επιπέδου σύνθεσης. Μεταξύ άλλων, παράμετροι όπως ο αριθμός των θυρών, το μέγεθος phit (μήκος λέξης), το μέγεθος flit, τα μεγέθη των buffer για την είσοδο/έξοδο και ο αριθμός των πακέτων καθυστέρησης είναι παραμετροποιήσιμα, μέσω ενός αρχείου λογισμικού. Επιπλέον, καθώς το προτεινόμενο πλαίσιο στοχεύει στη γρήγορη αξιολόγηση ετερογενών τρισδιάστατων συστημάτων NoC, η προτεινόμενη αρχιτεκτονική λύση, η οποία αποτελείται από συνδυασμούς δισδιάστατων και τρισδιάστατων δρομολογητών, περιγραφόμενη σε SystemC είναι πλήρως συνθέσιμη.

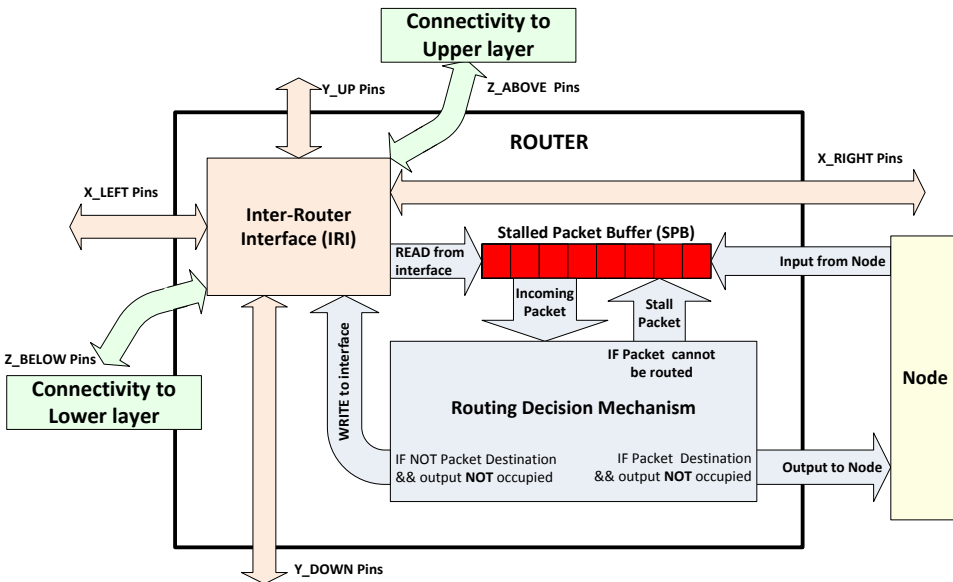
Το Σχήμα 3.17 απεικονίζει την αρχιτεκτονική του προτεινόμενου τρισδιάστατου δρομολογητή. Τα πακέτα εισόδου από τη διεπαφή καθώς και τα αντίστοιχα από τον συνημμένο κόμβο, αποθηκεύονται σε έναν buffer ακίνητων πακέτων. Καθώς μόνο ένα πακέτο δρομολογείται σε κάθε κύκλο ρολογιού ανά θύρα εξόδου, ο buffer των σταθερών πακέτων (SPB) περιέχει εκείνα τα πακέτα που δε μπορούν να δρομολογηθούν στις θύρες εξόδου του δρομολογητή. Παρόλο που



Σχήμα 3.16 Εναλλακτικές αρχιτεκτονικές 3-D NoCs για την εφαρμογή που απεικονίζεται στο Σχήμα 3.15.

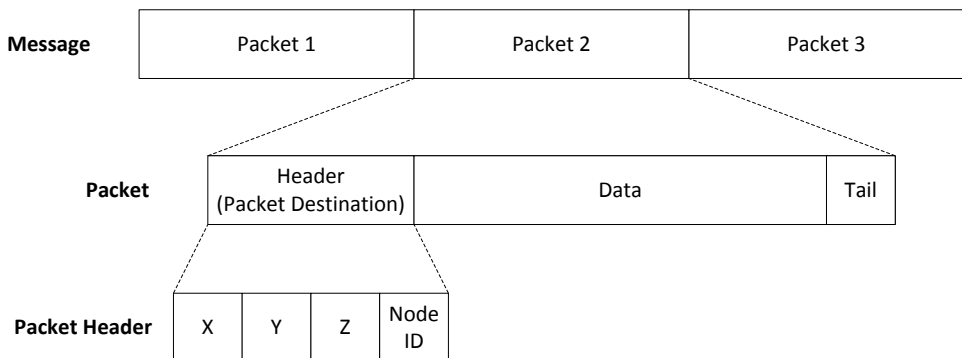
ένας δισδιάστατος δρομολογητής αποτελείται από λιγότερες θύρες (εφόσον δεν υπάρχουν θύρες για συνδεσιμότητα σε υψηλότερα/χαμηλότερα στρώματα), η αρχιτεκτονική οργάνωση ενός δισδιάστατου δρομολογητή είναι παρόμοια με αυτή που συζητήθηκε προηγουμένως. Οι μόνες διαφορές μεταξύ αυτών των δύο αρχιτεκτονικών στιγμιοτύπων επηρεάζουν το μηχανισμό δρομολόγησης (καθώς δεν υπάρχει συνδεσιμότητα σε υψηλότερα/χαμηλότερα στρώματα), καθώς και το μέγεθος του SPB buffer. Πιο συγκεκριμένα, καθώς ο δισδιάστατος δρομολογητής έχει λιγότερες θύρες εξόδου, είναι πιο πιθανό για ένα πακέτο να ακινητοποιηθεί σε αυτόν το δρομολογητή, σε σύγκριση με την αντίστοιχη τρισδιάστατη υλοποίηση. Συνεπώς, το μέγεθος του SPB buffer για το δισδιάστατο δρομολογητή είναι 30% μεγαλύτερο συγκρινόμενο με το αντίστοιχο μέγεθος ενός τρισδιάστατου δρομολογητή.

Επιπλέον, το προτεινόμενο μοντέλο SystemC του δρομολογητή ενσωματώνει ένα ευέλικτο Inter-Router Interface (IRI) προκειμένου να εξασφαλίσει σταθερή και έγκυρη κατεύθυνση δρομολόγησης, ακόμα και σε περιπτώσεις κάποιου σφάλματος. Η υλοποίηση της διεπαφής IRI βασίζεται σε ένα σύνολο από μεθόδους Read/Write σε γλώσσα C++, μία για κάθε διεύθυνση, που ορίζει τη συνδεσιμότητα των θυρών του δρομολογητή με γειτονικούς δρομολογητές. Από άποψη φυσικού επιπέδου σχεδίασης, η διεπαφή IRI παρέχει τους κατάλληλους ακροδέκτες επικοινωνίας με γειτονικούς δρομολογητές, οριζόμενοι ως *X_LEFT*, *X_RIGHT*, *Y_UP*, *Y_DOWN*, καθώς και ως *Z_ABOVE* and *Z_BELOW* για τον τρισδιάστατο δρομολογητή. Σε περίπτωση που ένας δρομολογητής δεν συνδέεται σε γειτονικό κόμβο για δεδομένη κατεύθυνση, τότε η αντίστοιχη μέθοδος ακυρώνει τις αντίστοιχες λειτουργίες Read/Write.



Σχήμα 3.17 Προτεινόμενο αρχιτεκτονικό πρότυπο 3-D δρομολογητή.

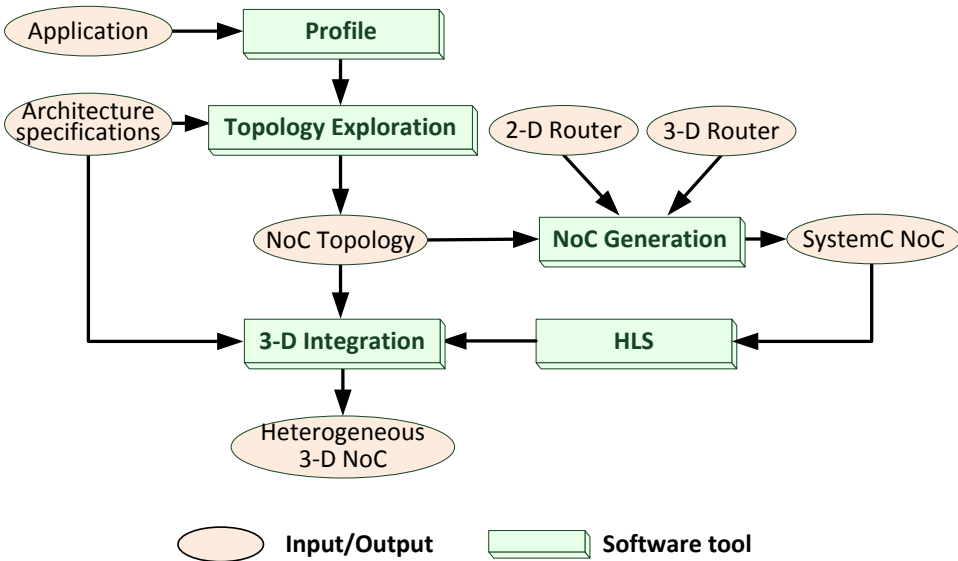
Η επιλογή της κατεύθυνσης εξόδου ενός εισερχόμενου πακέτου ορίζεται από το υποσύστημα “Routing Decision Mechanism” που βασίζεται σε ένα τροποποιημένο αλγόριθμο ZXY. Σύμφωνα με αυτόν τον αλγόριθμο, δίδεται προτεραιότητα στα ακίνητα πακέτα που αποθηκεύονται στον τοπικό SPB buffer του δρομολογητή. Καθώς το μέγεθος του buffer είναι περιορισμένο, ο αλγόριθμος δρομολόγησης ενσωματώνει ένα μηχανισμό για πρόληψη υπερχειλίσις του SPB buffer. Συγκεκριμένα, σε περίπτωση που το πακέτο δε μπορεί να δρομολογηθεί προς την κατεύθυνσή με την ελάχιστη Manhattan απόσταση και ο buffer SPB είναι πλήρως δεσμευμένος, τότε το πακέτο δρομολογείται στη πρώτη ελεύθερη διεύθυνση (θύρα) ώστε να αποφευχθεί απώλεια δεδομένων. Αντίθετα (π.χ. όταν υπάρχουν διαθέσιμες ελεύθερες θύρες εξόδου), ο z-άξονας έχει υψηλότερη προτεραιότητα συγκρινόμενος με τη δρομολόγηση πακέτου στο ίδιο επίπεδο. Επιπρόσθετα, ο υλοποιημένος αλγόριθμος δρομολόγησης υποστηρίζει την αποφυγή live-locks και dead-locks. Πιο λεπτομερώς, η αποφυγή των αδιεξόδων γίνεται ενσωματώνοντας τη τεχνική Turn Model [39]. Στη παρούσα εργασία, οι υλοποιημένοι δρομολογητές αποφεύγουν τα livelocks δίνοντας προτεραιότητα στη κίνηση πακέτων βασιζόμενοι σε μετρητές hop. Για κάθε κόμβο που διασχίζει κάποιο πακέτο, ο μετρητής hop αυξάνει. Αν διάφορα πακέτα ζητούν ένα κανάλι επικοινωνίας, τότε αποκτά πρόσβαση το πακέτο με τη μεγαλύτερη τιμή μετρητή hop. Με τον τρόπο αυτό, πακέτα που έχουν διανύσει μεγαλύτερη απόσταση στο δίκτυο, έχουν μεγαλύτερη προτεραιότητα εξυπηρέτησης και τελικά φτάνουν στο προορισμό. Τέλος, για να δηλωθεί η χωρική ανάθεση κάθε δρομολογητή στο πλέγμα του τρισδιάστατου NoC, ορίζονται επίσης οι συντεταγμένες του ως $\langle X, Y, Z \rangle$. Η πληροφορία για τις συντεταγμένες του δρομολογητή προορισμού βρίσκεται επίσης στην επικεφαλίδα κάθε πακέτου, όπως φαίνεται στο Σχήμα 3.18. Αναφέρεται ότι μπορούν να υποστηριχθούν περισσότερο ευέλικτοι αλγόριθμοι δρομολόγησης από τον προτεινόμενο τροποποιημένο αλγόριθμο ZXY. Η ανάπτυξη του προτεινόμενου πλαισίου σε γλώσσα SystemC διευκολύνει την επεκτασιμότητά του.



Σχήμα 3.18 Δομή του πακέτου της προτεινόμενης αρχιτεκτονικής NoC.

3.3.3. Προτεινόμενη μεθοδολογία σχεδιασμού 3-D NoC

Η υποενοότητα αυτή περιγράφει τη προτεινόμενη μεθοδολογία για τη γρήγορη αξιολόγηση ετερογενών τρισδιάστατων NoCs. Τα βήματα αυτής της μεθοδολογίας απεικονίζονται στο Σχήμα 3.19. Αρχικά, εφαρμόζεται μια προσέγγιση χαρακτηρισμού δύο-βημάτων στη περιγραφή υψηλού επιπέδου της εφαρμογής (π.χ. C/C++). Αυτή η διαδικασία χαρακτηρίζει το εύρος ζώνης επικοινωνίας μεταξύ των διεργασιών που ανατίθενται σε διαφορετικούς κόμβους. Συγκεκριμένα, το πρώτο βήμα αφορά την αλγοριθμική ανάλυση της εκάστοτε εφαρμογής, ώστε να οριστεί η ποσότητα δεδομένων που ανταλλάσσεται μεταξύ των επεξεργαστικών πυρήνων της εφαρμογής ανά κλήση συνάρτησης. Το δεύτερο βήμα στοχεύει στην ανάλυση του επεξεργαστικού φόρτου κάθε συνάρτησης του πηγαίου κώδικα. Για το βήμα αυτό χρησιμοποιείται το εργαλείο λογισμικού Callgrind, της βιβλιοθήκης λογισμικού χαρακτηρισμού εφαρμογών Valgrind [40]. Κατά το βήμα αυτό υπολογίζεται με ακρίβεια το συνολικό μέγεθος δεδομένων που στέλνονται/λαμβάνονται μεταξύ των διεργασιών της εφαρμογής.



Σχήμα 3.19 Προτεινόμενη μεθοδολογία σχεδιασμού 3-D NoCs.

Το αποτέλεσμα αυτής της ανάλυσης, σε συνάρτηση με τις προδιαγραφές του συστήματος (π.χ. επιθυμητή ρυθμαπόδοση, μέγιστη ανεκτή κατανάλωση ενέργειας, κλπ) τροφοδοτούνται ως είσοδοι στο εργαλείο διερεύνησης τοπολογίας (Topology Exploration). Στο βήμα αυτό αξιολογούνται διαφορετικά ετερογενή τρισδιάστατα NoCs ώστε να χαρακτηριστούν οι τοπολογίες οι οποίες είναι Pareto-κυρίαρχες. Επιπλέον, καθώς η εσωτερική επικοινωνία του δρομολογητή είναι πιο αποδοτική συγκρινόμενη με την επικοινωνία μεταξύ των δρομολογητών, το προτεινόμενο πλαίσιο υποστηρίζει την ομαδοποίηση πολλαπλών κόμβων σε ένα μόνο δρομολογητή.

Το πλαίσιο εξερεύνησης λαμβάνει υπόψιν ένα σύνολο τεχνολογικών χαρακτηριστικών, όπως ο περιορισμένος αριθμός TSVs (ένα σύνολο από TSVs καταλαμβάνει σημαντικό μέρος της επιφάνειας πυριτίου [41]) και ο μέγιστος αριθμός επιπέδων πυριτίου για τη τρισδιάστατη ολοκλήρωση. Τροφοδοτώντας αυτούς τους περιορισμούς στο εργαλείο εξερεύνησης, είναι δυνατός ο χαρακτηρισμός μόνο των τεχνολογικά βιώσιμων λύσεων. Η διαδικασία εξερεύνησης τοπολογίας αυτοματοποιείται με ένα εργαλείο λογισμικού, το οποίο έχει προταθεί στην εργασία [42]. Ωστόσο, για τους σκοπούς της παρούσας εργασίας, έχει τροποποιηθεί εκτενώς ο αλγόριθμος διερεύνησης ώστε να λαμβάνεται υπ' όψιν οι επιπλέον περιορισμοί συνδεσιμότητας που τίθενται από την ενσωμάτωση ετερογενών τρισδιάστατων NoCs, σε αντίθεση με την υποστήριξη μόνο ομοιογενών τρισδιάστατων NoCs που υποστηρίζει η υφιστάμενη προσέγγιση. Η τοπολογία ορίζει τη χωρική θέση κάθε δρομολογητή στο πλέγμα των NoCs, καθώς και τον τύπο (π.χ. δισδιάστατος, τρισδιάστατος με συνδεσιμότητα μόνο στο ανώτερο επίπεδο, τρισδιάστατος με συνδεσιμότητα μόνο στο κατώτερο επίπεδο, πλήρως τρισδιάστατος) για κάθε δρομολογητή. Συνδυάζοντας και συνδέοντας κατάλληλα ένα σύνολο δισδιάστατων και τρισδιάστατων δρομολογητών, το προτεινόμενο πλαίσιο χαρακτηρίζει αυτόματα ένα συνθέσιμο μοντέλο SystemC που περιγράφει το ετερογενές τρισδιάστατο NoC.

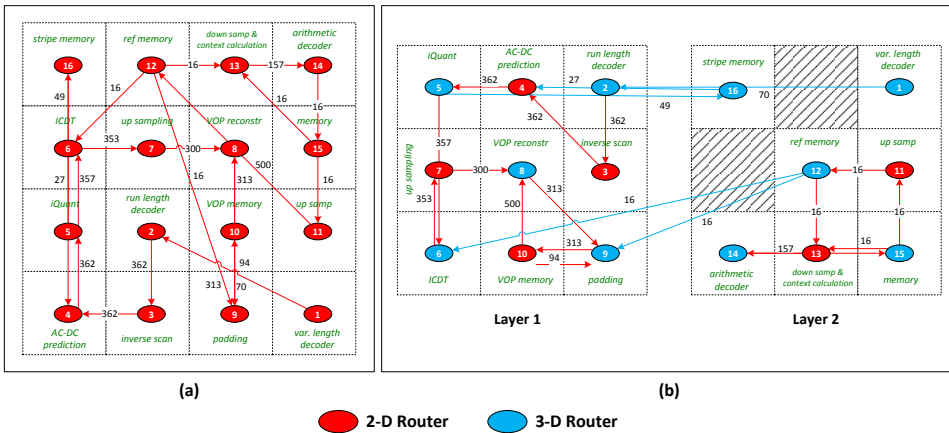
Το επόμενο βήμα αφορά τη φυσική υλοποίηση του σχεδιασμού στην επιλεγόμενη τρισδιάστατη τεχνολογία. Αρχικά, η περιγραφή του συστήματος σε γλώσσα SystemC μεταφράζεται σε επίπεδο RTL με τη χρήση ενός εργαλείου High-Level Synthesis (HLS). Στα πλαίσια της παρούσας εργασίας, χρησιμοποιείται το εργαλείο C-to-Silicon [43]. Τα υπόλοιπα βήματα της προτεινόμενης μεθοδολογίας αφορούν τη σύνθεση της αρχιτεκτονικής καθώς και την ποσοτικοποίηση πολλών μετρικών απόδοσης για την υλοποίηση του τρισδιάστατου συστήματος. Αυτό το πλαίσιο της 3-D υλοποίησης καλύπτεται λεπτομερώς στην Ενότητα 3.5, οπότε και ο αναγνώστης παραπέμπεται σε εκείνο το εδάφιο, ενώ περισσότερες λεπτομέρειες για την εφαρμογή του πλαισίου ολοκλήρωσης 3-D σχετικά με τον προτεινόμενο δρομολογητή μπορούν να βρεθούν στην εργασία [44].

3.3.4. Πειραματικά αποτελέσματα

Η ενότητα αυτή παρέχει ένα σύνολο από πειραματικά αποτελέσματα που αποδεικνύουν την αποδοτικότητα της προτεινόμενης λύσης. Στα πλαίσια της παρούσας εργασίας μελετήθηκαν τέσσερις εφαρμογές ψηφιακής επεξεργασίας σήματος (DSP). Η επιλογή των συγκρινόμενων εφαρμογών έγινε με βάση το γεγονός ότι η απόδοσή τους εξαρτάται σε μεγάλο βαθμό από τη μεταφορά δεδομένων μεταξύ διαφορετικών κόμβων αποθήκευσης και επεξεργασίας. Στη συνέχεια ακολουθεί μια σύντομη περιγραφή για τις υλοποιημένες εφαρμογές.

- Video Object Plane Decoder (VOPD): είναι μία εφαρμογή DSP για επεξεργασία υψηλής-ποιότητας βίντεο με μεγάλο εύρος ζώνης. Ο αποκωδικοποιητής VOPD αποτελείται από 16 κόμβους και εμπεριέχει δύο αποκωδικοποιητές μήκους (length decoders), ένα υποσύστημα πρόβλεψης AC-DC (AC-DC prediction), έναν επεξεργαστή ARM, δύο υποσυστήματα μνήμης και έναν επανα-συνθέτη VOP (VOP reconstructor) [45] [42].

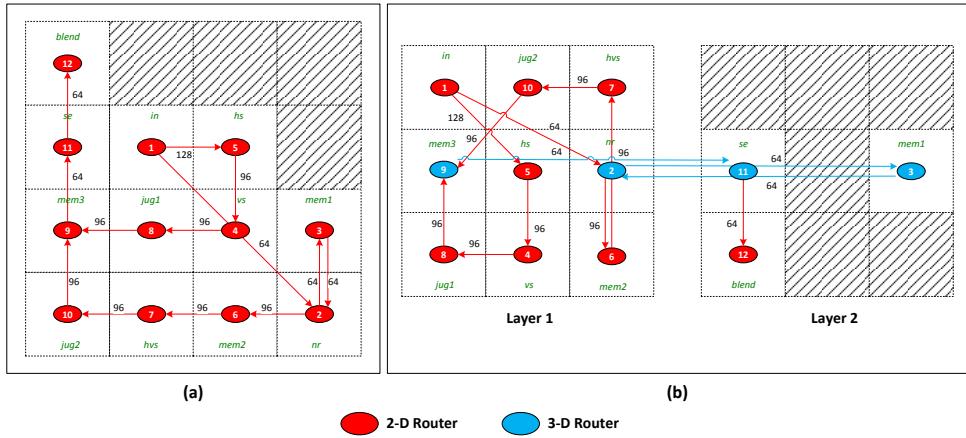
- Multi-Window Display (MWD): είναι μία εφαρμογή DSP κατάλληλη για υλοποίηση σε τοπολογίες NoC [46]. Αυτή η εφαρμογή αποτελείται από 12 κόμβους (επεξεργασίας και αποθήκευσης), καθέννας από τους οποίους αντιστοιχεί σε διαφορετικό δρομολογητή.
- MPEG-4: είναι ένα κοινώς χρησιμοποιούμενο πρωτόκολλο κωδικοποίησης/αποκωδικοποίησης ήχου και εικόνας [47]. Η υλοποίηση MPEG-4 που χρησιμοποιήθηκε σε αυτή την εργασία έχει 12 κόμβους, συμπεριλαμβανομένων διαφόρων επεξεργαστικών στοιχείων, όπως μία υπομονάδα βίντεο, μία υπομονάδα ήχου, ένας επεξεργαστής RISC, μία υπομονάδα CPU, μία υπομονάδα binary alpha και τρεις υπομονάδες μνήμης SRAM. Καθώς ο κωδικοποιητής/αποκωδικοποιητής αποτελείται από αρκετές διακριτές υπομονάδες, η αρχιτεκτονική προσέγγιση NoC είναι κατάλληλη για μελέτη στα πλαίσια της συγκεκριμένης εργασίας.
- MMS: είναι μία εφαρμογή πολυμέσων [33] που αποτελείται από 25 κόμβους, συμπεριλαμβανομένων υπομονάδων μνήμης και επεξεργαστές DSP.



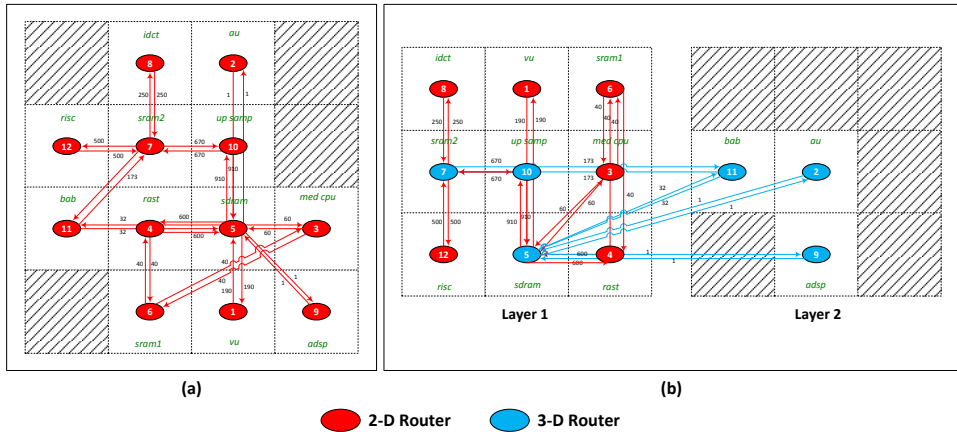
Σχήμα 3.20 Απεικόνιση της εφαρμογής VOPD σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.

Τα Σχήματα 3.20, 3.21, 3.22 και 3.23 απεικονίζουν την αντιστοιχία μεταξύ των τεσσάρων προαναφερθέντων εφαρμογών σε συμβατική διςδιάστατη αρχιτεκτονική NoC καθώς και στην προτεινόμενη ετερογενή τριςδιάστατη αρχιτεκτονική NoC με δύο επίπεδα πυριτίου. Στις υλοποιήσεις αυτές έχει επιλεγθεί ο μικρότερος αριθμός δρομολογητών που απαιτείται για την απεικόνιση των εφαρμογών στις τοπολογίες NoC, υποθέτοντας μία τοπολογία πλέγματος (mesh) και μόνο ένα κόμβο ανά δρομολογητή. Παρόλο που η ανάλυση που παρουσιάζεται στην ενότητα αυτή αφορά δρομολογητές με μόνο μία τοπική θύρα, η προτεινόμενη μεθοδολογία υποστηρίζει επίσης πιο εξελιγμένα σχήματα επικοινωνίας, όπως π.χ. ομαδοποίηση πολλαπλών κόμβων σε ένα μόνο δρομολογητή. Ομοίως

με τη διαδικασία διερεύνησης τοπολογίας, που παρουσιάστηκε στην υποενότητα 3.3.3, η αντιστοίχιση εφαρμογής-κόμβου εκτελείται με μια επεκτεταμένη έκδοση του εργαλείου λογισμικού που παρουσιάζεται στην εργασία [42].

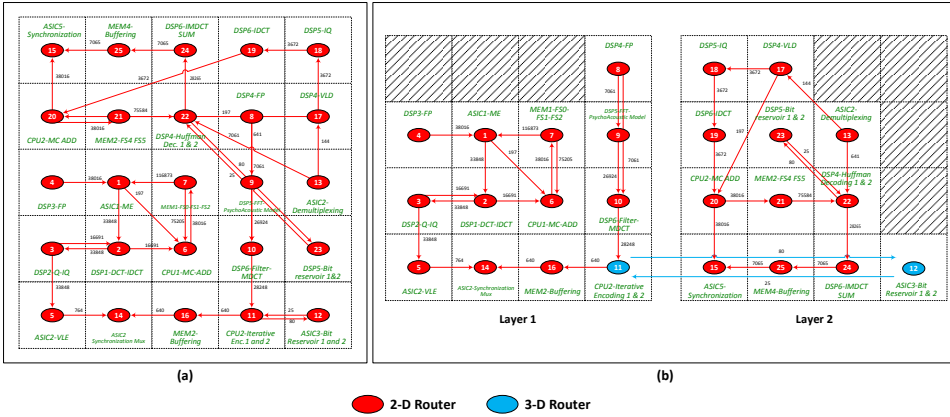


Σχήμα 3.21 Απεικόνιση της εφαρμογής MWD σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.



Σχήμα 3.22 Απεικόνιση της εφαρμογής MPEG-4 σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.

Οι κόκκινοι και μπλε χρωματισμένοι κύκλοι στα παραπάνω σχήματα απεικονίζουν διδιάστατους και τρισδιάστατους δρομολογητές αντίστοιχα. Ομοίως, τα κατευθυνόμενα κόκκινα και μπλε τόξα δείχνουν μεταφορές πακέτων μεταξύ δρομολογητών αφετηρίας και προορισμού που ανατίθενται στο ίδιο καθώς και σε διπλανά επίπεδα. Τα βάρη στα τόξα αυτά αντιστοιχούν στον αριθμό των πακέτων που πρέπει να σταλθούν από τον κόμβο αφετηρίας στο κόμβο προορισμού. Επιπλέον, στα Σχήματα αυτά, υπάρχουν κάποιες σκιασμένες τοποθεσίες



Σχήμα 3.23 Απεικόνιση της εφαρμογής MMS σε: (a) 2-D NoC και (b) την προτεινόμενη ετερογενή πλατφόρμα 3-D NoC.

χωροθέτησης. Σε αυτές τις τοποθεσίες, ο αλγόριθμος αντιστοίχισης δεν έχει αναθέσει κάποιο κόμβο.

προκειμένου να ποσοτικοποιηθεί η αποδοτικότητα των τρισδιάστατων ετερογενών αρχιτεκτονικών NoC, παρέχεται μία ανάλυση που στοχεύει στην αξιολόγηση των υλοποιημένων δισδιάστατων και τρισδιάστατων δρομολογητών καθώς αλλάζει ο αριθμός των τοπικών θυρών. Σημειώνεται ότι η περίπτωση $Node = 0$ αντιστοιχεί στο σενάριο όπου ο δρομολογητής είναι συνδεδεμένος σε τέσσερις δισδιάστατους ή έξι τρισδιάστατους γειτονικούς δρομολογητές. Σκοπός της ανάλυσης αυτής είναι να αποδείξει ότι τα μοντέλα SystemC για τρισδιάστατους δρομολογητές παρουσιάζουν παρόμοια απόδοση με τους αντίστοιχους δισδιάστατους δρομολογητές, παρόλο που η επιπλέον λογική που απαιτείται για τη πραγματοποίηση της δρομολόγησης πακέτου σε περισσότερες της μία διευθύνσεις (θύρες εξόδου) προσθέτει καθυστέρηση στην επικοινωνία. Συγκεκριμένα, το Σχήμα 3.24 απεικονίζει τον αριθμό των λογικά ισοδύναμων πυλών ύστερα από το στάδιο της σύνθεσης των δύο αυτών δρομολογητών, με τη χρήση της τεχνολογίας σχεδίασης φυσικού επιπέδου TMS320C45nm. Σύμφωνα με τα αποτελέσματα που συνοψίζονται στο σχήμα αυτό, συμπεραίνεται ότι, καθώς αυξάνει ο αριθμός των λογικών θυρών, υπάρχει μια γραμμική σχεδόν αύξηση του αριθμού των λογικών πυλών. Επιπλέον, οι υλοποιήσεις τόσο των δισδιάστατων όσο και των τρισδιάστατων δρομολογητών δεσμεύουν παρόμοιο αριθμό λογικών πυλών για δεδομένο αριθμό λογικών θυρών. Επίσης, δύο ακόμα ενδιαφέροντα συμπεράσματα μπορούν να εξαχθούν από το σχήμα αυτό. Συγκεκριμένα, παρόλο που ο τρισδιάστατος δρομολογητής εμπεριέχει επιπλέον λογική για την εκτέλεση δρομολόγησης πακέτου (δηλαδή έχει περισσότερους γείτονες συγκριτικά με τη δισδιάστατη υλοποίηση), εντούτοις, ενσωματώνει μικρότερου μεγέθους SPB buffers, όπως συζητήθηκε στην υποενότητα 3.3.2. Αυτή η διαπίστωση απεικονίζεται στο Σχήμα 3.24 για τρισδιάστατους δρομολογητές αποτελούμενους από μια ή δύο λογικές θύρες. Θεωρώντας μεγαλύτερο αριθμό λογικών πυλών, η πολυπλοκότητα

τητα του μηχανισμού “Routing Decision Mechanism” αυξάνεται και συνεπώς οι τρισδιάστατοι δρομολογητές χρειάζονται επιπλέον λογικές πύλες. Εκτός από τα αποτελέσματα που συνοφίζονται στο σχήμα αυτό, είναι εφικτή περαιτέρω μείωση του αριθμού των πυλών εφαρμόζοντας ένα εξειδικευμένο σχεδιασμό (full-custom) του δρομολογητή, αντί για την υλοποίηση με χρήση υψηλού επιπέδου μοντέλων SystemC. Ωστόσο σε αυτή την εργασία εξετάζεται η ανάπτυξη ενός πλαισίου γρήγορης αξιολόγησης ετερογενών τρισδιάστατων NoCs με τη χρήση εργαλείων HLS.

Σχήμα 3.24 Αξιολόγηση δισδιάστατων και τρισδιάστατων δρομολογητών αναφορικά προς τον αριθμό των δεσμευμένων λογικών πυλών, σε σχέση με τον αριθμό των θηρών E/E.

Η διαφορετική αρχιτεκτονική οργάνωση μεταξύ δισδιάστατων και τρισδιάστατων δρομολογητών, τόσο στο μηχανισμό δρομολόγησης όσο και στους SPB buffers, οδηγεί σε διακυμάνσεις της απόδοσης. Τα επόμενα δύο σχήματα ποσοτικοποιούν την απόδοση των δρομολογητών αυτών αναφορικά προς τη καθυστέρηση και την ενέργεια για διαφορετικό αριθμό λογικών θυρών, υποθέτοντας δρομολογητές με παρόμοια κίνηση για τη δισδιάστατη και τη τρισδιάστατη αρχιτεκτονική προσέγγιση. Σύμφωνα με το Σχήμα 3.25, συμπεραίνεται ότι ένας τρισδιάστατος δρομολογητής παρουσιάζει 32% μικρότερη καθυστέρηση συγκρινόμενος με τη δισδιάστατη υλοποίηση. Αυτό η μέτρηση δικαιολογείται λόγω του σημαντικά μικρότερου (30%) μεγέθους του SPB buffer που υλοποιείται στο τρισδιάστατο πεδίο. Στο σχήμα αυτό παρατηρούνται επίσης κάποιες αποκλίσεις (μη-μονοτονικές βελτιώσεις), οι οποίες αποδίδονται στη χρήση του εργαλείου HLS.

Το Σχήμα 3.26 απεικονίζει την κατανάλωση ενέργειας για τα διαφορετικά είδη δρομολογητών που παρουσιάστηκαν προηγουμένως. Από την ανάλυση αυτή, συμπεραίνεται ότι η μέση κατανάλωση ενέργειας για τρισδιάστατους και δισδιά-

Σχήμα 3.25 Αξιολόγηση δισδιάστατων και τρισδιάστατων δρομολογητών αναφορικά προς τη καθυστέρηση, σε σχέση με τον αριθμό των θηρών E/E.

στατους δρομολογητές, για τις εφαρμογές πολυμέσων που χρησιμοποιούνται σε αυτή την εργασία, είναι 36pJ και 49pJ, αντίστοιχα. Ομοίως με τη καθυστέρηση, η επιπλέον κατανάλωση ενέργειας κατά 38% στους δισδιάστατους δρομολογητές αποδίδεται στο γεγονός ότι οι τρισδιάστατες αρχιτεκτονικές ενσωματώνουν μικρότερους SPB buffers.

Στη συνέχεια ποσοτικοποιείται η φυσική υλοποίηση των εναλλακτικών αρχιτεκτονικών NoC. Για την ανάλυση αυτή, οι τέσσερις DSP εφαρμογές που παρουσιάστηκαν προηγουμένως συντέθηκαν με τη χρήση του προτεινόμενου πλαισίου και του εργαλείου Cadence SoC Encounter. Τρεις διαφορετικές αρχιτεκτονικές λύσεις NoC αξιολογούνται σε αυτή την ανάλυση, ως εξής: (i) ομοιόμορφο δισδιάστατο NoC, (ii) ομοιόμορφο τρισδιάστατο NoC και (iii) προτεινόμενο ετερογενές τρισδιάστατο NoC. Η φυσική υλοποίηση έγινε με χρήση της τεχνολογίας TMS320C45nm, ενώ οι παράμετροι για την υλοποιημένη βιβλιοθήκη TSV συνοψίζονται στον Πίνακα 3.2 [41]. Το Σχήμα 3.27 απεικονίζει τη χωροθέτηση, καθώς και τη φυσική υλοποίηση για μία από τις εφαρμογές αξιολόγησης, τον κωδικοποιητή/αποκωδικοποιητή MPEG-4. Στο Σχήμα 3.27(a), τονίζονται τα δυο *Εικονικά Επίπεδα* (Virtual Layers), ενώ οι αντίστοιχες συνδέσεις των φυσικών καλωδίων μεταξύ των επιπέδων αυτών απεικονίζονται με κίτρινο χρώμα στο Σχήμα 3.27(b).

Σε αντίθεση με την απεικόνιση της εφαρμογής στους κόμβους του NoC, όπου στόχος βελτιστοποίησης είναι η ελαχιστοποίηση του αριθμού των packet hops, κατά το στάδιο της φυσικής υλοποίησης η στόχευση είναι η ελαχιστοποίηση του κρίσιμου μονοπατιού καθυστέρησης, έτσι ώστε να επιτευχθεί η μέγιστη απόδοση επικοινωνίας. Συνεπώς, οι μετρικές απόδοσης για τα δύο είδη των τρισδιάστα-

Σχήμα 3.26 Αξιολόγηση δισδιάστατων και τρισδιάστατων δρομολογητών αναφορικά προς τη κατανάλωση ενέργειας, σε σχέση με τον αριθμό των θηρών E/E.

Πίνακας 3.2 Χαρακτηριστικά της επιλεγμένης τεχνολογίας TSV.

Διάμετρος:	1.2um
Ελάχιστο βήμα (Pitch):	4um
Αντίσταση:	0.35Ω
Χωρητικότητα:	2.5fF
Μήκος:	4-9um

των NoCs που αναλύονται σε αυτήν την εργασία (ομοιογενές και ετερογενές) διαφέρουν σημαντικά, κυρίως λόγω των διακυμάνσεων στη συνδεσιμότητα που επιβάλλονται από τον διαφορετικό αριθμό των χρησιμοποιημένων TSVs.

Ο Πίνακας 3.3 παρέχει ένα σύνολο παραμέτρων σχετικών με τις εφαρμογές αξιολόγησης, μετά από το στάδιο της σύνθεσης και της φυσικής υλοποίησης. Τόσο η τρισδιάστατη όσο και η προτεινόμενη ετερογενή λύση έχουν αναλογία διαστάσεων φυσικού επιπέδου ίση με δύο (1.99), προκειμένου να υλοποιηθούν στην ίδια ψηφίδα (die) τα δύο εικονικά στρώματα, όπως παρουσιάστηκαν προηγουμένως στο Σχήμα 3.27. Εφόσον το προτεινόμενο ετερογενές τρισδιάστατο NoC αποτελείται από ένα σύνολο δισδιάστατων και τρισδιάστατων δρομολογητών, έχει λιγότερα TSVs συγκρινόμενο με μια ομοιόμορφη τρισδιάστατη υλοποίηση NoC. Ωστόσο, αυτή η σχεδιαστική επιλογή δεν αυξάνει το συνολικό μήκος καλωδίου καθώς η προτεινόμενη ετερογενή προσέγγιση οδηγεί σχεδόν σε μέση μείωση κατά 57% του μήκους καλωδίου συγκρινόμενη με τη περίπτωση όπου όλοι οι δρομολογητές είναι τρισδιάστατοι.

Σχήμα 3.27 Φυσική απεικόνιση της προτεινόμενης ετερογενούς αρχιτεκτονικής 3-D NoC για την εφαρμογή MPEG-4.

Η μειωμένη συνδεσιμότητα στον κατακόρυφο άξονα, της προτεινόμενης αρχιτεκτονικής λύσης, οδηγεί σε μικρή αύξηση του αριθμού των packet hops, προκειμένου να επιτευχθεί δρομολόγηση μεταξύ των κόμβων αφετηρίας και προορισμού. Το Σχήμα 3.28 παρουσιάζει το συνολικό αριθμό των hops ανά εφαρμογή για τις τρεις εναλλακτικές τοπολογίες NoC. Με βάση αυτή την αξιολόγηση συμπεραίνεται ότι η αρχιτεκτονική λύση των ομοιόμορφων τρισδιάστατων NoCs, όπου όλοι οι δρομολογητές παρέχουν συνδεσιμότητα σε ανώτερο/κατώτερο επίπεδο, είναι η βέλτιστη τοπολογία, καθώς μειώνει τον αριθμό των packet hops σε σύγκριση με την αντίστοιχη υλοποίηση δισδιάστατου NoC, κατά 29%. Αντίθετα, το προτεινόμενο ετερογενές NoC οδηγεί σε αύξηση των packet hops συγκριτικά με το ομοιόμορφο τρισδιάστατο NoC κατά 8.5%.

Το Σχήμα 3.29 απεικονίζει τη μέγιστη συχνότητα λειτουργίας για τις εναλλακτικές τοπολογίες. Ως φαίνεται, το προτεινόμενο ετερογενές τρισδιάστατο NoC

Πίνακας 3.3 Μετρικές υλοποίησης των εφαρμογών αξιολόγησης.

Υλοποίηση		Μετρικές						
		Επιφάνεια (um ²)	Αν/γία	Μήκος καλ.	#TSVs	#κελιών	#καλωδ.	#E/E
MMS	2-D	991.5×990.3	0.99	10,261,064	0	458,823	462,825	3,251
	3-D	599.3×1,198.2	1.99	11,480,859	704	324,689	326,317	3,251
	Προστ/νη	762.8×1,523.3	1.99	7,864,344	130	534,731	539,864	3,251
MPEG-4	2-D	754.74×753.48	0.99	5,915,764	0	256,980	259,307	1,561
	3-D	432.03×861.84	1.99	6,373,170	576	167,911	168,709	1,561
	Προστ/νη	561.7×1,122.66	1.99	3,882,465	192	271,828	275,552	1,561
MWD	2-D	758.9×756.0	0.99	5,566,222	0	262,609	265,219	1,561
	3-D	435.4×870.6	1.99	5,794,055	1,170	171,979	172,762	1,561
	Προστ/νη	551.5×1,101.24	1.99	3,673,350	260	272,823	275,523	1,561
VOPD	2-D	781.3×777.4	0.99	6,156,379	0	283,098	285,961	2,081
	3-D	441.3×885.7	2.00	7,081,633	1,170	178,034	179,182	2,081
	Προστ/νη	585.8×1,170.5	1.99	4,121,908	650	318,237	320,404	2,081

Σχήμα 3.28 Αριθμός packet hops για διαφορετικές αρχιτεκτονικές προσεγγίσεις: (i) ομογενές 2-D NoC, (ii) ομογενές 3-D NoC και (iii) προτεινόμενο ετερογενές 3-D NoC.

έχει υψηλότερες επιδόσεις από τις υπόλοιπες υλοποιήσεις για όλες τις εφαρμογές αξιολόγησης. Ωστόσο, καθώς η τελική απόδοση εξαρτάται από τις εγγενείς ιδιότητες κάθε εφαρμογής, όπως επίσης και από την απεικόνιση της εφαρμογής στους κόμβους της αρχιτεκτονικής NoC, παρατηρούνται ιδιαίτερες διακυμάνσεις στην απόδοση της προτεινόμενης ετερογενούς λύσης και της ομοιόμορφης τρισδιάστατης προσέγγισης. Συγκεκριμένα, παρατηρείται διακύμανση της τάξης του 1.16× έως 1.43×. Ένα ακόμα αξιοσημείωτο συμπέρασμα που απεικονίζεται σε αυτό το σχήμα είναι ότι οι δύο αρχιτεκτονικές NoC επιτυγχάνουν, κατά μέσο όρο, την ίδια σχεδόν μέγιστη συχνότητα λειτουργίας. Αυτή η παρατήρηση δικαιολογείται από το γεγονός ότι ο αυξημένος αριθμός κάθετων συνδέσεων οδηγεί σε κορεσμό και η εισαγωγή επιπλέον TSVs δεν επιφέρει επιπλέον βελτίωση στην απόδοση της υλοποίησης.

Το Σχήμα 3.30 απεικονίζει τη κατανάλωση ενέργειας για εναλλακτικές τοπολογίες NoC. Από το Σχήμα αυτό συμπεραίνουμε ότι, κατά μέσο όρο, η προτεινόμενη ετερογενής λύση, αποτελούμενη από ένα σύνολο δισδιάστατων και τρισδιάστατων δρομολογητών, πετυχαίνει σημαντική εξοικονόμηση ενέργειας συγκριμένη με τα ομοιόμορφα δισδιάστατα και τρισδιάστατα NoCs. Συγκεκριμένα, κα-

Σχήμα 3.29 Μέγιστη συχνότητα λειτουργίας για διαφορετικές αρχιτεκτονικές προσεγγίσεις NoC.

θώς οι τρισδιάστατοι δρομολογητές έχουν λιγότερους SPB buffers, εξοικονομούν σημαντική ενέργεια. Επιπρόσθετα, τα ετερογενή τρισδιάστατα NoCs πλεονεκτούν αναφορικά προς το μήκος καλωδίου, όπως φαίνεται και στον Πίνακα 3.3. Το μικρότερο μήκος καλωδίου ελαχιστοποιεί επιπλέον τη κατανάλωση ενέργειας στα τρισδιάστατα NoCs.

Σχήμα 3.30 Power consumption for different instantiations of NoC.

3.3.5. Συμπεράσματα

Σε αυτήν την ενότητα παρουσιάστηκε ένα πλαίσιο που στοχεύει στη ταχεία αξιολόγηση ετερογενών τρισδιάστατων NoCs. Το εισαγόμενο αρχιτεκτονικό πρότυπο αποτελείται από συνδυασμό δισδιάστατων και τρισδιάστατων δρομολογητών, οι οποίοι προσαρμόζονται στις απαιτήσεις των εφαρμογών μέσω κατάλληλου λογισμικού διερεύνησης του χώρου λύσεων. Τα πειραματικά αποτελέσματα της προτεινόμενης προσέγγισης σε εφαρμογές DSP αποδεικνύουν την αποτελεσματικότητά της, σε σύγκριση με τις υφιστάμενες λύσεις ομοιόμορφων δισδιάστατων και τρισδιάστατων NoCs, καθώς επετεύχθη, κατά μέσο όρο, 25% μεγαλύτερη συχνότητα λειτουργίας και 39% λιγότερη κατανάλωση ενέργειας.

3.4. Συστήματα Διαστημικών Εφαρμογών

3.4.1. Εισαγωγή

Η εξερεύνηση του Άρη είναι ένας από τους κύριους στόχους της Εθνικής Υπηρεσίας Αεροναυτικής και Διαστήματος (NASA) και της Ευρωπαϊκής Υπηρεσίας Διαστήματος (ESA), όπως δείχνουν οι προηγούμενες και πρόσφατες δραστηριότητές τους. Τα τελευταία 16 χρόνια υπάρχουν πολλές αποστολές στον Άρη, τόσο σε τροχιά όσο και στην επιφάνειά του πλανήτη (π.χ., NASA's Mars Global Surveyor, Mars Odyssey, Phoenix, Mars Reconnaissance Orbiter, Mars Express) οι οποίες έχουν σημειώσει αξιοσημείωτα αποτελέσματα. Μία από τις μεγαλύτερες προκλήσεις σε αυτές τις αποστολές είναι ο σχεδιασμός αυτόνομων ρομποτικών συστημάτων. Αφού η αποδοτικότητα αυτών των συστημάτων είναι στενά συνυφασμένη με την ακρίβεια των αλγορίθμων που λαμβάνουν τις αποφάσεις κίνησης [48], υπάρχει μια συνεχής προσπάθεια για την ανάπτυξη ακόμα πιο αποδοτικών αλγορίθμων μηχανικής όρασης (computer vision, CV).

Παρόλο που προτείνονται συνεχώς ελκυστικότερες λύσεις, σημειώνοντας αξιοσημείωτη αποδοτικότητά όσον αφορά την ακρίβεια υπολογισμού, εντούτοις, η αυξημένη ζήτηση των λύσεων αυτών για υπολογιστικούς πόρους και αποθηκευτικό χώρο, τις κάνει ακατάλληλες για διαστημικές αποστολές, δεδομένης της περιορισμένης διαθέσιμης τεχνολογίας η οποία μπορεί να είναι λειτουργική σε εξωγήινα περιβάλλοντα. Αυτή η τεχνολογία χαρακτηρίζεται από μεγάλη κλίμακα ολοκλήρωσης και υψηλά επίπεδα τάσεων κατωφλίου ώστε να καθίσταται αδύνατη η μεταβολή των λογικών δυαδικών στάθμεων (0-1) από την κοσμική ακτινοβολία. Οπότε τα ολοκληρωμένα κυκλώματα που μπορούν να λειτουργήσουν στο διάστημα είναι περιορισμένης επιφάνειας πυριτίου και συνεπώς περιορισμένης συστημικής πολυπλοκότητας. Επίσης, η πλειοψηφία των αλγορίθμων μηχανικής όρασης υλοποιούνται κυρίως σε επίπεδο λογισμικού (π.χ. Matlab, C/C++) συνδυάζοντας σύνθετες βιβλιοθήκες μηχανικής όρασης (π.χ. OpenCV), και επομένως η αυτούσια μεταφορά τους σε υλικό είναι μία ιδιαίζουσα δύσκολη και χρονοβόρα διαδικασία.

Στην παρούσα εργασία εξετάζουμε μια καινούρια μεθοδολογία για αποδοτική υλοποίηση αλγορίθμων μηχανικής όρασης που στοχεύει στην υποστήριξη αυτόνομης πλοήγησης ρομποτικών συστημάτων για εφαρμογές διαστήματος. Η προτεινόμενη λύση είναι μέρος του έργου SPARTAN (SPARing Robotics Technologies for Autonomous Navigation) του Ευρωπαϊκού Οργανισμού Διαστήματος (ESA), ως μέρος της αποστολής ExoMars (προγραμματισμένη για το 2018) [49]. Αντί των υφιστάμενων προσεγγίσεων που επιχειρούν να λύσουν τα προβλήματα αυξημένης υπολογιστικής πολυπλοκότητας και χρόνου εκτέλεσης μόνο στο επίπεδο του λογισμικού, η προτεινόμενη λύση εισηγείται το συν-σχεδιασμό υλικού/λογισμικού σε μία ετερογενή πλατφόρμα που αποτελείται από μια επαναδιαμορφούμενη συσκευή (FPGA) και μια κεντρική επεξεργαστική μονάδα (CPU). Σκοπός του συστήματος είναι η επιτάχυνση των αλγορίθμων μηχανικής όρασης στην επαναδιαμορφούμενη πλατφόρμα, ενώ ο συντονισμός όλων των εργασιών πραγματοποιείται από τη κεντρική μονάδα επεξεργασίας. Οι προδιαγραφές και των δύο συσκευών, οι οποίες προσδιορίζονται από την ESA, αποτε-

λούν τις εκτιμήσεις της ίδιας για τη διαθεσιμότητα των αντίστοιχων συσκευών, σε τεχνολογία διαστήματος, για το έτος 2018, το οποίο αποτελεί ορόσημο εκκίνησης της αποστολής Exomars. Σύμφωνα επομένως με αυτές τις εκτιμήσεις, σε αυτή την εργασία επιλέχθηκε ως επαναδιαμορφούμενη συσκευή μία Virtex-6 FPGA πλατφόρμα και ως σύστημα γενικού σκοπού μία πλατφόρμα με ενσωματωμένο επεξεργαστή ισχύος 150MIPS, με το λειτουργικό σύστημα Robotic Operating System (ROS) [50], μία διανομή Ubuntu Linux ειδικά σχεδιασμένη για ρομποτικές εφαρμογές. Επιπλέον λεπτομέρειες για το έργο SPARTAN και το ανεπτυγμένο σύστημα μπορούν να βρεθούν στη δημοσιευμένη εργασία της ομάδας σχεδίασης [51].

Εφόσον η αρχιτεκτονική του συστήματος είναι μια ετερογενής πλατφόρμα, προτείνεται μια μεθοδολογία συν-σχεδιασμού υλικού/λογισμικού, που υποστηρίζεται από ένα σύνολο υφιστάμενων και νέων εργαλείων. Κρίσιμο σημείο της μεθοδολογίας αυτής είναι ο καθορισμός του μέσου επεξεργασίας στο οποίο θα ανατεθεί ο κάθε ένας εκ των αλγορίθμων CV. Προς αυτή τη κατεύθυνση εισηγείται μία νέα μεθοδολογία χαρακτηρισμού των αλγορίθμων, η οποία παρέχει συγκεκριμένες οδηγίες για τη σημασία τοποθέτησης των διαφορετικών αλγορίθμων CV, ανάλογα με τις σχεδιαστικές παραμέτρους. Συγκεκριμένα, οι επιλογές ανάθεσης βασίζονται σε συνδυασμό της υπολογιστικής πολυπλοκότητας, του τύπου των μαθηματικών υπολογισμών, του ίχνους μνήμης (χρόνος ζωής δεδομένων και χώρος αποθήκευσης), του μεγέθους των δεδομένων που πρέπει να μεταφερθεί μεταξύ αυτών των αλγορίθμων, και της εξάρτησης των δεδομένων για κάθε αλγόριθμο CV.

Παρόλο που η διαδικασία χαρακτηρισμού είναι ένα γνωστό πρόβλημα για τους σχεδιαστές συστημάτων, η κύρια διαφοροποίηση της προτεινόμενης λύσης, σε σχέση με σχετικές προσεγγίσεις, εκτός της εισηγούμενης διαδικασίας συν-σχεδιασμού που αναφέρεται στην προηγούμενη παράγραφο, σημειώνεται ως προς την ελαχιστοποίηση του χρόνου εκτέλεσης. Συγκεκριμένα, στα πλαίσια του έργου SPARTAN, διαπιστώθηκε ότι ο χρόνος μίας μεμονωμένης εξομοίωσης μπορεί να διαρκέσει από μερικά δευτερόλεπτα έως αρκετές ημέρες ανάλογα με τις παραμέτρους της εξομοίωσης. Λαμβάνοντας υπόψιν ότι ένα ρομποτικό σύστημα πρέπει να δοκιμασθεί σε διαφορετικές περιβαλλοντικές συνθήκες, ενώ επίσης πρέπει να εντοπισθούν οι παράμετροι των αλγορίθμων CV που μεγιστοποιούν την απόδοση όλου του συστήματος, το πλήθος των σεναρίων εξομοίωσης μπορεί να αυξηθεί σε τέτοιο επίπεδο, όπου οι σχεδιαστές πολλές φορές αναγκάζονται να προβούν σε εκπτώσεις του συνόλου των εξομοιώσεων, είτε δοκιμάζοντας λιγότερες παραμέτρους, είτε λιγότερα περιβαλλοντικά σενάρια λειτουργίας. Με αυτόν τον τρόπο περιορίζεται η βελτιστοποίηση του συστήματος καθώς υιοθετούνται υπο-βέλτιστες λύσεις. Στην προτεινόμενη εργασία γίνεται προσπάθεια μείωσης του χρόνου εξομοίωσης. Για τους αλγόριθμους CV που μελετήθηκαν σε αυτή την εργασία διαπιστώθηκε ότι ο μεγάλος χρόνος εξομοίωσης οφείλεται στην αυξημένη ανάλυση των εικόνων εισόδου. Συνεπώς εισηγείται μία τεχνική που εγγυάται ακριβή συμπεράσματα με τη χρήση μικρότερου μεγέθους εικόνων. Αυτή η προσέγγιση μειώνει σημαντικά το χρόνο εκτέλεσης κατά τη διαδικασία του χαρακτηρισμού των εφαρμογών.

Η υπόλοιπη εργασία οργανώνεται ως εξής: Η υποενότητα 3.4.2 περιγράφει τη βασική αρχιτεκτονική του συστήματος SPARTAN. Η προτεινόμενη μεθοδολογία περιγράφεται στην υποενότητα 3.4.3, ενώ τα πειραματικά αποτελέσματα που προκύπτουν εφαρμόζοντας τη προτεινόμενη μεθοδολογία στο σύστημα SPARTAN παρουσιάζονται στην υποενότητα 3.4.4. Τέλος, τα συμπεράσματα της παρούσας εργασίας συνοψίζονται στην υποενότητα 3.4.5.

3.4.2. Η αρχιτεκτονική του συστήματος SPARTAN

Στόχος του SPARTAN συστήματος είναι η μετατροπή της οπτικής πληροφορίας από τις κάμερες του ρομποτικού οχήματος σε τρισδιάστατους τοπικούς χάρτες καθώς επίσης και ο ακριβής υπολογισμός τοποθεσίας για τη διαδικασία πλοήγησης. Για αυτό το σκοπό, επιλέγονται κατάλληλοι αλγόριθμοι επεξεργασίας εικόνας για την τρισδιάστατη ανακατασκευή του χάρτη περιβάλλοντος (3D map reconstruction) και εντοπισμό (localization) και στη συνέχεια υλοποιούνται με - όσο το δυνατόν - παράλληλη επεξεργαστική ροή για την επίτευξη υψηλής απόδοσης, ενώ διατηρείται η αποτελεσματικότητα του σχεδιασμού αναφορικά προς την υπολογιστική πολυπλοκότητα, τη δέσμευση μνήμης και την κατανάλωση ενέργειας. Στα πλαίσια του έργου SPARTAN, η επιλογή των αλγορίθμων CV γίνεται από την ερευνητική ομάδα PME-DUTH [52].

Όπως αναφέρθηκε και στην υποενότητα 3.4.1, η αρχιτεκτονική του συστήματος SPARTAN αποτελείται από ένα σύστημα υποδοχής (PC host) με CPU μικρής ισχύος (έως 150MIPS) και μια συσκευή Virtex-6 FPGA. Πιο συγκεκριμένα, το σύστημα υποδοχής μέσω του λειτουργικού συστήματος ROS παρέχει όλες τις συντονιστικές υπηρεσίες που απαιτούνται για την πλοήγηση του ρομποτικού οχήματος. Συγκεκριμένα το λειτουργικό σύστημα είναι υπεύθυνο για την παροχή μηχανισμών επιπέδου αφαιρετικότητας του υλικού, την υποστήριξη χαμηλού επιπέδου έλεγχου των συσκευών μέσω προγραμμαμάτων οδήγησης, την μεταφορά μηνυμάτων μεταξύ των διεργασιών, και την ενιαία διαχείριση του λογισμικού, ενώ η FPGA λειτουργεί ως επιταχυντής για τους υπολογιστικά απαιτητικούς αλγορίθμους CV.

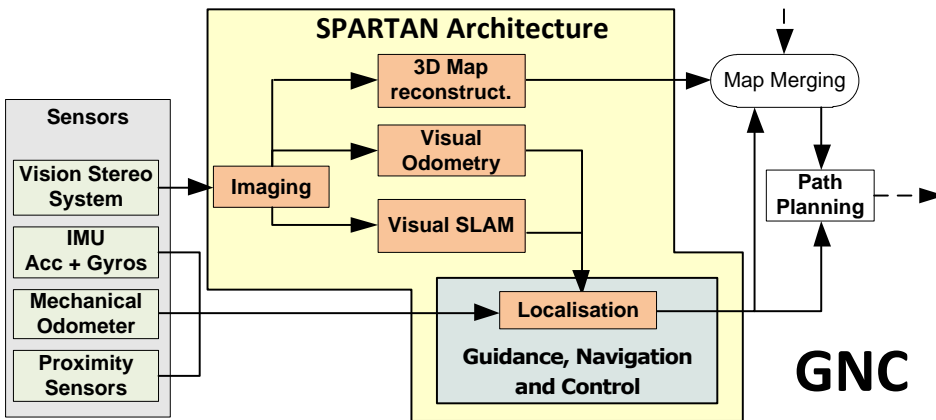
Η λειτουργικότητα του συστήματος SPARTAN διακρίνεται σε δυο διαφορετικούς τρόπους λειτουργίας, τη χαρτογράφηση και τον εντοπισμό. Όταν το σύστημα εκτελεί λειτουργία χαρτογράφησης, παράγει ένα τρισδιάστατο χάρτη του περιβάλλοντος του, ενώ όταν εκτελεί λειτουργία εντοπισμού, το σύστημα παράγει εκτιμήσεις θέσεώς του στον χαρτογραφημένο χάρτη. Ως είσοδος στο σύστημα SPARTAN, χρησιμοποιείται στερεοσκοπική πληροφορία εικόνας από δυο διαφορετικά ζεύγη καμερών. Τα μεγέθη των εικόνων από τις κάμερες αυτές είναι 1120×1120 και 512×384 για τη λειτουργία της χαρτογράφησης και του εντοπισμού, αντίστοιχα.

Γενικά το σύστημα SPARTAN χρησιμοποιεί τρεις βασικές μονάδες επεξεργασίας κατά την εκτέλεση των δύο βασικών λειτουργιών. Η πρώτη αναφέρεται ως τρισδιάστατη ανακατασκευή του χάρτη [53] και περιλαμβάνει το στάδιο του υπολογισμού του χάρτη βάθους (disparity map) και την εύρεση των συντεταγμένων των αντικειμένων στον τρισδιάστατο χώρο. Η δεύτερη μονάδα είναι η οπτική οδομετρία (visual odometry), η οποία περιλαμβάνει τον αλγόριθμο SURF [54]

για εξαγωγή χαρακτηριστικών από τις εικόνες και μία υπομονάδα εκτίμησης κίνησης [55]. Τέλος η τελευταία μονάδα αποτελείται από έναν αλγόριθμο πιθανολογικού οπτικού εντοπισμού και χαρτογράφησης (Simultaneous Localization And Mapping, SLAM) [55]. Η πρώτη μονάδα (3D reconstruction) χρησιμοποιείται και για τη λειτουργία της χαρτογράφησης και για τη λειτουργία του εντοπισμού, ενώ οι υπόλοιπες μονάδες χρησιμοποιούνται μόνο στην οπτική οδομετρία (visual odometry) στη λειτουργία του εντοπισμού.

Το Σχήμα 3.31 απεικονίζει ένα σχηματικό διάγραμμα της αρχιτεκτονικής του συστήματος SPARTAN, ενώ η λειτουργικότητα των χρησιμοποιούμενων αλγορίθμων CV συνοφίζεται ως εξής [51]:

- **Προ-επεξεργασία εικόνας (Imaging):** Εκτελείται επεξεργασία εικόνας για την δημιουργία κατάλληλης εισόδου στους υπόλοιπους αλγόριθμους.
- **Ανακατασκευή 3D χάρτη (3D Map Reconstruction):** Παραγωγή τρισδιάστατου χάρτη του περιβάλλοντος.
- **Οπτική οδομετρία (Visual Odometry):** Παροχή εκτίμησης μετατόπισης του ρομποτικού συστήματος.
- **Οπτικός εντοπισμός και χαρτογράφηση (Visual SLAM):** Καθορισμός της τρέχουσας θέσης του ρομποτικού συστήματος.
- **Εντοπισμός (Localization):** Εύρεση χωροθέτησης του ρομποτικού συστήματος στον τρισδιάστατο χάρτη.



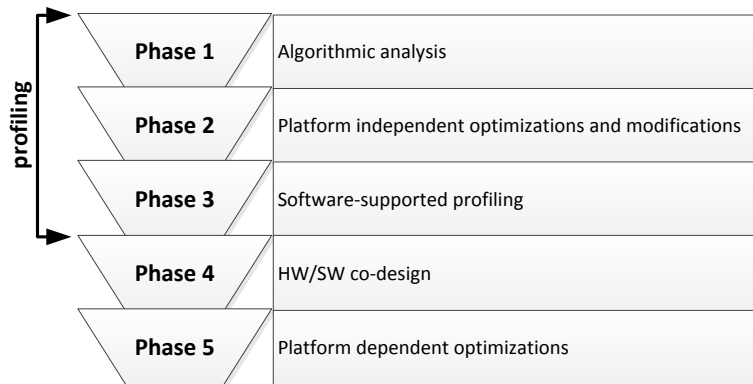
Σχήμα 3.31 Σχηματική απεικόνιση των λειτουργιών χαρτογράφησης και εντοπισμού του συστήματος SPARTAN.

Η αποτελεσματικότητα του συστήματος SPARTAN, σύμφωνα με τις προδιαγραφές της ESA, επαληθεύεται χρησιμοποιώντας δύο σενάρια: (I) ένα ρομποτικό σύστημα παρόμοιων προδιαγραφών με το ρομποτικό σύστημα της αποστολής Exomars, χρησιμοποιώντας τον ρεαλιστικό εξομοιωτή 3DROV [56] και (II) ένα

πραγματικό ρομποτικό σύστημα σε περιβάλλον εξωτερικού χώρου. Συγκεκριμένα, το πρώτο σενάριο περιλαμβάνει ένα ρομποτικό όχημα που εξομοιώνεται μέσω της χρήσης μοντέλου βασισμένου σε σχέδιο CAD του οχήματος της αποστολής Exomars, καθώς και ένα εικονικό τρισδιάστατο περιβάλλον που προσομοιώνει το ανάγλυφο του πλανήτη Άρη. Το δεύτερο σενάριο περιλαμβάνει ένα ρεαλιστικό μοντέλο ρομποτικού συστήματος, το οποίο αν και δεν αποτελεί αντίγραφο του συστήματος της αποστολής Exomars, εντούτοις παρέχει τη δυνατότητα ρεαλιστικής επίδειξης και απόδειξης της αποτελεσματικότητας της προτεινόμενης αρχιτεκτονικής για πραγματικά ρομποτικά οχήματα. Και στα δύο σενάρια, οι αλγόριθμοι CV τροφοδοτούνται τόσο με τεχνητές εικόνες, δημιουργημένες στο τρισδιάστατο περιβάλλον 3DROV, όσο και με ρεαλιστικές εικόνες από πραγματικές λήψεις στερεοσκοπικών καμερών σε διάφορα περιβάλλοντα εξωτερικού χώρου που παρουσιάζουν ομοιότητες με το ανάγλυφο του πλανήτη Άρη.

3.4.3. Μεθοδολογία συν-σχεδιασμού Υλικού/Λογισμικού.

Αυτή η υποενότητα περιγράφει τη προτεινόμενη μεθοδολογία για τον συν-σχεδιασμό υλικού/λογισμικού. Παρόλο που η προτεινόμενη μεθοδολογία είναι μια λύση γενικού σκοπού, πάραυτα, στο πλαίσιο του έργου SPARTAN, έχουν ληφθεί κατάλληλα υπόψη τα εγγενή αρχιτεκτονικά χαρακτηριστικά του υπό-ανάπτυξη συστήματος (FPGA και χαμηλής απόδοσης CPU) καθώς και οι περιορισμοί που σχετίζονται με την αυξανόμενη υπολογιστική πολυπλοκότητα των αλγορίθμων CV, όπως ορίζονται από τις προδιαγραφές της ESA. Το Σχήμα 3.32 απεικονίζει τη βασική ροή διεργασιών της προτεινόμενης σχεδιαστικής μεθοδολογίας.

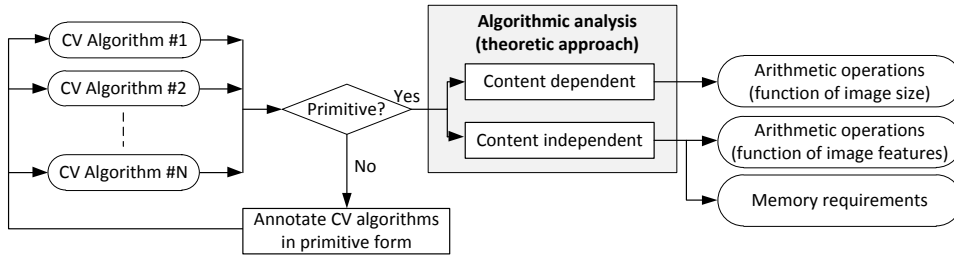


Σχήμα 3.32 Ροή διεργασιών της προτεινόμενης μεθοδολογίας συν-σχεδιασμού Υλικού/Λογισμικού.

3.4.3.1. Φάση 1: Αλγοριθμική ανάλυση

Η πρώτη φάση της προτεινόμενης λύσης, η οποία απεικονίζεται στο Σχήμα 3.33, περιλαμβάνει θεωρητική μελέτη των υλοποιημένων αλγορίθμων CV. Η ανάλυση των αλγορίθμων είναι ένα σημαντικό μέρος της προτεινόμενης μεθοδολογίας

εφόσον επιτρέπει θεωρητικές εκτιμήσεις για τις απαιτήσεις υπολογιστικών των αλγορίθμων CV.



Σχήμα 3.33 Διεργασίες που εκτελούνται κατά τη φάση της αλγοριθμικής ανάλυσης.

Προς την κατεύθυνση αυτή, αρχικά ελέγχεται εάν οι αλγόριθμοι περιγράφονται σε “πρωτογενή” μορφή. Με τον όρο “πρωτογενή” αναφερόμαστε σε μια αλγοριθμική περιγραφή μέσω πηγαίου κώδικα, ο οποίος είναι απαλλαγμένος από κλήσεις σε εξωτερικές βιβλιοθήκες λογισμικού. Αυτή η πιστοποίηση είναι πολύ σημαντική ώστε να υπάρξει ακριβής χαρακτηρισμός. Ένα παράδειγμα μη πρωτογενής περιγραφής είναι ο κώδικας σε υψηλή γλώσσα προγραμματισμού, π.χ. Matlab, ο οποίος κάνει χρήση ενσωματωμένων συναρτήσεων είτε του προγραμματιστικού περιβάλλοντος, είτε του συστήματος, προς επιτάχυνση της εκτέλεσης. Σε αυτή τη περίπτωση ο χρόνος εκτέλεσης για ενσωματωμένες/εξωτερικές κλήσεις συναρτήσεων στο Matlab μπορεί να είναι λίγα δευτερόλεπτα, ενώ η πραγματική αναλογία υπολογιστικού κόστους να είναι πολύ μεγαλύτερη. Συνεπώς αναγνωρίζοντας αυτή την ιδιαιτερότητα, το προτεινόμενο πλαίσιο διατηρεί ένα προ-επεξεργαστικό βήμα στο οποίο εξασφαλίζεται η πρωτογενής μορφή των διαθέσιμων αλγορίθμων CV. Στο βήμα αυτό καταβάλλεται προσπάθεια από την ομάδα σχεδίασης προς τροποποίηση του πηγαίου κώδικα (code refactoring) σε πρωτογενή μορφή χωρίς να επηρεαστεί η λειτουργικότητά των αλγορίθμων.

Η διαδικασία ανάλυσης των αλγορίθμων καθορίζει τους υπολογιστικούς πόρους, όπως οι μαθηματικές πράξεις και ο χώρος αποθήκευσης των δεδομένων, που απαιτούνται για την εκτέλεση του αλγορίθμου. Εκκινώντας από την πρωτογενή μορφή του πηγαίου κώδικα διακρίνονται δύο προσεγγίσεις ανάλυσης των αλγορίθμων, βάση της εξάρτησής τους από την εισόδου τους. Συγκεκριμένα, στη περίπτωση που η πολυπλοκότητα των αλγορίθμων CV δεν εξαρτάται από το περιεχόμενο της εισόδου, π.χ. εικόνας, τότε η θεωρητική ανάλυση παρέχει ακριβή αποτελέσματα για τον αριθμό των μαθηματικών πράξεων που απαιτούνται για να εκτελεστούν αυτοί οι αλγόριθμοι σε μία αφαιρετική υπολογιστική πλατφόρμα. Αντίθετα, στη περίπτωση που η εκτέλεση του αλγορίθμου εξαρτάται από το περιεχόμενο της εισόδου, η θεωρητική ανάλυση αναφέρει τον αριθμό των μαθηματικών πράξεων συναρτήσει των χαρακτηριστικών της εισόδου, π.χ. της εικόνας.

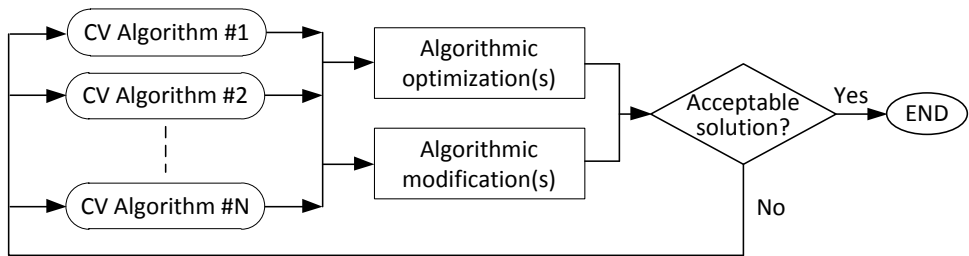
Στην παρούσα εργασία τα μεγέθη εικόνας που χρησιμοποιούνται ως είσοδο στο σύστημα SPARTAN ορίζονται από τις προδιαγραφές της ESA και είναι σταθερά και συνεπώς δεν υπάρχει διακύμανση ως προς τον αριθμό των μαθηματι-

κών πράξεων. Ωστόσο η προτεινόμενη αλγοριθμική ανάλυση παρέχει επίσης τις απαραίτητες διαδικασίες για την εκτίμηση των υπολογιστικών μετρικών, ακόμα και την περίπτωση όπου η είσοδος είναι δυναμική. Συγκεκριμένα χρησιμοποιείται κατάλληλη διαδικασία με τη μέθοδο της παρεκβολής (extrapolation), η οποία και συζητείται στην επόμενη υποενότητα.

Τα συμπεράσματα που προκύπτουν από αυτή την ανάλυση παρέχουν πολύτιμες ενδείξεις για την υπολογιστική πολυπλοκότητα των αλγορίθμων CV, τις απαιτήσεις απόδοσής τους, τις απαιτήσεις μνήμης/αποθήκευσης των δεδομένων, καθώς και το ρυθμό επικοινωνίας μεταξύ τους. Αυτές οι πληροφορίες μπορούν να διαχειριστεί κατάλληλα προκειμένου να δώσουν κάποιες πιθανές αλγοριθμικές βελτιστοποιήσεις για τους αλγορίθμους CV και να παρέχουν συγκεκριμένες οδηγίες στη διαδικασία συν-σχεδιασμού υλικού/λογισμικού.

3.4.3.2. Φάση II: Βελτιστοποιήσεις και τροποποιήσεις ανεξαρτήτως της πλατφόρμας υλοποίησης

Η δεύτερη φάση της προτεινόμενης μεθοδολογίας εφαρμόζει σε αλγοριθμικό επίπεδο έναν αριθμό από βελτιστοποιήσεις και τροποποιήσεις ανεξαρτήτως της πλατφόρμας υλοποίησης. Αυτό επιτυγχάνεται μέσω τροποποιήσεων πηγαίου κώδικα σε πηγαίο κώδικα (source-to-source refactoring). Το Σχήμα 3.34 απεικονίζει τις εργασίες που εκτελούνται στην δεύτερη φάση.



Σχήμα 3.34 Προτεινόμενη ροή εργασιών για βελτιστοποιήσεις και τροποποιήσεις ανεξαρτήτως της πλατφόρμας υλοποίησης.

Είναι σύνθηες για τους αλγόριθμους CV να αναπτύσσονται σε υψηλού επιπέδου προγραμματισμό προκειμένου να αποκρύπτεται η επίπονη συγγραφή των λεπτομερών μαθηματικών πυρήνων επεξεργασίας. Τυπικές γλώσσες αλγορίθμων CV, όπως η OpenCV, διαθέτουν επίσης στους προγραμματιστές ειδικές συναρτήσεις, μαθηματικές βιβλιοθήκες και τύπους δεδομένων οι οποίες κάνουν ακόμα πιο εύκολη τη συγγραφή και διατήρηση μεγάλων και πολύπλοκων ροών επεξεργασίας εικόνας και βίντεο. Ωστόσο το τίμημα αυτής της αφαιρετικότητας είναι η δυσκολία μεταφοράς του πηγαίου κώδικα σε διαφορετικές πλατφόρμες υλοποίησης από αυτές για τις οποίες είναι σχεδιασμένες και βελτιστοποιημένες οι γλώσσες υψηλού επιπέδου. Αυτό το πρόβλημα γίνεται ακόμα μεγαλύτερο όταν η πλατφόρμα υλοποίησης είναι ένα ετερογενές σύστημα αποτελούμενο από διαφορετικά υποσυστήματα εκτέλεσης, όπως είναι η περίπτωση της πλατφόρμας CPU-FPGA του έργου SPARTAN. Σε τέτοιου τύπου συστήματα, η χρήση

γλωσσών υψηλού επιπέδου δεν λαμβάνει υπόψιν θέματα που σχετίζονται με τη παραλληλοποίηση της εφαρμογής, καθώς και με το περιορισμένο μέγεθος της μνήμης αποθήκευσης που βρίσκετε σε ενσωματωμένες πλατφόρμες. Συνεπώς, αυτοί οι αλγόριθμοι πρέπει να βελτιστοποιηθούν κατάλληλα προκειμένου αν συμπεριλάβουν έναν αριθμό αρχιτεκτονικών περιορισμών. Σημειωτέον ότι κατά τη διάρκεια αυτών των βελτιστοποιήσεων πρέπει να διατηρηθεί η λειτουργικότητα των αλγορίθμων CV. Στα πλαίσια αυτής της εργασίας υιοθετήθηκαν αλγοριθμικές βελτιστοποιήσεις που αναφέρονται στην εργασία [57].

Εκτός από αυτές τις βελτιστοποιήσεις, κατά τη διάρκεια της δεύτερης φάσης της εισαγόμενης μεθοδολογίας, εφαρμόζεται επίσης ένας αριθμός αλγοριθμικών τροποποιήσεων. Σκοπός αυτών των τροποποιήσεων είναι η σημαντική μείωση της πολυπλοκότητας σε επίπεδο υπολογισμών, αποθήκευσης και ρυθμού επικοινωνίας, για τους υλοποιημένους αλγορίθμους CV, με ένα ελεγχόμενο κόστος της ακρίβειάς τους, αρκεί το τελευταίο να διατηρεί το συνολικό σύστημα εντός των προδιαγραφών του έργου. Σε αντίθεση με τις αλγοριθμικές βελτιστοποιήσεις, οι εφαρμοσμένες τροποποιήσεις συνήθως επηρεάζουν τη λειτουργικότητα των αλγορίθμων CV. Σε αυτή τη περίπτωση, η απόδοση των αλγορίθμων CV ποσοτικοποιείται αναφορικά προς την απόδοση και την επιθυμητή ακρίβεια. Σε περίπτωση που η συμπληρωματική λύση δεν ικανοποιεί τις προδιαγραφές του συστήματος, υπάρχει ένας βρόχος ανατροφοδότησης για επιπλέον βελτιώσεις. Εναλλακτικά ακολουθείται η τρίτη φάση της προτεινόμενης μεθοδολογίας.

3.4.3.3. Φάση III: Χαρακτηρισμός εφαρμογής μέσω λογισμικού

Ο χαρακτηρισμός (profiling) μίας εφαρμογής είναι μία εκ των κρίσιμων διαδικασιών του συν-σχεδιασμού υλικού/λογισμικού, και η ουσιαστική σημασία τονίζεται ακόμα περισσότερο σε περιπτώσεις όπου η πλατφόρμα υλοποίησης αποτελείται από ετερογενείς επεξεργαστικούς πυρήνες, όπως είναι το σύστημα SPARTAN.

Η προτεινόμενη μεθοδολογία χαρακτηρισμού των εφαρμογών εκτελείται σε διαφορετικά επίπεδα αφαίρεσης του συστήματος. Συγκεκριμένα, αναφορικά προς το έργο SPARTAN, διακρίθηκαν δύο επίπεδα αφαίρεσης: (i) αδρομερές και (ii) λεπτομερές. Αυτή η προσέγγιση επιτρέπει στους σχεδιαστές να εκτελέσουν γρήγορη αξιολόγηση σύνθετων συστημάτων, όπου ένα μόνο υποσύνολο των αλγορίθμων πρέπει να μελετηθεί σε λεπτομερές επίπεδο, ενώ το υπόλοιπο σύστημα χαρακτηρίζεται σε πιο αδρομερή συμπεριφορά. Συγκεκριμένα, παρόλο που η λεπτομερής ανάλυση εγγυάται να βρει τα ακριβή και λεπτομερή στοιχεία για εκείνους τους υπολογιστικούς πυρήνες που κατέχουν το μεγαλύτερο ποσοστό εκτέλεσης του συστήματος, ωστόσο, δεν είναι είναι κλιμακούμενη αναφορικά προς το μέγεθος του πηγαίου κώδικα. Οπότε, προκειμένου να εφαρμοστεί ανάλυση σε πολυσύνθετα και ετερογενή περιβάλλοντα το προτεινόμενο πλαίσιο χρησιμοποιεί μία υβριδική προσέγγιση χαρακτηρισμού, κατά την οποία επιλέγονται για λεπτομερή ανάλυση οι πυρήνες με το μεγαλύτερο κόστος εκτέλεσης.

Αναλυτικότερα, κατά το στάδιο του χαρακτηρισμού εφαρμογής μέσω λογισμικού εφαρμόζεται αρχικά μια αδρομερής ανάλυση σε όλη τη ροή εκτέλεσης του συστήματος, προκειμένου να αναγνωριστούν οι “κρίσιμοι” πυρήνες. Με τον

όρο αυτό χαρακτηρίζονται τα τμήματα πηγαίου κώδικα τα οποία σημειώνουν αυξημένο κόστος εκτέλεσης στο σύστημα. Το κόστος εκτέλεσης είναι μία μετρική που αξιολογείται ως προς το ποσοστό εκτέλεσης, τη δέσμευση μνήμης και το κόστος επικοινωνίας. Στη συνέχεια αυτοί οι πυρήνες αναλύονται με τη χρήση λεπτομερούς χαρακτηρισμού προκειμένου να καταμετρηθούν οι παράμετροι σχεδιασμού που επηρεάζουν την υλοποίηση του συστήματος. Ανάμεσα σε αυτούς η εργασία αυτή προτείνει την υπολογιστική πολυπλοκότητα, τα ίχνη μνήμης, τη διάρκεια ζωής των δεδομένων και το ρυθμό επικοινωνίας.

Για τον χαρακτηρισμό πολύπλοκων συστημάτων η προτεινόμενη μεθοδολογία χρησιμοποιεί έναν σύνολο από κατάλληλα εργαλεία λογισμικού, προκειμένου να εξασφαλίσει εγκυρότητα αποτελεσμάτων αλλά και αυτοματισμό της ροής σχεδίασης. Συγκεκριμένα, η αδρομερής ανάλυση εκτελείται με τη χρήση των συναρτήσεων μέτρησης χρόνου που διαθέτει η κάθε γλώσσα προγραμματισμού. Τέτοιες συναρτήσεις μπορεί να είναι ο συνδυασμός *Tic-Toc* στη γλώσσα Matlab και η *time* στις γλώσσες C/C++. Η λεπτομερής ανάλυση εκτελείται με αλγοριθμική ανάλυση καθώς και με εργαλεία λογισμικού εξειδικευμένα για αυτή τη διαδικασία (π.χ. Matlab profiler, VTune, Valgrind, κλπ). Η προαναφερθείσα ταξινόμηση απεικονίζεται στο Σχήμα 3.35.

3

Σχήμα 3.35 Προτεινόμενα εργαλεία χαρακτηρισμού εφαρμογών.

Στη συνέχεια, συνοψίζουμε τα κύρια πλεονεκτήματα και μειονεκτήματα των εναλλακτικών υποστηριζόμενων μεθόδων ανάλυσης:

- *Tic-Toc*: Αυτή η μέθοδος εφαρμόζεται στους αλγορίθμους CV που περιγράφονται με γλώσσα Matlab/Octave. Η έξοδος από την *Tic-Toc* ανάλυση παρέχει μια επισκόπηση του χρόνου εκτέλεσης για διαφορετικά μέρη του αλγορίθμου. Αφού η μέθοδος αυτή εφαρμόζεται επιλεκτικά μόνο στους “κρίσιμους” πυρήνες, η ευελιξία της και η αποδοτικότητά της μειώνεται όσο το μέγεθος του κώδικα αυξάνει. Η μέτρηση με αυτή τη μέθοδο επιφέρει ένα σφάλμα στον απόλυτο υπολογισμό χρόνου, λόγω του υψηλού επιπέδου αφαιρετικότητας των γλωσσών Matlab/Octave.

- *Time*: Η μέθοδος *Time* αναφέρει το χρόνο εκτέλεσης για αλγορίθμους CV που περιγράφονται σε γλώσσες C/C++ καθώς και σε υψηλού επιπέδου γλώσσες που αποτελούν παράγωγα αυτών, όπως π.χ. η OpenCV. Όπως και στη μέθοδο *Tic-Toc*, επίσης αυτή η μέθοδος εισάγει ένα κόστος υπολογισμού λόγω του λειτουργικού συστήματος. Προκειμένου να ελαχιστοποιηθεί η επίδραση αυτού του κόστους προτείνεται ο καθολικός χαρακτηρισμός των αλγορίθμων στο ίδιο σύστημα.
- *Αλγοριθμική ανάλυση (Algorithmic analysis)*: Αυτή η μέθοδος παρέχει τη μέγιστη ακρίβεια αλλά δεν είναι κατάλληλη για αυξημένου μεγέθους κώδικες επειδή εφαρμόζεται με μη αυτόματο τρόπο. Επιπρόσθετα, η *αλγοριθμική ανάλυση* δεν λαμβάνει υπόψιν της τα κόστη που τίθενται από την επιλεγόμενη γλώσσα περιγραφής (π.χ. C++, VHDL).
- *Matlab/Octave/C/C++ profiler*: Η τελευταία μέθοδος περιλαμβάνει αλγοριθμικό χαρακτηρισμό με τη χρήση εξειδικευμένων εργαλείων (profilers). Αυτά τα εργαλεία εισάγουν κατάλληλες εντολές στον πηγαίο κώδικα των αλγορίθμων και επιτρέπουν την εξαγωγή χρήσιμων πληροφοριών από την εκτέλεσή τους. Στα πλαίσια αυτής της εργασίας προτείνεται η χρήση των μετρικών του χρόνου εκτέλεσης, του αριθμού κλήσεων των συναρτήσεων, του αριθμού συσχέτισης συναρτήσεων γονέων/παιδιών και του αριθμού εκτέλεσης κάθε γραμμής πηγαίου κώδικα. Παρόλο που η προσέγγιση αυτή εξάγει τη μέγιστη δυνατή πληροφορία από το βήμα χαρακτηρισμού, ωστόσο έχει μεγάλο κόστος εκτέλεσης αναφορικά προς το χρόνο επιβάρυνσης στην εκτέλεση του αλγορίθμου. Επίσης επιβάλλει την ύπαρξη του πηγαίου κώδικα αναγκαστικά σε πρωτογενή μορφή, διαφορετικά δεν μπορούν να εξαχθούν ακριβή συμπεράσματα.

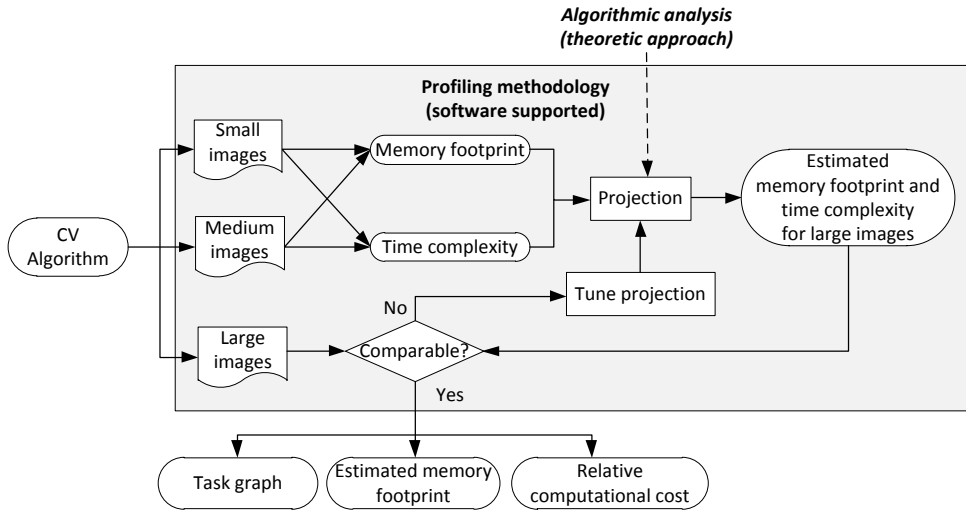
Όπως έχει ήδη αναφερθεί, η διαδικασία του χαρακτηρισμού επιφέρει μεγάλη επιβάρυνση στο χρόνο εκτέλεσης του αλγορίθμου. Επίσης, λόγω του ότι είναι μια διαδικασία που πρέπει να κρατά πληροφορίες για όλες τις μεταβλητές του εκτελέσιμου κώδικα, καθ όλη τη διάρκεια του εκτελούμενου στιγμιότυπου, ειδικά για την περίπτωση των αλγορίθμων CV, οι οποίοι έχουν αυξημένες απαιτήσεις σε μέγεθος μεταβλητών, η διαδικασία χαρακτηρισμού μπορεί να οδηγήσει σε δέσμευση υπερμεγέθους μνήμης συστήματος. Για παράδειγμα στο έργο SPARTAN, η λειτουργία εντοπισμού χρησιμοποιεί τρία ζεύγη σταθερών εικόνων, κάθε ένα από τα οποία έχει μέγεθος 1120×1120 pixels με 200 επίπεδα βάθους, ενώ η φάση εντοπισμού χρησιμοποιεί ένα ζεύγος σταθερών εικόνων με μέγεθος 512×384 pixels.

Λόγω των αυξημένων μεγεθών των εικόνων ο συμβατικός χαρακτηρισμός είναι μια χρονοβόρα διαδικασία. Προκειμένου να ξεπεραστεί αυτός ο περιορισμός, η προτεινόμενη μεθοδολογία ενσωματώνει μια τεχνική εκτίμησης του τελικού χαρακτηρισμού, η οποία βασίζεται σε εικόνες μικρότερου μεγέθους. Η λύση αυτή εκτιμά με αξιοπιστία έναν αριθμό αρχιτεκτονικών και αλγοριθμικών παραμέτρων αλλά σε σημαντικά μικρότερο χρόνο εκτέλεσης.

Έχοντας ως είσοδο τις πληροφορίες από τα παραπάνω βήματα, κατά το στάδιο χαρακτηρισμού εκτιμάται το ίχνος μνήμης και η πολυπλοκότητα χρόνου

για κάθε ένα από τους αλγόριθμους CV, καθώς και για ολόκληρο το σύστημα SPARTAN. Αυτές οι τιμές χρησιμοποιούνται ως είσοδοι σε μία μέθοδο παρεκβολής [58] προκειμένου να εκτιμηθούν οι αντίστοιχες τιμές αναφορικά προς το πραγματικό μέγεθος εικόνων για το σύστημα SPARTAN. Καθώς είναι πιθανό να υπάρξουν αποκλίσεις μεταξύ των εκτιμώμενων τιμών και των πραγματικών αποτελεσμάτων υλοποίησης, προτείνεται ένα επιπλέον βήμα, κατά το οποίο μπορούν να προσαρμοστούν συγκεκριμένες παράμετροι της μεθόδου παρεκβολής. Σημειώνεται ότι για το έργο SPARTAN, αυτό το βήμα προσαρμογής εκτελέστηκε μόνο μια φορά, ενώ οι ρυθμισμένες παράμετροι χρησιμοποιήθηκαν και στο υπόλοιπο τμήμα του χαρακτηρισμού, χωρίς μεγάλες αποκλίσεις.

Το Σχήμα 3.36 απεικονίζει την προτεινόμενη μεθοδολογία για τον χαρακτηρισμό αλγορίθμων CV μέσω λογισμικού.

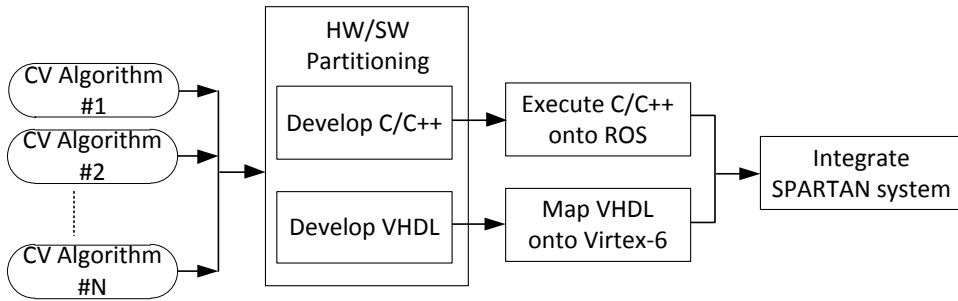


Σχήμα 3.36 Προτεινόμενη μεθοδολογία για τον χαρακτηρισμό αλγορίθμων CV μέσω λογισμικού.

3.4.3.4. Φάση IV: Συν-σχεδιασμός Υλικού/Λογισμικού

Η προτεινόμενη μεθοδολογία συν-σχεδιασμού υλικού/λογισμικού απεικονίζεται στο Σχήμα 3.37. Ένα από τα πιο κρίσιμα σημεία αυτής της μεθοδολογίας αφορά την κατάτμηση (partitioning) ενός αλγορίθμου σε πόρους υλικού, σύμφωνα με τα αποτελέσματα του χαρακτηρισμού που ανακτήθηκαν στο προηγούμενο βήμα. Όπως έχει ήδη αναφερθεί, η αξιοπιστία των συμπερασμάτων του χαρακτηρισμού εξαρτάται από τη, πρώιμη ή μη, περιγραφική μορφή των αλγορίθμων CV.

Έχοντας ως είσοδο τις αποφάσεις από το βήμα κατάτμησης του αλγορίθμου, το επόμενο βήμα της μεθοδολογίας είναι η ανάπτυξη και βελτιστοποίηση κώδικα σε γλώσσες C/C++ και VHDL για τους αλγορίθμους CV. Σύμφωνα με τις προδιαγραφές του έργου SPARTAN, ο αναπτυσσόμενος C/C++ κώδικας ενσωματώνεται σε ένα υπολογιστικό σύστημα το οποίο χρησιμοποιεί το λειτουργικό



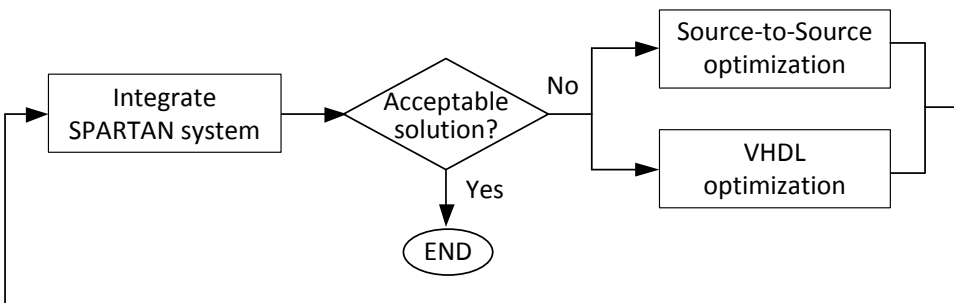
Σχήμα 3.37 Προτεινόμενη ροή εργασιών για συν-σχεδιασμό υλικού/λογισμικού.

σύστημα ROS, ενώ ο κώδικας VHDL αφορά μία συσκευή Virtex-6 FPGA board.

Οι αναπτυσσόμενοι αλγόριθμοι αξιολογούνται με τη χρήση τριών διαφορετικών σεναρίων: (i) κάθε αλγόριθμος αξιολογείται μεμονωμένα (ii) η συνολική απόδοση της ροής εκτέλεσης της λειτουργίας χαρτογράφησης αξιολογείται σε σχέση με τις προδιαγραφές του έργου για την τρισδιάστατη ανακατασκευή χάρτη και (iii) η συνολική απόδοση της ροής εκτέλεσης της λειτουργίας εντοπισμού αξιολογείται σε σχέση με τις προδιαγραφές του έργου για την εκτίμηση τοποθεσίας του ρομποτικού οχήματος.

3.4.3.5. Φάση V: Βελτιστοποίηση εξαρτώμενη από την πλατφόρμα υλοποίησης

Κατά την τελευταία φάση της προτεινόμενης μεθοδολογίας, η οποία απεικονίζεται στο Σχήμα 3.38, εφαρμόζεται ένας αριθμός από βελτιστοποιήσεις σε σχέση με την πλατφόρμα υλοποίησης, οι οποίες στοχεύουν στην επιπλέον βελτίωση της απόδοσης του τελικού συστήματος. Για το σκοπό αυτό, τόσο η αποτελεσματικότητα κάθε αλγορίθμου CV, όσο και το συνολικό σύστημα συγκρίνονται με τις προδιαγραφές που θέτει η ESA, στα πλαίσια του έργου SPARTAN.



Σχήμα 3.38 Προτεινόμενη ροή εργασιών προς βελτιστοποίηση του συστήματος SPARTAN, λαμβάνοντας υπόψιν την πλατφόρμα υλοποίησης.

Σε έργα μακροπρόθεσμης υλοποίησης, όπως η πλειοψηφία αυτών που αφορούν το διάστημα και στα οποία οι προδιαγραφές συστήματος αφορούν συγκεκριμένες υπολογιστικές επιδόσεις, είναι σύνθηρες να μην είναι διαθέσιμο το

τελικό πρωτότυπο υλικό του τελικού συστήματος στο οποίο θα εκτελεστούν οι αλγόριθμοι CV, κατά τη φάση σχεδίασης του έργου. Για παράδειγμα για το έργο SPARTAN, παρόλο που ο οργανισμός ESA έχει θέσει ως προδιαγραφή ένα επεξεργαστή με ισχύ 150 MIPS, εντούτοις, το πραγματικό επεξεργαστικό σύστημα δεν ήταν διαθέσιμο κατά τη φάση υλοποίησης. Προκειμένου λοιπόν να προχωρήσει η ανάπτυξη του έργου, η παρούσα εργασία προτείνει μία διαδικασία εκτίμησης της συνολικής απόδοσης του συστήματος για τον επεξεργαστή ισχύος 150 MIPS, όταν αυτό εκτελείται σε έναν συμβατικό επεξεργαστή, π.χ. αρχιτεκτονικής x86 με μεγαλύτερη επεξεργαστική ισχύ. Συγκεκριμένα, αρχικά μετράται ο χρόνος εκτέλεσης των πυρήνων επεξεργασίας που εκτελούνται στον συμβατικό επεξεργαστή CPU T_{cpu} και ο χρόνος εκτέλεσης των πυρήνων επεξεργασίας που εκτελούνται στη συσκευή FPGA T_{fpga} . Έπειτα υπολογίζεται ο συνολικός χρόνος εκτέλεσης $T_{total_projected}$ αφού προβληθεί ο χρόνος του συμβατικού επεξεργαστή CPU σε έναν επεξεργαστή CPU με ισχύ 150 MIPS, σύμφωνα με την εξίσωση 3.3.

$$T_{total_projected} = T_{fpga} + T_{cpu} \times \frac{150}{MIPS} \quad (3.3)$$

Στη περίπτωση που οι προκύπτουσες αποφάσεις για την κατάτμηση υλικού-λογισμικού, καθώς και την υλοποίηση των CV αλγορίθμων, δεν ικανοποιούν τις προδιαγραφές του έργου, τότε ένα επιπλέον βήμα βελτιστοποίησης απαιτείται προκειμένου να βελτιωθεί η συνολική απόδοση του συστήματος. Σε αντίθεση με τις προηγούμενες διαδικασίες, στη φάση αυτή εφαρμόζονται βελτιστοποιήσεις εξαρτώμενες από την πλατφόρμα και οι οποίες λαμβάνουν υπόψιν τα εγγενή αρχιτεκτονικά χαρακτηριστικά της πλατφόρμας υλοποίησης. Παραδείγματα τέτοιων βελτιστοποιήσεων αποτελούν ο παραλληλισμός και η τεχνική σωλήνωσης σε επίπεδο διεργασιών (task level parallelism) για πυρήνες επεξεργασίας που έχουν ανατεθεί στη συσκευή FPGA, λαμβάνοντας υπόψιν το κόστος επικοινωνίας PC-FPGA.

3.4.4. Πειραματικά Αποτελέσματα

Αυτή η υποενότητα παρέχει ένα σύνολο αποτελεσμάτων που επαληθεύουν την αποδοτικότητα της προτεινόμενης μεθοδολογίας. Συγκεκριμένα, αρχικά αναφέρονται τα συμπεράσματα που προέκυψαν από την εφαρμογή της διαδικασίας χαρακτηρισμού των αλγορίθμων CV, οι οποίοι υλοποιούνται στα πλαίσια του προγράμματος SPARTAN [53] [54] [55] [59].

3.4.4.1. Γενικευμένα αποτελέσματα χαρακτηρισμού των αλγορίθμων CV

Το Σχήμα 3.39 απεικονίζει τα αποτελέσματα χαρακτηρισμού που προέρχονται από τη μελέτη των αλγορίθμων CV, οι οποίοι υλοποιούνται για τους σκοπούς του έργου SPARTAN, τόσο για τη φάση της χαρτογράφησης (Σχήμα 3.39(α)), όσο και για τη φάση του εντοπισμού (Σχήμα 3.39(β)). Συγκεκριμένα, αυτό το σχήμα αναφέρει τα πειραματικά αποτελέσματα που ανακτώνται από (i) τον Matlab/Octave profiler, (ii) την μέθοδο “Tic-Toc” και (iii) την αλγοριθμική ανάλυση. Ο κάθετος άξονας του σχήματος αυτού αντιστοιχεί στη πολυπλοκότητα

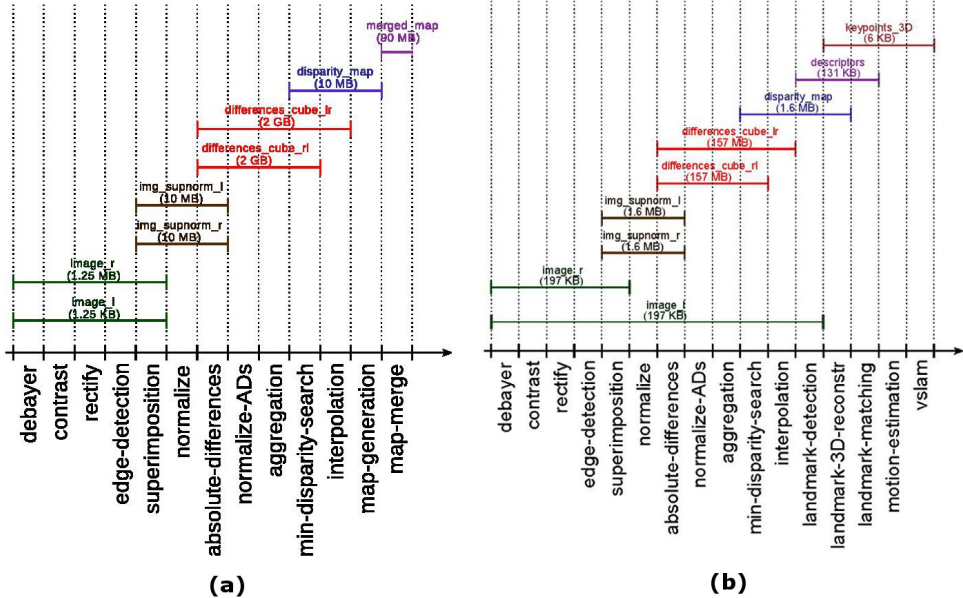
του χρόνου. Για λόγους παρουσίασης, ο άξονας αυτός δίνει το σχετικό κόστος ανάμεσα στους διαφορετικούς αλγορίθμους CV και μεθόδους χαρακτηρισμού, ως ένα ποσοστό του συνολικού χρόνου εκτέλεσης, υποθέτοντας ξεχωριστή εκτέλεση των λειτουργιών χαρτογράφησης και εντοπισμού.

Σύμφωνα με τα αποτελέσματα που απεικονίζονται στο Σχήμα 3.39, είναι δυνατό να οριστεί το ποσοστό χρόνου εκτέλεσης κάθε αλγορίθμου επί του συνόλου του χρόνου εκτέλεσης για όλο το σύστημα. Συγκεκριμένα, οι αλγόριθμοι *rec3d*, *disparity* και *aggregation* είναι οι πιο χρονοβόροι πυρήνες στη λειτουργία χαρτογράφησης, ενώ όσον αφορά τη λειτουργία εντοπισμού ο πιο χρονοβόρος πυρήνας είναι ο πυρήνας *aggregation*. Αυτά τα συμπεράσματα παρέχουν στους σχεδιαστές πολύτιμη πληροφορία για τους πυρήνες που έχουν αυξημένη συστημική πολυπλοκότητα και συνεπώς πρέπει είναι υποψήφιοι για ανάθεση εκτέλεσης στην συσκευή FPGA προκειμένου να επιταχυνθεί η εκτέλεσή τους. Ακόμα σημειώνεται ότι οι τρεις μέθοδοι χαρακτηρισμού έχουν μικρές αποκλίσεις ως προς τα ποσοστά χρόνου εκτέλεσης των αλγορίθμων. Με αυτό τον τρόπο επιβεβαιώνεται ότι η προτεινόμενη μεθοδολογία μπορεί να εξάγει συμπεράσματα υψηλής αξιοπιστίας.

Σχήμα 3.39 Αποτελέσματα χαρακτηρισμού των αλγορίθμων CV για τις λειτουργίες (α) χαρακτηρισμού και (β) εντοπισμού.

3.4.4.2. Χρονική διάρκεια ζωής δεδομένων

Στη συνέχεια παρέχονται μερικά πειραματικά αποτελέσματα για τη διάρκεια ζωής των δεδομένων κατά την εκτέλεση των αλγορίθμων CV. Αυτά εξήχθησαν με αλγοριθμική ανάλυση και απεικονίζονται στα Σχήματα 3.40(α) και 3.40(β), αντίστοιχα για τις λειτουργίες της χαρτογράφησης και του εντοπισμού. Οι τιμές αυτές υπολογίστηκαν μελετώντας τον πηγαίο κώδικα των αλγορίθμων και υποθέτοντας σειριακή εκτέλεση αυτών. Σημειώνεται ότι κατά τη διάρκεια της ανάλυσης αυτής δεν λαμβάνεται υπόψιν κάποιο πιθανό σενάριο παραλληλισμού.



Σχήμα 3.40 Χρονική διάρκεια ζωής δεδομένων για τις λειτουργίες (α) της χαρτογράφησης και (β) του εντοπισμού.

Σύμφωνα με τα αποτελέσματα που απεικονίζονται στο Σχήμα 3.40, συμπεραίνεται ότι δύο μεταβλητές, οι οποίες ονομάζονται ως *differences cube left* και *differences cube right*, επιδεικνύουν τη μέγιστη ζήτηση για αποθήκευση (on-chip memory). Συγκεκριμένα, αυτές οι δύο μεταβλητές αποθηκεύουν σχεδόν 4GB δεδομένων, ενώ αυτή η πληροφορία πρέπει να είναι διαθέσιμη κατά τη φάση εκτέλεσης πέντε αλγορίθμων CV, των *absolute differences*, *normalized Ads*, *aggregation*, *min disparity search* και *interpolation*.

Αυτός ο απαιτητικός αριθμός αποθήκευσης, ο οποίος δεν είναι διαθέσιμος στο σύστημα SPARTAN επιβάλλει την εφαρμογή ενός αριθμού από αλγοριθμικές βελτιστοποιήσεις και τροποποιήσεις. Συγκεκριμένα για το έργο SPARTAN, όπου το τελικό σύστημα επιτρέπεται να χρησιμοποιήσει έως 500MB συνολικού χώρου αποθήκευσης, χρησιμοποιήθηκαν τεχνικές ελαχιστοποίησης του μεγέθους δέσμευσης μνήμης, όπως η βελτιστοποίηση μήκους λέξης και η επεξεργασία της εικόνας ανά τμήματα. Πληροφορίες για τη δεύτερη τεχνική, αναφορικά προς την χρήση της σε έναν από τους αλγορίθμους του έργου SPARTAN, μπορούν να βρεθούν στην εργασία [60]

3.4.4.3. Εκτιμήσεις σχετικά με την κατάτμηση υλικού/λογισμικού

Στα πλαίσια αυτής της εργασίας, παρέχονται επίσης εκτιμήσεις σχετικά με την κατάτμηση υλικού/λογισμικού. Συγκεκριμένα, εφαρμόζοντας την προτεινόμενη μεθοδολογία συν-σχεδίασης υλικού/λογισμικού, αναζητείται η ανάθεση των αλγορίθμων στους διαθέσιμους υπολογιστικούς πόρους του συστήματος. Για παράδειγμα, για το έργο SPARTAN, η διαδικασία αυτή καλείται να διαχωρίσει τους

αλγορίθμους που θα ανατεθούν είτε στον επεξεργαστή CPU είτε στη συσκευή FPGA. Η είσοδος της διαδικασίας αυτής είναι τα αποτελέσματα χαρακτηρισμού που εξήχθησαν από τη προηγούμενη φάση. Μεταξύ άλλων, λαμβάνονται υπόψιν οι σχεδιαστικές παράμετροι που επηρεάζουν την υπολογιστική πολυπλοκότητα, τις απαιτήσεις μνήμης (μέγεθος αποθήκευσης και διάρκεια ζωής δεδομένων) καθώς και το κόστος επικοινωνίας. Τα αποτελέσματα από αυτή την ανάλυση για τις λειτουργίες χαρτογράφησης και εντοπισμού απεικονίζονται στα Σχήματα 3.41 και 3.42, αντίστοιχα.

Σύμφωνα με αυτή την ανάλυση, οι αλγοριθμικοί πυρήνες για τη λειτουργία της χαρτογράφησης ομαδοποιούνται ως εξής:

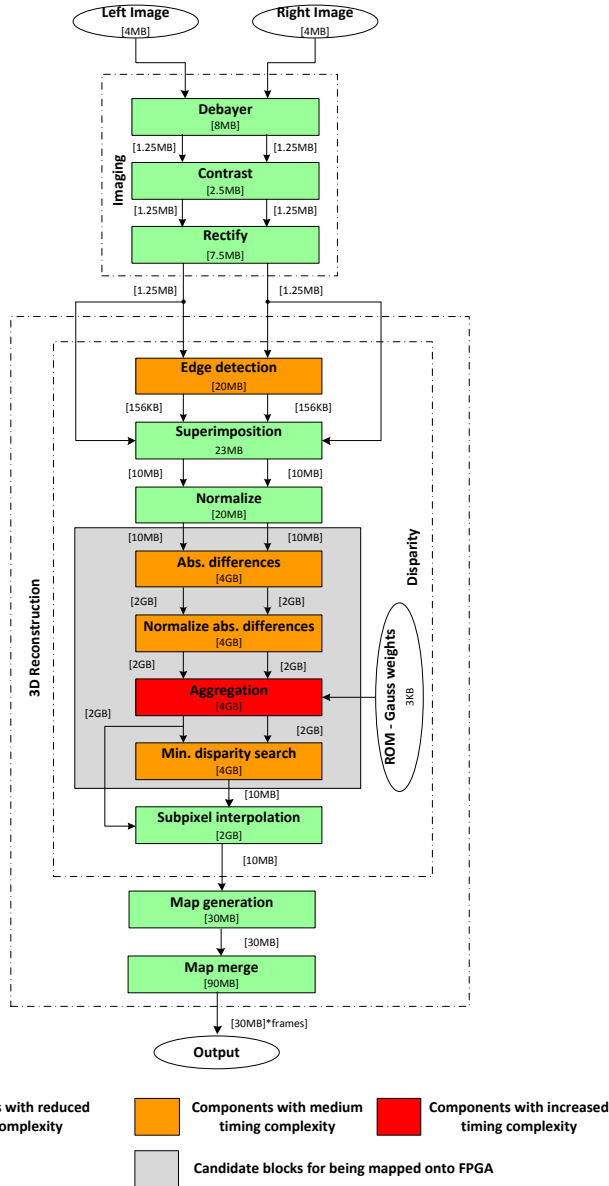
- Πυρήνες με μειωμένη εκτελεστική πολυπλοκότητα²: *debayer, contrast, rectify, superposition, normalize, subpixel interpolation, mapgen and mapmerge*.
- Πυρήνες με μέτρια εκτελεστική πολυπλοκότητα: *edge detection, absolute differences, normalized absolute differences and minimum disparity search*.
- Πυρήνες με αυξημένη εκτελεστική πολυπλοκότητα: *aggregation*.

Ομοίως, σχετικά με τη λειτουργία εντοπισμού, η αντίστοιχη ομαδοποίηση είναι η εξής:

- Πυρήνες με μειωμένη εκτελεστική πολυπλοκότητα: *debayer, contrast, rectify, superposition, normalize, subpixel interpolation, landmark 3D reconstruction, motion estimation and VSLAM*.
- Πυρήνες με μέτρια εκτελεστική πολυπλοκότητα: *edge detection, absolute differences, normalized absolute differences, minimum disparity search, landmark detection and landmark matching*.
- Πυρήνες με αυξημένη εκτελεστική πολυπλοκότητα: *aggregation*.

Η προαναφερθείσα ανάλυση δείχνει ότι το ποσοστό του χρόνου εκτέλεσης ενός αλγορίθμου δεν είναι απαραίτητα το μοναδικό κριτήριο για τη διαδικασία του συν-σχεδιασμού υλικού/λογισμικού σε ετερογενή πλατφόρμες. Για παράδειγμα, από το Σχήμα 3.41 φαίνεται ότι βάση της πολυπλοκότητας χρόνου εκτέλεσης, ο πυρήνας που πρέπει να ανατεθεί στο FPGA προς επιτάχυνση είναι ο πυρήνας *aggregation*. Ωστόσο η μεμονωμένη εξαγωγή του από τη ροή εκτέλεσης του επεξεργαστή και η ταυτόχρονη τοποθέτησή του στη συσκευή FPGA, επιφέρει στο σύστημα κόστος μεταφοράς τουλάχιστον 4 GB δεδομένων. Δεδομένης της Ethernet σύνδεσης PC-FPGA, καθίσταται σχεδόν σίγουρο ότι το κέρδος χρόνου εκτέλεσης από την επιτάχυνση του αλγορίθμου *aggregation* στο FPGA επισκιάζεται από την ανάγκη για είσοδο/έξοδο δεδομένων μεταξύ του PC και της συσκευής FPGA. Για αυτό το λόγο το προτεινόμενο πλαίσιο παρέχει ένα

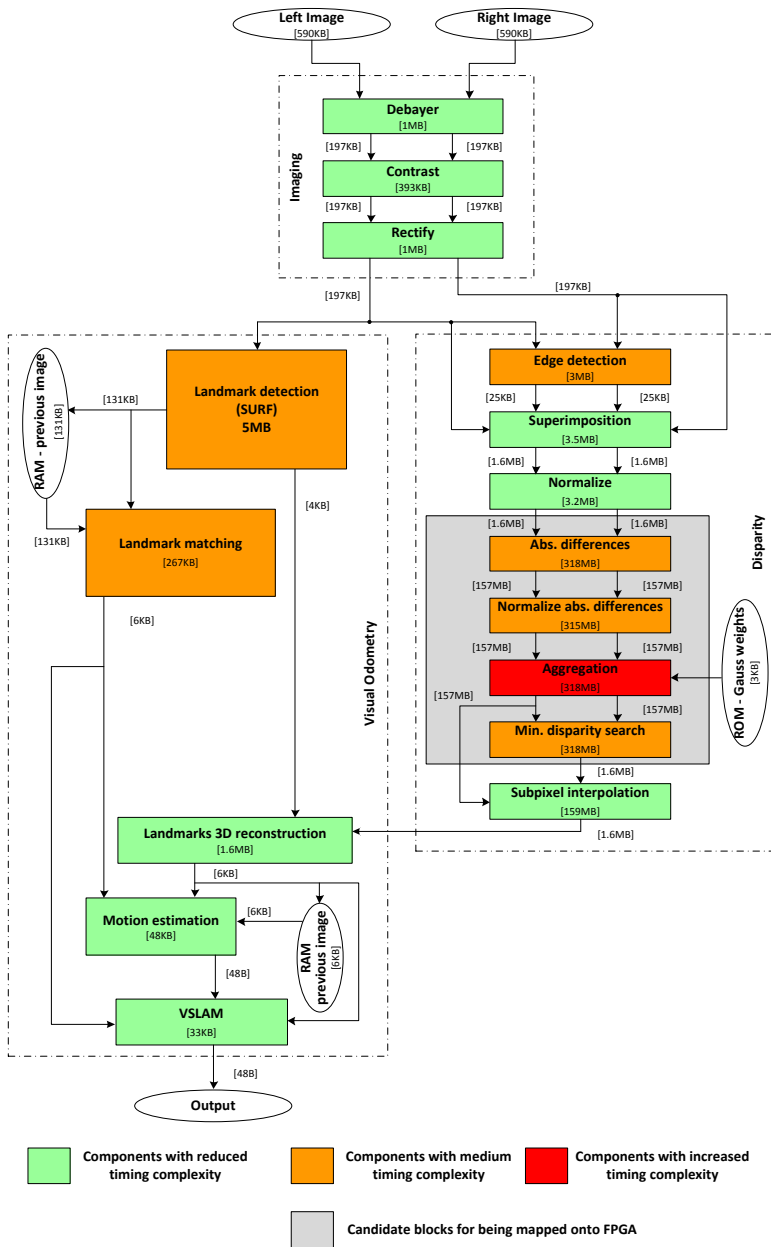
²Ο όρος “εκτελεστική πολυπλοκότητα” ενός αλγορίθμου αναφέρεται στο συνολικό κόστος υλοποίησης του αλγορίθμου αυτού αναφορικά με τη συνολική ροή εκτέλεσης. Αποτιμάται υπολογίζοντας το χρόνο εκτέλεσης, τη δέσμευση μνήμης και το μέγεθος και τη συχνότητα των δεδομένων εισόδου/εξόδου του αλγορίθμου.



Σχήμα 3.41 Αδρομερής κατάτμηση υλικού/λογισμικού για τη λειτουργία της χαρτογράφησης του ρομπωτικού οχήματος του έργου SPARTAN.

σύνολο από εναλλακτικές επιλογές κατάτμησης των αλγορίθμων οι οποίες αξιολογούνται ως προς το κόστος επιβάρυνσης στο χρόνο επεξεργασίας, στο χρόνο επικοινωνίας και στο μέγεθος δέσμμευσης μνήμης για το σύστημα συνολικά.

Το αποτέλεσμα από τη μελέτη αυτή τονίζεται στα Σχήματα 3.41 και 3.42



Σχήμα 3.42 Αδρομερής κατάτμηση υλικού/λογισμικού για τη λειτουργία του εντοπισμού του ρομποτικού σχήματος του έργου SPARTAN.

με γκρι χρώμα. Συγκεκριμένα, αναθέτοντας τους πυρήνες *absolute differences*, *normalize absolute differences*, *aggregation* και *minimum disparity search* στην συσκευή

FPGA, η προκύπτουσα αρχιτεκτονική απαιτεί μόνο 30MB δεδομένων μεταφοράς από τη CPU στην FPGA. Επίσης καθώς η FPGA έχει περιορισμένη ενσωματωμένη μνήμη, έχει σχεδιαστεί κατάλληλη αρχιτεκτονική σωλήνωσης η οποία επεξεργάζεται την εικόνα κατά τμήματα, ενώ παράλληλα χρησιμοποιείται τεχνική προανάκλησης, ώστε να επισιαίζεται το κόστος επικοινωνίας των τμημάτων εικόνας με την επεξεργασία αυτών. Ακριβείς λεπτομέρειες για αυτή την αρχιτεκτονική μπορούν να βρεθούν στην εργασία [60]. Επίσης σε αυτό το στάδιο αξιολογούνται τεχνικές επαναχρησιμοποίησης των δεδομένων για τους πυρήνες που ενδεχομένως επεξεργάζονται κοινά δεδομένα σε ίδιες, η ακόμα και διαφορετικές, χρονικές στιγμές.

3.4.4.4. Ανάλυση ροής δεδομένων

Η ανάλυση της ροής δεδομένων (data-flow analysis) για τις λειτουργίες χαρτογράφησης και εντοπισμού απεικονίζεται στα Σχήματα 3.41 και 3.42, αντίστοιχα. Όσον αφορά τη λειτουργία χαρτογράφησης, κανένας από τους πυρήνες της αλυσίδας εκτέλεσης δεν μπορεί να εκτελεστεί παράλληλα, καθώς υπάρχει σαφής εξάρτηση δεδομένων ανάμεσά τους. Πάραυτα, το συμπέρασμα αυτό επιβάλλει την -κατά το περισσότερο δυνατό- εκμετάλλευση του εγγενή παραλληλισμού που βρίσκεται μέσα σε κάθε πυρήνα, προκειμένου να ικανοποιηθούν οι χρονικές προδιαγραφές του έργου. Τεχνικές όπως η σωλήνωση (pipelining) και η προανάκληση (prefetching) μπορούν να βοηθήσουν ουσιαστικά στη περίπτωση που είναι αδύνατος ο παραλληλισμός σε επίπεδο διεργασιών (task-level pipelining). Αντίθετα, από το Σχήμα 3.42, το οποίο αναφέρεται στη λειτουργία εντοπισμού, είναι σαφές ότι οι πυρήνες *disparity* και *visual odometry* είναι υποψήφιοι για παράλληλη εκτέλεση, καθώς δεν υπάρχει εξάρτηση δεδομένων ανάμεσά τους. Ωστόσο, εκτός από αυτή τη διαπίστωση, όπως και στη λειτουργία χαρτογράφησης, είναι δυνατή η περαιτέρω επιτάχυνση της εκτέλεσης εφαρμόζοντας τις προαναφερθείσες τεχνικές.

3.4.5. Συμπεράσματα

Η ενότητα αυτή αποτυπώνει ένα μεθοδολογικό πλαίσιο για τον αποτελεσματικό συν-σχεδιασμό υλικού/λογισμικού και την κατάτμηση των επεξεργαστικών διεργασιών στους διαθέσιμους υπολογιστικούς πόρους ενός ετερογενούς συστήματος, αποτελούμενο από επεξεργαστές γενικού σκοπού και συσκευές επαναδιαμορφώμενης αρχιτεκτονικής, βάση μιας διαδικασίας χαρακτηρισμού των εφαρμογών. Η προτεινόμενη συστηματική μεθοδολογία εξετάζει την υπολογιστική πολυπλοκότητα, τις απαιτήσεις μνήμης (μέγεθος αποθήκευσης και διάρκεια ζωής δεδομένων) καθώς και το κόστος επικοινωνίας, ώστε να βρεθεί η βέλτιστη αρχιτεκτονική αδρομερών επιταχυντών υλικού ως μέσο επιτάχυνσης της συνολικής απόδοσης του συστήματος. Η προσέγγιση αυτή χρησιμοποιήθηκε στη σχεδίαση ενός ετερογενούς συστήματος, το οποίο υποστηρίζει αυτόνομη πλοήγηση σε διαστημικά ρομποτικά οχήματα μέσω χρήσης αλγορίθμων μηχανικής όρασης και το υλοποιημένο σύστημα κατάφερε να επιτύχει τις προδιαγραφές που τέθηκαν από τον οργανισμό ESA, στα πλαίσια του ερευνητικού έργου SPARTAN [61].

3.5. 3-D SoC Αρχιτεκτονικές Ψηφιακής Επεξεργασίας Σήματος

3.5.1. Εισαγωγή

Τα τελευταία χρόνια η τρισδιάστατη τεχνολογία ολοκλήρωσης ημιαγωγών 3-D κερδίζει συνεχές ενδιαφέρον από την ερευνητική και βιομηχανική κοινότητα, λόγω της υπόσχεσης για μεγάλη υπολογιστική ολοκλήρωση ανά μονάδα επιφάνειας πυριτίου καθώς και ετερογενή ολοκλήρωση τεχνολογιών στο ίδιο ολοκληρωμένο κύκλωμα [62, 63]. Ακολουθώντας τις τεχνολογικές εξελίξεις στην ολοκλήρωση 3-D, αντίστοιχα έχουν δημοσιευτεί πλήθος εργασιών που αφορούν το φυσικό τρισδιάστατο πρόβλημα σχεδιασμού. Μεταξύ άλλων, έχουν προταθεί εργαλεία για κατάτμηση (partitioning), χωροθέτηση (floor-plan), τοποθέτηση (partitioning) και δρομολόγηση (routing) τρισδιάστατων αρχιτεκτονικών. Οι προσεγγίσεις αυτές βασίζονται σχεδόν αποκλειστικά σε ακαδημαϊκά εργαλεία. Αντίθετα, στη βιομηχανική κατεύθυνση, το μόνο γνωστό εμπορικό πλαίσιο για υποστήριξη σχεδιασμού τρισδιάστατων SoCs παρέχεται από την εταιρεία R3Logic Corp.[64]. Σε αυτή την εργασία εισάγεται ένα νέο πλαίσιο για την υποστήριξη ταχείας αξιολόγησης τρισδιάστατων SoCs με τη χρήση υφιστάμενων εργαλείων σχεδίασης CAD. Το προτεινόμενο πλαίσιο είναι απαραίτητο ακόμα και πριν από τη διαθεσιμότητα εμπορικών εργαλείων φυσικού σχεδιασμού 3-D, εφόσον παρέχει μια καλή εκτίμηση για τις αναμενόμενες βελτιστοποιήσεις στις μετρικές απόδοσης σχεδίασης, από την υιοθέτηση της τρισδιάστατης τεχνολογίας ολοκλήρωσης.

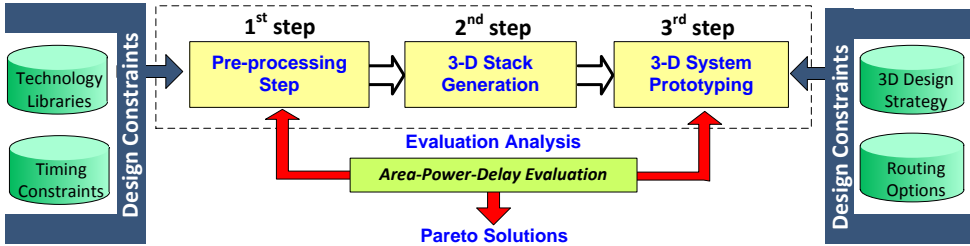
Σημειώνεται ότι ένα τμήμα της μεθοδολογίας σχεδιασμού που παρουσιάζεται στη παρούσα εργασία αποτελεί επίσης τμήμα της σχεδιαστικής ροής του πλαισίου *Plug&Chip* που παρουσιάστηκε στο Κεφάλαιο 2 και αναφέρεται στην υποενότητα 2.3.2. Ωστόσο η έτερη εργασία αποτελεί ένα προαιρετικό βήμα για την μεθοδολογία ταχείας προτυποποίησης που προτείνεται από τη διδακτορική διατριβή, ενώ η παρούσα εργασία αποτελεί μία συστηματική μελέτη για τη σχεδίαση 3-D αρχιτεκτονικών επεξεργασίας σήματος, οπότε και προκύπτει ο εννοιολογικός και συγγραφικός διαχωρισμός αυτών.

3.5.2. Προτεινόμενο μεθοδολογικό πλαίσιο για 3-D SoC

Αυτή η υποενότητα παρουσιάζει το προτεινόμενο πλαίσιο για ταχεία εκτίμηση της απόδοσης σχεδιασμών 3-D SoCs. Το πλαίσιο, το οποίο απεικονίζεται στο Σχήμα 3.43, αποτελείται από τρία διακριτά στάδια προκειμένου να συστηματοποιήσει τη διαδικασία σχεδίασης, χρησιμοποιώντας τόσο υφιστάμενα συμβατικά εργαλεία σχεδιασμού 2-D, όσο και νέα εργαλεία που επιτρέπουν τη μετάβαση στη τρίτη διάσταση σχεδίασης. Συγκεκριμένα, τα στάδια σχεδίασης συνοψίζονται ως εξής:

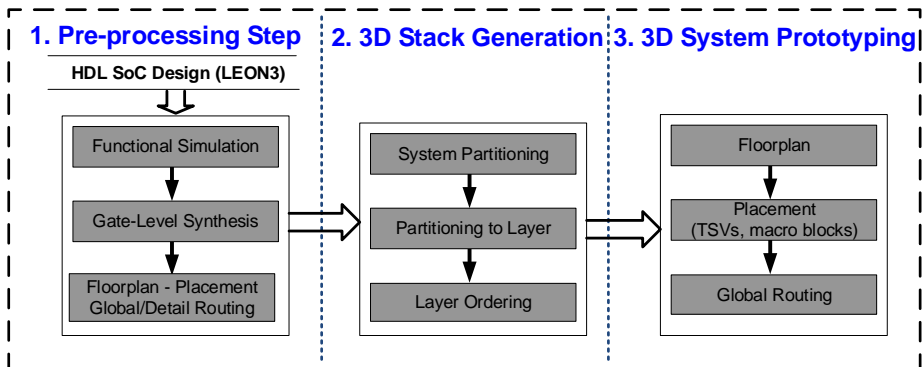
- *Προ-επεξεργασία*: Επαλήθευση της λειτουργικής ακεραιότητας του σχεδιασμού και εξαγωγή σε περιγραφή XML.
- *Δημιουργία 3-D στοίβας*: Δημιουργία της 3-D στοίβας και καθορισμός επικοινωνίας μεταξύ των επιπέδων πυριτίου.

- *Συστημική προτυποποίηση 3-D*: Υλοποίηση σε φυσικό επίπεδο μάσκας πυριτίου του 3-D SoC και επαλήθευση ορθής λειτουργίας του τελικού συστήματος.



Σχήμα 3.43 Προτεινόμενο μεθοδολογικό πλαίσιο για τη υποστήριξη ταχείας αξιολόγησης 3-D SoC συστημάτων.

Το Σχήμα 3.44 απεικονίζει μία λεπτομερή ροή εκτέλεσης των βημάτων του κάθε σταδίου της μεθοδολογίας.



Σχήμα 3.44 Αναλυτική ροή εκτέλεσης εργασιών του προτεινόμενου 3-D πλαισίου.

Το πρώτο βήμα της προτεινόμενης μεθοδολογία αφορά την λειτουργική επαλήθευση του σχεδιασμού. Συνεπώς είσοδος της ροής εργασιών αποτελεί η περιγραφή του σχεδιασμού, σε συνθέσιμη μορφή, σε μία γλώσσα HDL. Ο σχεδιασμός εξομοιώνεται με διάφορους περιορισμούς (όπως π.χ. η περίοδος του ρολογιού, η οργάνωση των μνημών, η ιεραρχία του σχεδιασμού κτλ.) προκειμένου να επαληθευτεί η λειτουργικότητα του συστήματος. Για το σκοπό αυτό, χρησιμοποιείται ο εξομοιωτής *Cadence NC-sim*.

Στη συνέχεια, καθορίζεται η ιεραρχία της 3-D αρχιτεκτονικής. Διαφορετικά σενάρια επιπέδων ιεραρχίας υποστηρίζονται από το προτεινόμενο πλαίσιο, το καθένα από τα οποία παρουσιάζει πλεονεκτήματα και μειονεκτήματα. Για παράδειγμα, η περιγραφή ενός συστήματος σε επίπεδο λειτουργικής μονάδας οδηγεί

γεί σε αδρομερή ιεραρχία, ενώ η περιγραφή σε επίπεδο πύλης οδηγεί σε λεπτομερή ιεραρχία. Όσο πιο λεπτομερής είναι η ιεραρχία του σχεδιασμού, τόσο μεγαλύτερη ευελιξία μπορεί να επιτευχθεί από τα εργαλεία σχεδίασης, καθώς όλος ο σχεδιασμός επεξεργάζεται εξολοκλήρου δίχως επί μέρους περιορισμούς. Ωστόσο αυτή η προσέγγιση αυξάνει την υπολογιστική πολυπλοκότητα για την εξερεύνηση της αρχιτεκτονικής 3-D. Πρέπει να σημειωθεί ότι λόγω της ιδιαιτερότητας αυτής της επιλογής, πρέπει να εφαρμοστεί προσεκτική μελέτη για την επιλογή της ιεραρχίας, καθώς μία υπο-βέλτιστη λύση θα μπορούσε να επισιχιάσει τη βελτίωση των επιδόσεων που προσφέρει η 3-D ολοκλήρωση.

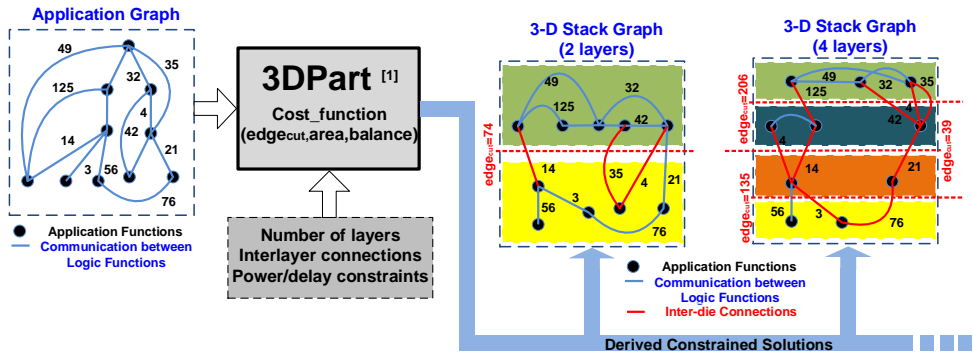
Επειτα από τον καθορισμό της ιεραρχίας του συστήματος, ο σχεδιασμός συντίθεται με το εργαλείο Synopsys Design Compiler. Εφόσον σε αυτό το στάδιο επιτευχθούν οι ζητούμενοι περιορισμοί, η έξοδος της σύνθεσης μεταφράζεται σε μία αντίστοιχη περιγραφή XML. Αυτή η διαδικασία υλοποιείται μέσω ενός νέου εργαλείου, υπό την ονομασία Net2XML. Η προκύπτουσα XML περιγραφή χρησιμοποιείται ως είσοδος στο δεύτερο βήμα του προτεινόμενου πλαισίου που ασχολείται με τη δημιουργία της τρισδιάστατης στοίβας, σύμφωνα με επιλεγμένους σχεδιαστικούς περιορισμούς.

Αρχικά, ο σχεδιασμός σε επίπεδο RTL netlist κατατάσσεται σε έναν αριθμό υποσυνόλων (system partitioning). Διαφορετικοί στόχοι βελτιστοποίησης μπορούν να τεθούν κατά τη διάρκεια αυτής της διαδικασίας, όπως για παράδειγμα η ελαχιστοποίηση των συνδέσεων μεταξύ των επιπέδων, τηρώντας ταυτόχρονα κάποιους περιορισμούς, όπως είναι η διατήρηση της DRAM και της λογικής σε διαφορετικά διαμερίσματα. Προηγούμενες μελέτες έχουν δείξει ότι η διαδικασία της κατάτμησης μπορεί να προσφέρει αυξημένη ευελιξία στη συνολική σχεδίαση, όταν ο αριθμός των υποσυνόλων είναι υψηλότερη σε σύγκριση με τον αντίστοιχο αριθμό των επιπέδων πυριτίου [65].

Το επόμενο βήμα είναι η ανάθεση των υποσυνόλων στα επίπεδα πυριτίου του 3-D συστήματος (partitioning to layer assignment). Κατά τη διάρκεια αυτού του σταδίου, σημαντικό ρόλο έχουν η διαδικασία παραγωγής μέσω της αξιοποιήσιμης ποσότητας πυριτίου ανά παραγωγική μονάδα (yield) και οι παράμετροι κόστους μέσω του αριθμού των TSVs. Συγκεκριμένα, σε κάθε επίπεδο πυριτίου μπορούν να ανατεθούν μόνο τεχνολογικά συμβατές μονάδες, ενώ κάθε ένα από τα επίπεδα πρέπει να υφίσταται επαρκής αξιοποιήσιμη επιφάνεια πυριτίου. Ακολουθεί η κατασκευή ενός πρωτότυπου του σχεδιασμού εφόσον ληφθεί απόφαση για τη διάταξη της σειράς των επιπέδων στη 3-D στοίβα (layer ordering) και για την επιλογή της 3-D τεχνολογίας συγκόλλησης. Παρόλο που το προτεινόμενο πλαίσιο μπορεί να αξιολογήσει εναλλακτικές τεχνολογίες συγκόλλησης, όπως η TSV (Face-to-back), η Face-to-Face, καθώς και η wire-bond, στη παρούσα εργασία παρέχονται αποτελέσματα μόνο για την προσέγγιση TSV, λόγω του ότι οδηγεί στη μέγιστη βελτίωση των επιδόσεων [29], [37].

Οι διαδικασίες της κατάτμησης του σχεδιασμού, της ανάθεσης των υπομονάδων στα επίπεδα πυριτίου και της διάταξης της σειράς των επιπέδων στη 3-D στοίβα, υποστηρίζονται από νέα εργαλεία που βασίζονται στον προτεινόμενο αλγόριθμο *Tabu*. Μια πρώτη έκδοση αυτού του αλγορίθμου παρουσιάστηκε στην εργασία [66], αλλά για τις ανάγκες της παρούσας εργασίας έχει επεκταθεί

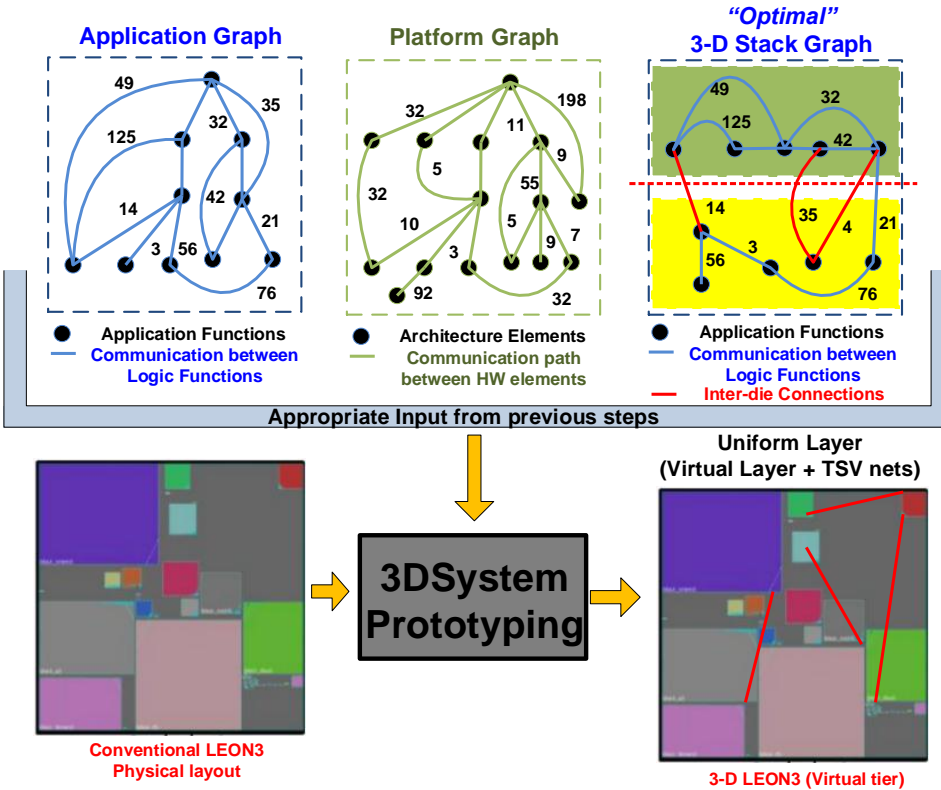
προκειμένου να καλύπτει τα ειδικά χαρακτηριστικά βελτιστοποίησης των 3-D ASIC πλατφορμών. Συγκεκριμένα το τροποποιημένο εργαλείο δέχεται ως είσοδο ένα σχεδιασμό σε επίπεδο διασύνδεσης πύλης (gate netlist) και κατασκευάζει το γράφο της εφαρμογής (application graph). Οι κόμβοι του γράφου αποτελούνται από τις λογικές μονάδες του σχεδιασμού, ανάλογα με την επιλογή ιεραρχίας που έχει καθοριστεί σε προγενέστερο στάδιο (αδρομερής - λεπτομερής). Αντίστοιχα οι ακμές του γράφου χαρακτηρίζονται από τον αριθμό των συνδέσεων μεταξύ των κόμβων του γράφου. Το προτεινόμενο εργαλείο παίρνει ως είσοδο τον γράφο της εφαρμογής και ένα σύνολο σχεδιαστικών περιορισμών, όπως ο αριθμός των επιπέδων πυριτίου, ο μέγιστος αριθμός των TSVs μεταξύ των επιπέδων καθώς και περιορισμούς σχετικά με την κατανάλωση ενέργειας και τη χρονική καθυστέρηση της εφαρμογής και προσπαθεί να βρεί μία ευριστική λύση κατάτμησης και ανάθεσης των κόμβων στα επίπεδα πυριτίου του σχεδιασμού. Το σύνολο αυτής της ροής σχεδίασης απεικονίζεται στο Σχήμα 3.45



Σχήμα 3.45 Δημιουργία της 3-D στοίβας με το εργαλείο 3DPart, χρησιμοποιώντας τον αλγόριθμο βελτιστοποίησης *Tabu Search* [66].

Η διαδικασία κατάτμησης παρέχει ως έξοδο πληροφορίες σχετικά με το μέγεθος του σχεδιασμού που υλοποιείται σε κάθε επίπεδο πυριτίου, καθώς και την αντίστοιχη συνδεσιμότητα μεταξύ των επιπέδων. Αυτές οι πληροφορίες εισέρχονται στο νέο εργαλείο *XML2Net*, το οποίο αναθέτει μια σειρά από TSVs στα σήματα που διατρέχουν πολλαπλά επίπεδα πυριτίου. Σημειώνεται ότι για την δρομολόγηση ενός σήματος μεταξύ δύο διαφορετικών επιπέδων πυριτίου i και j , δεσμεύεται σημαντική επιφάνεια πυριτίου και στα δύο επίπεδα. Το προτεινόμενο πλαίσιο μπορεί να διαχειριστεί τόσο διακριτές συνδέσεις TSVs όσο και ομαδοποιημένες. Στην παρούσα εργασία μελετούνται οι δεύτερες, καθώς θέτουν λιγότερους περιορισμούς στο εργαλείο δρομολόγησης και επομένως προσφέρουν ταχύτερη και αποδοτικότερη ροή σχεδίασης. Οι συστοιχίες TSV που παρέχουν δυνατότητα ομαδικής σύνδεσης μεταξύ των γειτονικών επιπέδων i και j , συνδέονται μέσω μονοπατιών ειδικού σκοπού δρομολόγησης, τα οποία ονομάζονται **δίκτυα TSV**, (*TSV networks*). Όπως αναλύεται στην επόμενη υποενότητα, τα δίκτυα σχεδιάζονται στα συμβατικά εργαλεία με πρόσθετα στρώματα μετάλλου, τα οποία εμφανίζουν παραμετροποιησιμα χαρακτηριστικά *RLC*, έτσι ώστε να

προσομοιώνουν τα αντίστοιχα ηλεκτρικά χαρακτηριστικά των TSVs συγκεκριμένων βιβλιοθηκών και μοντέλων 3-D τεχνολογιών. Το Σχήμα 3.46 απεικονίζει τη διαδικασία της συστημικής προτυποποίησης 3-D.



Σχήμα 3.46 Συστημική προτυποποίηση 3-D.

Το τελευταίο στάδιο του προτεινόμενου πλαισίου ασχολείται με την υλοποίηση σε φυσικό επίπεδο μάσκας πυριτίου της στοίβας 3-D. Συγκεκριμένα, κατά τη διάρκεια αυτού του σταδίου πραγματοποιείται η χωροθέτηση των μονάδων του σχεδιασμού (floor-planning), η δημιουργία του δικτύου διανομής ενέργειας στο τσιπ (power/ground network generation), η τοποθέτηση των κελιών του σχεδιασμού, βάση της βιβλιοθήκης της τεχνολογίας (physical library cells placement) και η γενική/ειδική δρομολόγηση των σημάτων του σχεδιασμού (global/detail signal routing). Αυτές οι διαδικασίες επιτελούνται από το εργαλείο *Cadence SoC Encounter*. Στο προτεινόμενο πλαίσιο χρησιμοποιούνται συμβατικά βιομηχανικά εργαλεία σχεδίασης 2-D, και συνεπώς δεν είναι δυνατόν να τροποποιηθεί ο πηγαίος κώδικάς τους και κατ'επέκταση η λειτουργικότητά τους προκειμένου να υποστηρίξουν 3-D σχεδιασμούς. Ως εκ τούτου, εισάγονται νέες έννοιες, εργα-

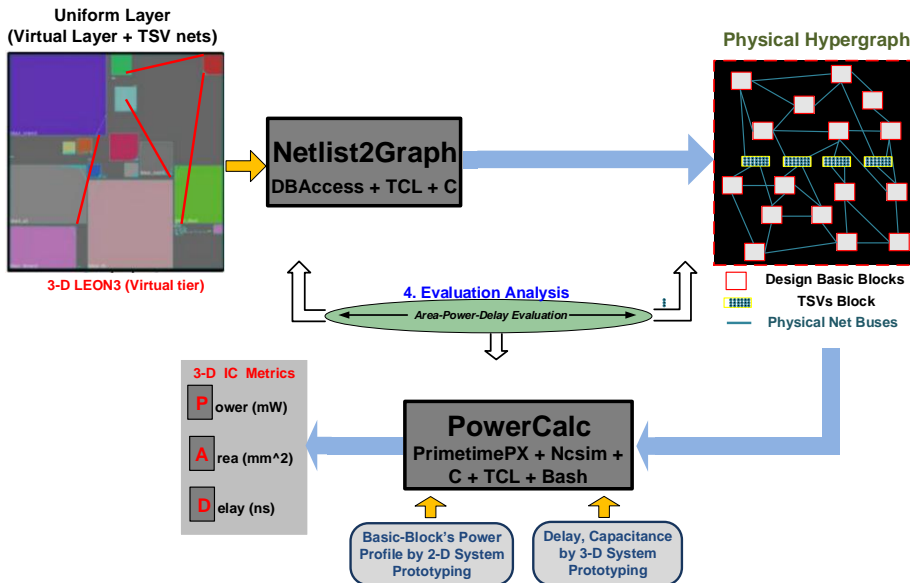
λεία και βιβλιοθήκες σχεδίασης έτσι ώστε να υποστηρίξουν την τρίτη διάσταση σχεδίασης, ως ακολούθως:

- *Εικονικά επίπεδα (Virtual layers)*: Το προτεινόμενο πλαίσιο ορίζει την έννοια του εικονικού επιπέδου προκειμένου να χαρακτηρίσει τις μονάδες υλικού που τοποθετούνται σε διαφορετικά επίπεδα του 3-D SoC. Συνεπώς καθώς το εργαλείο *Cadence SoC Encounter* αναγνωρίζει μόνο ένα επίπεδο πυριτίου, η έννοια των εικονικών επιπέδων επιτρέπει την συνύπαρξη στον ίδιο 2-D σχεδιασμό πολλαπλών επιπέδων πυριτίου, προσομοιώνοντας ένα 3-D σχεδιασμό.
- *Δίκτυα TSV*: Αυτά τα δίκτυα αναφέρονται στα μονοπάτια δρομολόγησης των μονάδων υλικού που ανήκουν σε διαφορετικά επίπεδα πυριτίου του 3-D σχεδιασμού. Εφόσον το εργαλείο *Cadence SoC Encounter* αναγνωρίζει μόνο ένα επίπεδο πυριτίου, οι συνδέσεις μεταξύ των εικονικών επιπέδων πραγματοποιούνται μέσω των ειδικών δικτύων TSV. Αυτό επιτυγχάνεται μέσω της διατήρησης της τοποθέτησης των δικτύων TSV στην ίδια σχετική γεωμετρική θέση (x, y) ανάμεσα στα γειτονικά εικονικά επίπεδα πυριτίου. Πρακτικά τα δίκτυα TSV υλοποιούνται ως επιπλέον επίπεδα μετάλλου στην βιβλιοθήκη τεχνολογίας, στα οποία αποδίδονται συγκεκριμένα ηλεκτρικά χαρακτηριστικά αντίστασης (R), χωρητικότητας (C) και επαγωγής (L) ανάλογα με τις παραμέτρους RLC της τεχνολογίας TSV που χρησιμοποιείται. Στα πλαίσια αυτής της εργασίας χρησιμοποιήθηκαν τα ηλεκτρικά χαρακτηριστικά TSV που βρίσκονται στην εργασία [67].

Ύστερα από το στάδιο της προτυποποίησης 3-D, ο σχεδιασμός αξιολογείται ως προς την λειτουργικότητά του με τη διαδικασία της ανάλυσης χρονισμού (timing analysis). Για το σκοπό αυτό χρησιμοποιείται το εργαλείο *Cadence Static Timing Analysis Engine*, ενώ για λόγους πληρότητας η ανάλυση γίνεται τόσο εκ των προτέρων, καθώς και μετά τη σύνθεση του δικτύου διανομής ρολογιού (clock tree synthesis) και τη δρομολόγηση του σχεδιασμού.

Σε αυτό το σημείο επαληθεύεται η λειτουργική ακεραιότητα του φυσικού σχεδιασμού εφαρμόζοντας εξομοίωση μετά τη φυσική τοποθέτηση του σχεδιασμού σε επίπεδο μάσκας (post-layout simulation), χρησιμοποιώντας το εργαλείο *Cadence Incisive Simulator*. Προς αυτή τη κατεύθυνση, εξάγεται η παρασιτική καθυστέρηση για όλα τα μονοπάτια δρομολόγησης της αρχιτεκτονικής σε μορφή αναπαράστασης SDF (Standard Delay Format) και με αυτό τον τρόπο ανακτώνται οι τιμές καθυστέρησης για τα δίκτυα TSV. Για τους σκοπούς αυτής της εργασίας, τα ηλεκτρικά χαρακτηριστικά των δικτύων TSV υιοθετούνται από δημοσιευμένα μοντέλα στις εργασίες [68] [69]. Το επόμενο βήμα της μεθοδολογίας είναι η αξιολόγηση των τρισδιάστατων SoCs αναφορικά προς την κατανάλωση ενέργειας. Για αυτή τη διαδικασία χρησιμοποιείται το υφιστάμενο εργαλείο *Synopsys PrimeTime PX*. Το εργαλείο αυτό μπορεί να εκτιμήσει την κατανάλωση ενέργειας ενός σχεδιασμού, βάση της φυσικής τοποθέτησης σε επίπεδο μάσκας πυριτίου. Επίσης δίδεται η δυνατότητα εισαγωγής της δραστηριότητας εκτέλεσης κάθε σήματος και λογικής μονάδας του σχεδιασμού, προκειμένου η

εκτίμηση κατανάλωσης ενέργειας να είναι περισσότερο ακριβής αναφορικά προς την εκτέλεση των εφαρμογών στο τελικό σύστημα. Για αυτό το λόγο αναπτύχθηκαν νέα κατάλληλα εργαλεία λογισμικού τα οποία αναλύουν τρία διαφορετικά αρχεία του σχεδιασμού, τα οποία έχουν παραχθεί σε διαφορετικές φάσεις σχεδίασης, και για κάθε λογική μονάδα και σήμα του σχεδιασμού, εισάγουν στο εργαλείο *Synopsys PrimeTime PX* την χρονική καθυστέρηση και την ηλεκτρική παρασιτική χωρητικότητα. Τα αντίστοιχα αρχεία εισόδου που αναλύονται, είναι το αρχείο μορφής VCD (Value Change Dump), καθώς και το αρχείο μορφής SPEF (Standard Parasitic Exchange Format). Οι παραπάνω διεργασίες και η ροή εκτέλεσης αυτών απεικονίζονται στο Σχήμα 3.47. Το σύνολο των νέων εργαλείων που αυτοματοποιούν τη διαδικασία αξιολόγησης ονομάζεται *PowerCalc*.



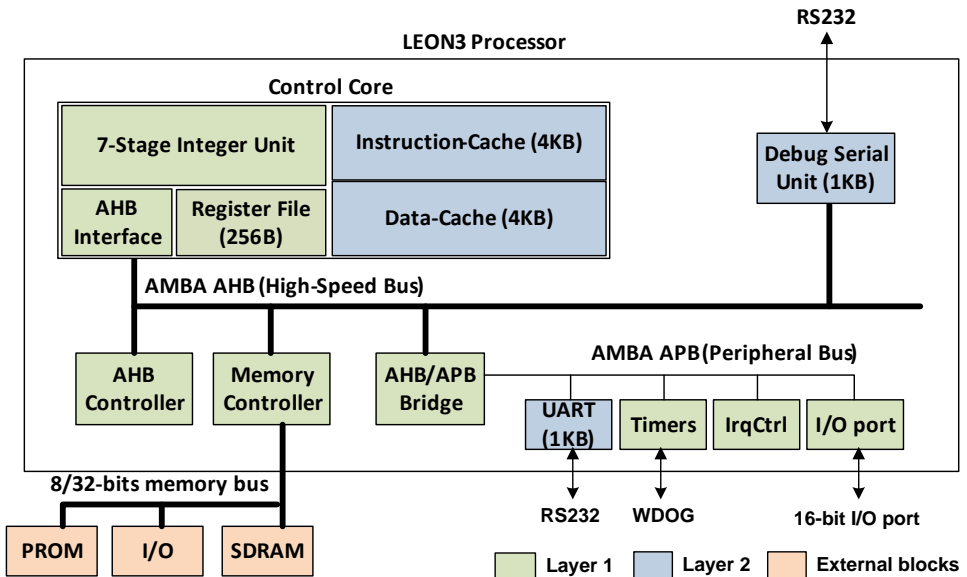
Σχήμα 3.47 Αξιολόγηση μετρικών απόδοσης σχεδιασμών 3-D με υφιστάμενα εμπορικά και νέα ακαδημαϊκά εργαλεία CAD.

3.5.3. Πειραματικά αποτελέσματα

Η ενότητα αυτή αναφέρει τα αποτελέσματα από την εφαρμογή του προτεινόμενου πλαισίου στο σχεδιασμό μιας τρισδιάστατης αρχιτεκτονικής, δύο επιπέδων πυριτίου, ενός συστήματος SoC βασισμένο στον επεξεργαστή LEON3 [18]. Ο επεξεργαστής LEON3, ο οποίος απεικονίζεται στο Σχήμα 3.48, είναι ένας ανοικτού κώδικας 32-bit επεξεργαστής πλήρως συνθέσιμος σε περιγραφή RTL και συμβατός με την αρχιτεκτονική SPARC V8 [70], ενώ παράλληλα είναι παραμετροποιήσιμος μέσω, VHDL generics, σε ένα πλήθος αρχιτεκτονικών και συστημικών παραμέτρων.

Στα πλαίσια αυτής της εργασίας, χρησιμοποιήθηκε ένα ενσωματωμένο σύ-

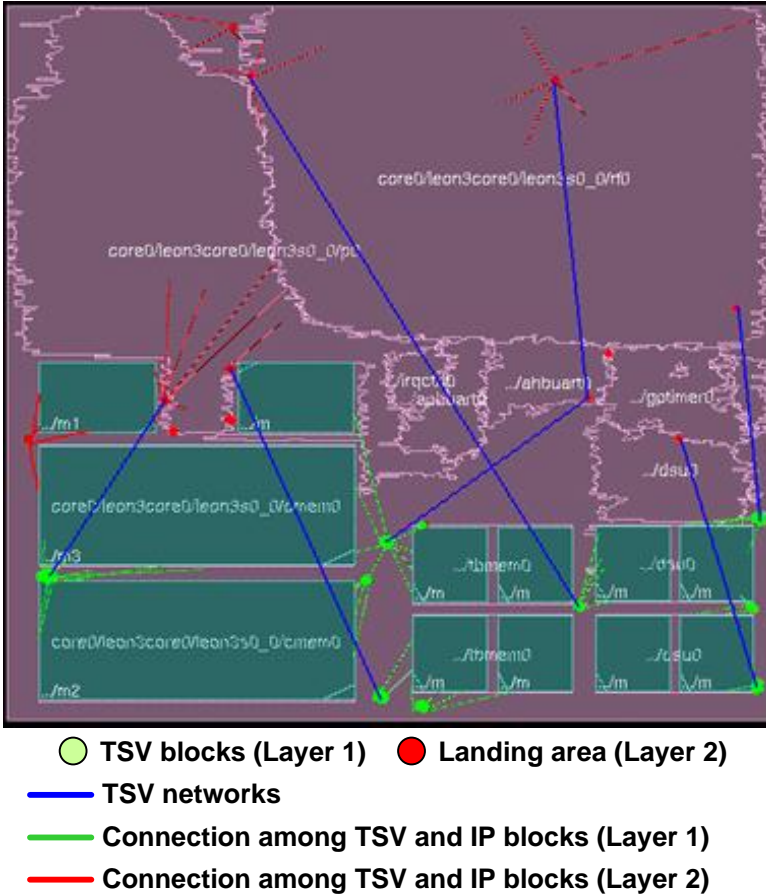
στημα βασισμένο σε ένα μονοπύρηνιο επεξεργαστή LEON3 προσκολλημένο, ως κύριο συντονιστή (master), σε ένα δίαυλο AMBA AHB (Advanced High-performance Bus). Ο επεξεργαστής έχει 7 στάδια σωλήνωσης, και ξεχωριστή κρυφή μνήμη δεδομένων και διευθύνσεων (αρχιτεκτονική Harvard), μεγέθους 4KB έκαστη. Εφόσον η επιλεγμένη αρχιτεκτονική αφορά ένα ενσωματωμένο σύστημα, η προτινόμενη μεθοδολογία έθεσε ως στόχο την διερεύνηση 3-D σχεδιαστικών λύσεων με χαμηλότερη κατανάλωση ενέργειας. Το Σχήμα 3.48 απεικονίζει μία κατάτμηση του υπο-σχεδίαση συστήματος SoC, όπως προκύπτει μετά από το στάδιο δημιουργίας τρισδιάστατης στοιβάς, σύμφωνα με την ανάλυση που έγινε στην υποενότητα 3.5.2. Τα διαφορετικά χρώματα στο σχήμα αυτό δείχνουν λογικές μονάδες που ανατίθενται σε διαφορετικά εικονικά επίπεδα. Ωστόσο, αν και από το εργαλείο 3DPart μπορούν να προκύψουν εναλλακτικές λύσεις τρισδιάστατης χωροθέτησης, η επιλεγόμενη αντιστοιχεί στη βέλτιστη λύση Pareto, δηλαδή στη λύση που αποδίδει τη μέγιστη απόδοση, με το μικρότερο κόστος παραγωγής, αναφορικά προς την επιφάνεια πυριτίου κάθε επιπέδου και τον αριθμό των TSVs.



Σχήμα 3.48 Ενσωματωμένο σύστημα βασισμένο στον επεξεργαστή Leon3 [18].

Η σύνθεση του επεξεργαστή LEON3 πραγματοποιήθηκε με το εργαλείο *Synopsys Design Compiler* σε τεχνολογία CMOS 130nm με χρονικό περιορισμό 4.35ns, δηλαδή συχνότητα λειτουργίας 230MHz. Το προκύπτων σύστημα σε επίπεδο φυσικού σχεδιασμού αποτελείται από 38,988 λογικά κελιά, 42,626 δίκτυα συνδέσεων, και 110 θύρες εισόδου/εξόδου. Το Σχήμα 3.49 απεικονίζει το αποτέλεσμα χωροθέτησης, υποθέτοντας ένα τρισδιάστατο σύστημα SoC δύο επιπέδων πυριτίου. Στο σχήμα αυτό, οι κόκκινοι και πράσινοι κύκλοι απεικονίζουν τις συστοιχίες από TSVs καθώς και την επιφάνεια πυριτίου που δεσμεύεται γύρω

από αυτά για τα εικονικά επίπεδα $layer_1$ και $layer_2$, αντίστοιχα, ενώ τα δίκτυα TSV απεικονίζονται με μπλε γραμμές. Ομοίως, οι κόκκινες και πράσινες γραμμές αντιστοιχούν στις συνδέσεις μεταξύ των TSVs συστοιχιών με τις υπόλοιπες μονάδες υλικού οι οποίες βρίσκονται στα εικονικά επίπεδα $layer_1$ και $layer_2$, αντίστοιχα.



Σχήμα 3.49 Παράδειγμα 3-D σχεδιασμού για ένα ενσωματωμένο σύστημα βασισμένο στον επεξεργαστή LEON3 με το συμβατικό εργαλείο 2-D σχεδίασης Cadence SoC Encounter [24].

Ο Πίνακας 4.5 δίνει κάποιες τεχνικές λεπτομέρειες για τη φυσική υλοποίηση του επεξεργαστή LEON3. Βασισμένοι στα αποτελέσματα αυτά, συμπεραίνουμε ότι η προκύπτουσα τρισδιάστατη αρχιτεκτονική μειώνει το συνολικό μήκος καλωδίου κατά 36%, συγκρινόμενη με την αντίστοιχη συμβατική δυσδιάστατη υλοποίηση. Επειδή σε αυτή την εργασία, στόχος της σχεδίασης 3-D είναι η επίτευξη χαμηλής κατανάλωσης ενέργειας του ενσωματωμένου συστήματος, η μείωση του μήκους του καλωδίου αναμένεται να οδηγήσει και σε εξοικονόμηση ενέργειας, χωρίς ωστόσο να επηρεάσει την απόδοση της υλοποίησης αναφορικά προς τη

Πίνακας 3.4 Μετρικές σχετικά με τη φυσική υλοποίηση σε επίπεδο μάσκας πυριτίου.

Χαρακτηριστικά	Σύστημα 2-D	Σύστημα 3-D	
		Επίπεδο 1 (Λογική)	Επίπεδο 2 (Μνήμη)
Μήκος καλωδίου (μm)	855,637	530,443	100,440
Ημιπερίμετρος (μm)	823,033	495,010	97,527
Αριθμός TSVs	0.00	817	817
Επιφάνεια πυριτίου για TSVs (μm^2)	0.00	24.7×24.7	24.7×24.7
Αναλογία διαστάσεων	1.00	1.00	1.00
Επιφάνεια κάθε επιπέδου (mm^2)	2.89	1.30	1.53
Συχνότητα λειτουργίας (MHz)	230		230
Κατανάλωση ενέργειας (mWatt)	53.96		43.30

συχνότητα λειτουργίας.

Κατά τη φάση αξιολόγησης της προτεινόμενης μεθοδολογίας σχεδίασης, χρησιμοποιήθηκε ένα σύνολο από εφαρμογές, οι οποίες αποτελούν βασικούς υπολογιστικούς πυρήνες διάφορων εφαρμογών επεξεργασίας σήματος, όπως MPEG-4, JPEG, ψηφιακά φίλτρα και H.263. Συγκεκριμένα, χρησιμοποιήθηκαν πέντε αλγόριθμοι εκτίμησης κίνησης: full search (FS), hierarchical search (HS), three step logarithmic search (3SLOG), parallel hierarchical one-dimensional search (PHODS) και spiral search (SS). Σημειώνεται ότι η πολυπλοκότητά αυτών κυμαίνεται σε ποσοστά 60 έως 80% της συνολικής πολυπλοκότητας κωδικοποίησης βίντεο με το πρότυπο MPEG-4 [71]. Επιπλέον, χρησιμοποιήθηκε μονοδιάστατος μετασχηματισμός wavelet, ένας πυρήνας cavity detector και ένα πυρήνας μετασχηματισμού Fast Fourier (FFT). Τέλος στιστο στάδιο αξιολόγησης ενσωματώθηκαν επίσης βασικοί πυρήνες επεξεργασίας δεδομένων που συναντώνται σε πολλά προβλήματα υπολογιστικής πολυπλοκότητας όπως ο πολλαπλασιασμός πινάκων και ο αλγόριθμος ταξινόμησης Bubblesort.

Οι παραπάνω εφαρμογές υλοποιήθηκαν σε γλώσσα C και μεταγλωττίστηκαν σε εκτελέσιμα αρχεία του επεξεργαστή LEON3 με τον μεταγλωττιστή BCC[72]. Τα εκτελέσιμα αρχεία αρχικά εκτελέστηκαν στον εξομοιωτή, του επεξεργαστή LEON3, TSIM [73] σε επίπεδο εντολών για λειτουργική επαλήθευση της εκάστοτε εφαρμογής. Επίσης η διαδικασία εξομοίωσης επιλέχθηκε κατά τη φάση επαλήθευσης προκειμένου να εκτιμηθεί ο χρόνος εκτέλεσης και η δέσμευση πόρων προκειμένου να εκτελεστεί ορθά το στιγμιότυπο της εφαρμογής που επιλέχθηκε για αξιολόγηση. Διαφορετικά, σε πολλές περιπτώσεις θα ήταν αδύνατη η εξαγωγή συμπερασμάτων από στιγμιότυπα εκτέλεσης που είτε θα απαιτούσαν περισσότερους υπολογιστικούς πόρους από τους διαθέσιμους του ενσωματωμένου συστήματος, είτε ο χρόνος εκτέλεσης θα ήταν απαγορευτικά μεγάλος για το στάδιο αξιολόγησης της παρούσας εργασίας. Για παράδειγμα, η εξομοίωση σε επίπεδο RTL, ενός πυρήνας Cavity detector, έχοντας ως είσοδο μία εικόνα μεγέθους 64×64 pixels, απαιτεί χρόνο προσομοίωσης 69 ώρες, καθώς και 58GB χώρο σκληρού δίσκου για την αποθήκευση των ίχνων εκτέλεσης VCD και 800MB φυσική μνήμη RAM, σε ένα μηχάνημα υποδοχής με τον επεξεργαστή Intel Core2

Πίνακας 3.5 Χρόνος εξομοίωσης και κατανάλωση ενέργειας για διαφορετικούς αλγορίθμους επεξεργασίας σήματος.

Εφαρμογή	Κύκλοι clk	Χρόνος Εξ/ση(ms)	Ενέργεια 2D (mW)	Ενέργεια 3D (mW)	Κέρδος σε 3D (%)
3SLOG	19,748,213	157	59.828	48.405	19.09%
FS	7,368,117	71	43.763	35.260	19.43%
PHODS	13,334,375	120	59.515	48.144	19.11%
SS	7,439,388	72	57.585	46.944	18.48%
CAVITY	23,777,500	216	51.639	42.094	18.48%
HS	9,603,875	93	53.047	42.198	20.45%
MMUL	5,894,285	57	43.763	35.260	19.43%
FFT	8,428,398	91	51.065	40.685	20.33%
WAVELET	6,035,932	56	55.081	43.293	21.40%
BUBBLESORT	5,595,728	46	64.314	50.752	21.09%
Μέσος όρος:	10,722,581	97.9	53.960	43.304	19.73%

Duo και 4GB φυσική μνήμη RAM. Αντίθετα η εξομοίωση στον εξομοιωτή TSIM ολοκληρώνεται σε λιγότερο από 3 λεπτά. Συνεπώς, το μέγεθος των δεδομένων εισόδου για κάθε αλγόριθμο επιλέχθηκε κατάλληλα ώστε η εκτέλεσή του να ανταποκρίνεται σε λογικά πλαίσια του χρόνου επαλήθευσης των αποτελεσμάτων της παρούσας εργασίας.

Ο Πίνακας 3.4 παρέχει αποτελέσματα για την απόδοση και τη μέση κατανάλωση ενέργειας του συστήματος SoC, βασισμένου στον επεξεργαστή LEON3. Σημειώνεται ότι η απόδοση, αναφορικά προς τη συχνότητα λειτουργίας, μεταξύ δυοδιάστατων και τρισδιάστατων υλοποιήσεων είναι σταθερή, εφόσον στη παρούσα εργασία στόχος ήταν η ελαχιστοποίηση της κατανάλωσης ενέργειας. Η μείωση του μήκους καλωδίου που επιβλήθηκε από τη χρήση της τρισδιάστατης ολοκλήρωσης οδηγεί σε μέση εξοικονόμηση ενέργειας κατά ποσοστό 20%, συγκρινόμενη με την αντίστοιχη δυοδιάστατη υλοποίηση. Ο Πίνακας 4.5 παρέχει το χρόνο προσομοίωσης και την εξοικονόμηση ενέργειας για κάθε υλοποιημένη εφαρμογή.

3.5.4. Συμπεράσματα

Στη παρούσα ενότητα προτάθηκε ένα νέο πλαίσιο για την υποστήριξη γρήγορης αξιολόγησης τρισδιάστατων συστημάτων SoCs. Για τους σκοπούς της εργασίας αυτής, η προτεινόμενη μεθοδολογία εφαρμόζεται σε ένα ενσωματωμένο σύστημα βασισμένο στον επεξεργαστή LEON3. Σκοπός της εργασίας είναι η ανάπτυξη ενός μεθοδολογικού πλαισίου και των αντίστοιχων εργαλείων σχεδίασης προκειμένου να εκτιμηθεί το κέρδος σχεδιασμού από την υιοθέτηση της τρισδιάστατης τεχνολογίας ολοκλήρωσης. Πειραματικά αποτελέσματα από διαφορετικούς αλγορίθμους επεξεργασίας σήματος αποδεικνύουν την αποτελεσματικότητα της προτεινόμενης λύσης, καθώς αυτή οδηγεί σε μέση εξοικονόμηση ενέργειας 19.7% δίχως υποβάθμιση της απόδοσης, αναφορικά προς τη συχνότητα λειτουργίας του συστήματος.

Βιβλιογραφία

- [1] T. Ulversoy, *Software defined radio: Challenges and opportunities*, [Communications Surveys Tutorials](#), **IEEE 12**, 531 (2010).
- [2] J. Glossner, E. Hokenek, and M. Moudgill, *The sandbridge sandblaster communications processor*, in [Software Defined Radio](#) (John Wiley and Sons, Ltd, 2004) pp. 129–159.
- [3] Y. Lin, H. Lee, M. Woh, Y. Harel, S. Mahlke, T. Mudge, C. Chakrabarti, and K. Flautner, *Soda: A low-power architecture for software radio*, in [Computer Architecture, 2006. ISCA '06. 33rd International Symposium on](#) (2006) pp. 89–101.
- [4] A. K. Coskun, T. S. Rosing, K. Mihic, G. De Micheli, and Y. Leblebici, *Analysis and optimization of mp soc reliability*, [Journal of Low Power Electronics](#) **2**, 56 (2006).
- [5] V. Gektin, A. Bar-Cohen, and J. Ames, *Coffin-manson fatigue model of underfilled flip-chips*, [Components, Packaging, and Manufacturing Technology, Part A](#), [IEEE Transactions on](#) **20**, 317 (1997).
- [6] Y. Liu, H. Yang, R. P. Dick, H. Wang, and L. Shang, *Thermal vs energy optimization for ddfs-enabled processors in embedded systems*, in [Symp. on Quality Electronic Design \(ISQED07\), \(International Symposium on Quality Electronic Design, 2007. ISQED '07. 8th\):pp. 204–209](#) (2007).
- [7] M. Harchol-Balter and A. B. Downey, *Exploiting process lifetime distributions for dynamic load balancing*, [ACM Trans. Comput. Syst.](#) **15**, 253 (1997).
- [8] J. Yang, X. Zhou, M. Chrobak, Y. Zhang, and L. Jin, *Dynamic thermal management through task scheduling*, in [Performance Analysis of Systems and software, 2008. ISPASS 2008. IEEE International Symposium on](#) (2008) pp. 191–201.
- [9] D. Atienza, P. Del Valle, G. Paci, F. Poletti, L. Benini, G. De Micheli, and J. Mendias, *A fast hw/sw fpga-based thermal emulation framework for multi-processor system-on-chip*, in [Design Automation Conference, 2006 43rd ACM/IEEE](#) (2006) pp. 618–623.
- [10] M. M. Sabry, J. Ayala, and D. Atienza, *Thermal-aware compilation for system-on-chip processing architectures*, in [Proc. of 20th ACM Great Lakes Symposium on VLSI \(GLSVLSI'10\)](#) (2010).
- [11] M. Monchiero, R. Canal, and A. Gonzalez, *Power/performance/thermal design-space exploration for multicore architectures*, [Parallel and Distributed Systems](#), [IEEE Transactions on](#) **19**, 666 (2008).
- [12] J. Rabaey, *Low Power Design Essentials*, 1st ed. (Springer Publishing Company, Incorporated, 2009).

- [13] H. Li, P. Liu, Z. Qi, L. Jin, W. Wu, S.-D. Tan, and J. Yang, *Efficient thermal simulation for run-time temperature tracking and management*, in *Computer Design: VLSI in Computers and Processors, 2005. ICCD 2005. Proceedings. 2005 IEEE International Conference on* (2005) pp. 130–133.
- [14] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, *Temperature-aware microarchitecture: Extended discussion and results*, in *In Proceedings of the 30th Annual International Symposium on Computer Architecture* (2003) pp. 2–13.
- [15] D. Brooks, V. Tiwari, and M. Martonosi, *Wattch: a framework for architectural-level power analysis and optimizations*, in *Computer Architecture, 2000. Proceedings of the 27th International Symposium on* (2000) pp. 83–94.
- [16] D. Brooks and M. Martonosi, *Dynamic thermal management for high-performance microprocessors*, in *High-Performance Computer Architecture, 2001. HPCA. The Seventh International Symposium on* (2001) pp. 171–182.
- [17] Proteas-Software-Repository, (2013), <http://proteas.microlab.ntua.gr/ksiop/software>.
- [18] Gaisler, *Leon3*. ().
- [19] W. Huang, K. Sankaranarayanan, K. Skadron, R. Ribando, and M. Stan, *Accurate, pre-rtl temperature-aware design using a parameterized, geometric thermal model*, *Computers, IEEE Transactions on* **57**, 1277 (2008).
- [20] K. Sankaranarayanan, S. Velusamy, M. Stan, C. L, and K. Skadron, *A case for thermal-aware floorplanning at the microarchitectural level*, *Journal of ILP* **7** (2005).
- [21] A. Raghunathan, N. K. Jha, and S. Dey, *High-Level Power Analysis and Optimization* (Kluwer Academic Publishers, Norwell, MA, USA, 1998).
- [22] F. Zanini, D. Atienza, and G. De Micheli, *A control theory approach for thermal balancing of mp soc*, in *Design Automation Conference, 2009. ASP-DAC 2009. Asia and South Pacific* (2009) pp. 37–42.
- [23] S. Inc., <http://www.synopsys.com>, ().
- [24] Cadence, *Cadence inc.* .
- [25] P. PX, http://www.europpractice.stfc.ac.uk/vendors/primetime_px_brief.pdf, .
- [26] S. Sapatnekar, *Rc interconnect optimization under the elmore delay model*, in *Design Automation, 1994. 31st Conference on* (1994) pp. 387–391.
- [27] S. Canumalla and P. Viswanadham, *Portable Consumer Electronics: Packaging, Materials, and Reliability* (PennWell Corporation, 2010).

- [28] M. Guthaus, J. Ringenberg, D. Ernst, T. Austin, T. Mudge, and R. Brown, *Mibench: A free, commercially representative embedded benchmark suite*, in *Workload Characterization, 2001. WWC-4. 2001 IEEE International Workshop on* (2001) pp. 3–14.
- [29] ITRS, *International technology roadmap for semiconductors*, (2012).
- [30] J. Srinivasan, S. Adve, P. Bose, and J. Rivers, *Lifetime reliability: toward an architectural solution*, *Micro, IEEE* **25**, 70 (2005).
- [31] G. De Micheli and L. Benini, *Networks on chips: technology and tools* (Academic Press, 2006).
- [32] F. Gebali, H. Elmiligi, and M. W. El-Kharashi, *Networks-on-Chips: Theory and Practice*, 1st ed. (CRC Press, Inc., Boca Raton, FL, USA, 2009).
- [33] J. Hu and R. Marculescu, *Energy- and performance-aware mapping for regular noc architectures*, *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* **24**, 551 (2005).
- [34] E. Rijpkema, K. Goossens, A. Radulescu, J. Dielissen, J. van Meerbergen, P. Wielage, and E. Waterlander, *Trade offs in the design of a router with both guaranteed and best-effort services for networks on chip*, in *Design, Automation and Test in Europe Conference and Exhibition, 2003* (2003) pp. 350–355.
- [35] I. Anagnostopoulos, A. Bartzas, and D. Soudris, *Application-specific temperature reduction systematic methodology for 2d and 3d networks-on-chip*, in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation*, Lecture Notes in Computer Science, Vol. 5953, edited by J. Monteiro and R. van Leuken (Springer Berlin Heidelberg, 2010) pp. 86–95.
- [36] A. Richard, D. Milojevic, F. Robert, A. Bartzas, A. Papanikolaou, K. Siozios, and D. Soudris, *Fast design space exploration environment applied on noc's for 3d-stacked mpsoc's*, in *Architecture of Computing Systems (ARCS), 2010 23rd International Conference on* (2010) pp. 1–6.
- [37] V. F. Pavlidis and E. G. Friedman, *Three-dimensional Integrated Circuit Design* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2009).
- [38] CADENCE, *Cadence 3d-ic solution*, (2013).
- [39] C. J. Glass and L. M. Ni, *Fault-tolerant wormhole routing in meshes without virtual channels*, *Parallel and Distributed Systems, IEEE Transactions on* **7**, 620 (1996).
- [40] V. suite, <http://valgrind.org>, .
- [41] S. Gupta, M. Hilbert, S. Hong, and R. Patti, *Techniques for producing 3d ics with high-density interconnect*, in *Proceedings of the 21st International VLSI Multilevel Interconnection Conference* (2004).

- [42] S. Murali and G. De Micheli, *Bandwidth-constrained mapping of cores onto noc architectures*, in *Design, Automation and Test in Europe Conference and Exhibition, 2004. Proceedings*, Vol. 2 (2004) pp. 896–901 Vol.2.
- [43] C. S. Compiler, http://www.cadence.com/products/sd/silicon_compiler/pages/default.aspx, .
- [44] E. Sotiriou-Xanthopoulos, D. Diamantopoulos, K. Siozios, G. Economakos, and D. Soudris, *A framework for rapid evaluation of heterogeneous 3-d noc architectures*, *Microprocessors and Microsystems* **38**, 292 (2014).
- [45] V.-D. Ngo, H.-N. Nguyen, and H.-W. Choi, *The optimum network on chip architectures for video object plane decoder design*, in *Parallel and Distributed Processing and Applications*, Lecture Notes in Computer Science, Vol. 4330, edited by M. Guo, L. Yang, B. Di Martino, H. Zima, J. Dongarra, and F. Tang (Springer Berlin Heidelberg, 2006) pp. 75–85.
- [46] D. Bertozzi, A. Jalabert, S. Murali, R. Tamhankar, S. Stergiou, L. Benini, and G. De Micheli, *Noc synthesis flow for customized domain specific multiprocessor systems-on-chip*, *Parallel and Distributed Systems, IEEE Transactions on* **16**, 113 (2005).
- [47] I. E. Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia* (John Wiley & Sons, 2004).
- [48] M. Di Marco, A. Garulli, S. Lacroix, and A. Vicino, *Set membership localization and mapping for autonomous navigation*, *International Journal of robust and nonlinear control* **11**, 709 (2001).
- [49] *Esa nasa exomars programme*, .
- [50] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, *Ros: an open-source robot operating system*, *ICRA workshop on open source software* **3**, 5 (2009).
- [51] K. Siozios, D. Diamantopoulos, I. Kostavelis, E. Boukas, L. Nalpantidis, D. Soudris, A. Gasteratos, M. Aviles, and I. Anagnostopoulos, *Spartan project: Efficient implementation of computer vision algorithms onto reconfigurable platform targeting to space applications*, in *Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC), 2011 6th International Workshop on* (2011) pp. 1–9.
- [52] R. lab Duth., <http://robotics.pme.duth.gr>, .
- [53] L. Nalpantidis, G. Sirakoulis, and A. Gasteratos, *A dense stereo correspondence algorithm for hardware implementation with enhanced disparity selection*, in *Artificial Intelligence: Theories, Models and Applications*, Lecture Notes in Computer Science, Vol. 5138, edited by J. Darzentas, G. Vouros, S. Vosinakis, and A. Arnellos (Springer Berlin Heidelberg, 2008) pp. 365–370.

- [54] H. Bay, T. Tuytelaars, and L. Van Gool, *Surf: Speeded up robust features*, in *Computer Vision – ECCV 2006*, Lecture Notes in Computer Science, Vol. 3951, edited by A. Leonardis, H. Bischof, and A. Pinz (Springer Berlin Heidelberg, 2006) pp. 404–417.
- [55] L. Nalpantidis, G. C. Sirakoulis, and A. Gasteratos, *Non-probabilistic cellular automata-enhanced stereo vision simultaneous localization and mapping*, *Measurement Science and Technology* **22**, 114027 (2011).
- [56] P. Poulakis, L. Joudrier, S. Wailliez, and K. Kapellos, *3drov: A planetary rover system design, simulation and verification tool*, in *10th International Symposium on Artificial Intelligence, Robotics and Automation in Space, i-SAIRAS-08, Hollywood, USA* (2008).
- [57] D. F. Bacon, S. L. Graham, and O. J. Sharp, *Compiler transformations for high-performance computing*, *ACM Comput. Surv.* **26**, 345 (1994).
- [58] S. Cabay and L. Jackson, *A polynomial extrapolation method for finding limits and antilimits of vector sequences*, *SIAM Journal on Numerical Analysis* **13**, 734 (1976).
- [59] ROS, *The robot operating system*, .
- [60] G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, and M. Rodrigalvarez, *Hardware implementation of stereo correspondence algorithm for the exomars mission*, in *Field Programmable Logic and Applications (FPL), 2012 22nd International Conference on* (2012) pp. 667–670.
- [61] I. Kostavelis, L. Nalpantidis, E. Boukas, M. A. Rodrigalvarez, I. Stamoulias, G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, and A. Gasteratos, *Spartan: Developing a vision system for future autonomous space exploration robots*, *Journal of Field Robotics* **31**, 107 (2014).
- [62] A. Pizzagalli, T. Buisson, and R. Beica, *3d technology applications market trends amp; key challenges*, in *Advanced Semiconductor Manufacturing Conference (ASMC), 2014 25th Annual SEMI* (2014) pp. 78–81.
- [63] P. Garrou, M. Koyanagi, and P. Ramm, *Handbook of 3D Integration: Volume 3-3D Process Technology* (John Wiley & Sons, 2014).
- [64] R3LOGIC, *R3logic inc.* (2013).
- [65] N. Selvakkumaran and G. Karypis, *Multiobjective hypergraph-partitioning algorithms for cut and maximum subdomain-degree minimization*, *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* **25**, 504 (2006).
- [66] K. Siozios and D. Soudris, *A tabu-based partitioning and layer assignment algorithm for 3-d fpgas*, *Embedded Systems Letters, IEEE* **3**, 97 (2011).

- [67] I. Savidis, S. M. Alam, A. Jain, S. Pozder, R. E. Jones, and R. Chatterjee, *Electrical modeling and characterization of through-silicon vias (tsvs) for 3-d integrated circuits*, *Microelectronics Journal* **41**, 9 (2010).
- [68] 3d-performance Lancaster University, *Tools for design space exploration of 3-d integrated circuits*, [Http://3d-performance.lancs.ac.uk](http://3d-performance.lancs.ac.uk).
- [69] A. Jantsch, M. Grange, and D. Pamunuwa, *The promises and limitations of 3-d integration*, in *3D Integration for NoC-based SoC Architectures* (Springer, 2011) pp. 27–44.
- [70] S. M. Inc., *Sparc v8 architecture*, ().
- [71] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, 2nd ed. (Kluwer Academic Publishers, Norwell, MA, USA, 1997).
- [72] Gaisler, *The bcc cross-compiler*, ().
- [73] Gaisler, *Tsim simulator*, ().

4

Εργαλεία Σχεδιασμού για Επαναδιαμορφούμενες Πλατφόρμες

Στο παρόν κεφάλαιο παρουσιάζονται τα προτεινόμενα εργαλεία σχεδιασμού για επαναδιαμορφούμενες πλατφόρμες. Αναλύεται η τεχνική χαρακτηρισμού και ταξινόμησης των εφαρμογών βάσει των εγγενών χαρακτηριστικών τους, κατά τη φάση της σύνθεσης. Επίσης περιγράφεται αναλυτικά ο γενετικός αλγόριθμος τοποθέτησης πολυ-κριτηριακής βελτιστοποίησης, ο οποίος αξιοποιεί κατευθύνει την τοποθέτηση προς βέλτιστα σημεία Pareto ανάλογα με την ταξινόμηση της εφαρμογής. Το κεφάλαιο 4 συνδέεται με τη γενικότερη μεθοδολογία της διατριβής που παρουσιάζεται στην Ενότητα 1.3, ως προς τη συνεισφορά του στη κατηγορία “Επαναδιαμορφούμενες Πλατφόρμες - Εργαλεία EDA για FPGAs”.

4.1. Εισαγωγή

Κατά την τελευταία δεκαετία οι επαναδιαμορφούμενες αρχιτεκτονικές, και πιο συγκεκριμένα τα FPGAs, έχουν φέρει επανάσταση στον τρόπο με τον οποίο τα ψηφιακά συστήματα σχεδιάζονται και κατασκευάζονται λόγω του εγγενούς χαρακτηριστικού τους για επαναπρογραμματισμό. Επιπλέον, τα σύγχρονα FPGAs προσφέρουν σχεδόν τόσο αποτελεσματικές εναλλακτικές λύσεις σχεδίασης για συγκεκριμένες εφαρμογές, όσο και τα παραδοσιακά ολοκληρωμένα κυκλώματα ASICs. Για παράδειγμα, η πρόσφατη Xilinx Virtex Ultrascale XCVU440 συσκευή FPGA, η οποία σχεδιάζεται στον τεχνολογικό κόμβο των 20nm, περιέχει σχεδόν μέχρι είκοσι δισεκατομμύρια τρανζίστορ [1].

Ακόμα κι αν οι προαναφερθείσες αρχιτεκτονικές επιλογές πληρούν τις απαιτήσεις των περισσότερων σύγχρονων σχεδιαστών FPGA, ωστόσο περιπλέκουν

ακόμη περισσότερο το έργο που ανατίθεται στα εργαλεία σχεδίασης (CAD) για FPGAs [2]. Παρά το γεγονός ότι είναι πλέον διαθέσιμα εργαλεία σχεδίασης (EDA) που υποστηρίζουν αυτοματοποιημένη υλοποίηση εφαρμογών σε πλατφόρμες υλικού, εντούτοις η εκτέλεση τους εξακολουθεί να είναι χρονοβόρα. Το πρόβλημα αυτό γίνεται περισσότερο εμφανές, λαμβάνοντας υπόψη ότι η πυκνότητα σε λογικούς πόρους, αυξάνεται σταθερά σε ρυθμό που προβλέπεται από το νόμο του Moore. Ως εκ τούτου, τα εργαλεία EDA πρέπει να είναι ικανά να συνθέσουν ολοένα και περισσότερους λογικούς πόρους μέσω δικτύων διασύνδεσης για κάθε νέας γενιάς πλατφόρμα. Ωστόσο, δεδομένης της αυξανόμενης πολυπλοκότητας των εφαρμογών που υλοποιούνται σε FPGAs, αναμένεται ότι μεγάλη βαρύτητα πρέπει να δοθεί στα εργαλεία φυσικού σχεδιασμού, όπως για παράδειγμα στην τοποθέτηση και τη δρομολόγηση (P&R), προκειμένου να προκύψουν ιδιαίτερα βελτιστοποιημένες λύσεις σε δεδομένο χρόνο εκτέλεσης.

Ένας βασικός λόγος για αυτή τη πρόβλεψη είναι το γεγονός ότι η βελτιστοποίηση στη σχεδίαση των επεξεργαστών γενικού σκοπού δεν έχει λάβει υπόψη της την αύξηση πυκνότητας λογικών πόρων των FPGAs ήδη από τα μέσα του 2000, όπως φαίνεται και στην Εικόνα 4.1.

Σχήμα 4.1 Σύγκριση της απόδοσης των επεξεργαστών γενικού σκοπού με τα FPGAs, σε σχέση με τη κλιμάκωση των λογικών πόρων.[3].

Η υλοποίηση εφαρμογών σε FPGA πλατφόρμες μπορεί να διαρκέσει αρκετές ώρες, ή ακόμη και ημέρες, ανάλογα με την πολυπλοκότητα του σχεδιασμού. Ένα από τα πιο χρονοβόρα βήματα στη ροή σχεδιασμού FPGA είναι η τοποθέτηση των λογικών πόρων της εφαρμογής στους φυσικούς διαθέσιμους πόρους της συσκευής FPGA (placement). Η τοποθέτηση καλής ποιότητας είναι απαραίτητη για την συνολική ποιότητα του σχεδιασμού, καθώς επηρεάζει μεταξύ άλλων την

καθυστέρηση της διασύνδεσης, τη συμφόρηση των σημάτων, το συνολικό μήκος καλωδίου, καθώς και την ισχύ κατανάλωσης [4]. Ενώ υπάρχει εκτενής έρευνα σε αλγόριθμους σχετικά με το πρόβλημα της τοποθέτησης για τη βελτιστοποίηση της μέγιστης συχνότητας λειτουργίας, της κατανάλωσης ισχύος/ενέργειας και του χώρου που καταλαμβάνεται από ένα κύκλωμα για την συνολική καλωδίωση, ωστόσο, ελάχιστοι ερευνητές έχουν ως πρωταρχικό τους στόχο την ελαχιστοποίηση του χρόνου εκτέλεσης των εργαλείων σχεδίασης. Πρόσφατα ο χρόνος μεταγλώττισης έχει αναγνωριστεί ως ένα από τα πλέον σημαντικά ζητήματα για τα FPGAs [5], ενώ παράλληλα υπάρχουν σχεδιαστές που είναι πρόθυμοι να ανεχθούν μια σχετικά μικρή μείωση της ποιότητας των αποτελεσμάτων (π.χ. μια μικρότερη συχνότητα λειτουργίας του τελικού σχεδιασμού) προκειμένου να επιτύχουν μεγαλύτερη ταχύτητα στο χρόνο φυσικής σχεδίασης [4]. Επιπλέον, δεδομένου ότι η ικανότητα των συσκευών FPGA και το μέγεθος των σχεδιασμών συνεχώς αυξάνονται, υπάρχει έντονο ενδιαφέρον τόσο στην ερευνητική κοινότητα όσο και στη βιομηχανία επαναδιαμορφούμενων αρχιτεκτονικών προς ταχύτατες μεθόδους φυσικής σχεδίασης.

Προκειμένου να ανταποκριθούν στο πρόβλημα των μεγάλων χρόνων φυσικής σχεδίασης, δύο από τις μεγαλύτερες εταιρείες κατασκευής FPGAs αναφορικά με τη χωρητικότητα που προσφέρουν, οι εταιρείες Xilinx και Altera, συνεχώς βελτιστοποιούν τα ήδη υπάρχοντα εργαλεία CAD. Ακόμα κι αν αυτή η επιλογή αποτελεί μερική λύση στην πίεση του χρόνου εκτέλεσης, είναι σχεδόν απίθανο ότι οι υπάρχοντες αλγόριθμοι σχεδιασμού μπορούν να ακολουθήσουν την κλιμάκωση του ρυθμού ανάπτυξης που επιβάλλει ο νόμος του Moore. Η συνεχής τεχνολογική κλιμάκωση χωρίς την ανάλογη κλιμάκωση του χρόνου σχεδίασης για τη φυσική υλοποίηση των εφαρμογών σε FPGAs αναμένεται να οδηγήσει σταδιακά σε κρίση του χρόνου σχεδίασης. Αυτή η κρίση, μεταξύ άλλων, εκδηλώνεται ως μείωση της παραγωγικότητας και ως εκ τούτου αντίστοιχη αύξηση του κόστους παραγωγής. Με βάση τις πρόσφατες σχετικές ερευνητικές προσεγγίσεις, υπάρχουν τρεις τρόποι προκειμένου να μειωθεί ο χρόνος σχεδίασης με εργαλεία CAD, οι οποίοι μπορούν να ταξινομηθούν ως εξής:

- Μετάβαση από τον ολοκληρωτικό τρόπο σχεδιασμού σε ιεραρχική και τμηματική ανάπτυξη των επί μέρους τμημάτων που απαρτίζουν το συνολικό σχεδιασμό. Παρά το γεγονός ότι αυτή η προσέγγιση επιτυγχάνει εν μέρει τη μείωση του χρόνου εκτέλεσης, εντούτοις αυξάνει τη σχεδιαστική πολυπλοκότητα, ενώ επίσης αποτρέπει τις βελτιστοποιήσεις μεταξύ των επί μέρους σχεδιαστικών τμημάτων.
- Εύρεση ταχύτερων σειριακών και μονονηματικών αλγορίθμων, οι οποίοι μπορούν να επιτύχουν αξιοσημείωτη επιτάχυνση του χρόνου εκτέλεσης, με απώλεια, όμως, στην ποιότητα της υλοποίησης της εφαρμογής [4] [6]. Αυτή η λύση μπορεί να οδηγήσει σε αξιοσημείωτα ποσοστά σχεδιαστικής επιτάχυνσης, ωστόσο, δεν είναι μία κοινώς αποδεκτή λύση, καθώς δεν δύναται να ακολουθήσει την εκθετική αύξηση των λογικών πόρων των σύγχρονων FPGAs.

Ανάπτυξη κανοτόμων παραλλήλων αλγορίθμων, οι οποίοι θα αξιοποιούν

τους υπάρχοντες αλλά και τους μελλοντικούς επεξεργαστές πολλαπλών πυρήνων [7] [8] [9] [10] [11] [12] [13] [14]. Ο τεχνολογικός περιορισμός που παρατηρείται στη συνεχή αύξηση της συχνότητας των επεξεργαστών, οδήγησε την έρευνα και τη βιομηχανία σε σχεδιασμούς επεξεργαστών πολλαπλών πυρήνων, συνεπεία δε τούτου, η χρήση παραλλήλων CAD αλγορίθμων αποτελεί μία ελκυστική λύση στο πρόβλημα του μεγάλου χρόνου σχεδίασης. Αυτοί οι αλγόριθμοι επιτρέπουν την ταυτόχρονη κλιμάκωση τόσο του μεγέθους των FPGA συσκευών όσο και του αριθμού των εν λειτουργία επεξεργαστικών πυρήνων. Προς αυτή την κατεύθυνση, οι εταιρείες Xilinx και Altera έχουν αρχίσει να υλοποιούν παράλληλες εκδόσεις των σχεδιαστικών τους εργαλείων, οι οποίες προσφέρουν αξιόλογη επιτάχυνση του χρόνου εκτέλεσης και ακριβώς η προσέγγιση αυτή ερευνάται στον παρόν κεφάλαιο.

4

Η λειτουργικότητα της πλειοψηφίας των διαθέσιμων αλγορίθμων για το πρόβλημα της τοποθέτησης βασίζεται στην τεχνική της προσομοιωμένης ανόπτησης. [15] [7] [8] [12] [7] [11] [8]. Σε αυτές τις προσεγγίσεις, η εύρεση της βέλτιστης ποιότητας τοποθέτησης, εδράζεται σε ακολουθιακές ανταλλαγές τυχαίων θέσεων των λογικών δομών, αποδεχόμενες ολόένα και εγγύτερες μετακινήσεις δυνάμει της εκθετικά μειούμενης πιθανότητας της τρέχουσας θερμοκρασίας και μιας μεταβλητής δ της συνολικής συνάρτησης κόστους. [16]. Η θερμοκρασία στην προσομοιωμένη ανόπτηση μειώνεται μετά από μία διαδικασία ψύξης. Στην αρχή αυτής της διαδικασίας, σχεδόν όλες οι μετακινήσεις είναι επιτρεπτές. Κατά τη διαδικασία ψύξης, μόνο οι μετακινήσεις που θα βελτιώσουν το συνολικό συναρτησιακό κόστος θα γίνουν εν τέλει αποδεκτές. Προηγούμενες εργασίες (λ.χ [15] [17]) έχουν δείξει ότι οι αλγόριθμοι τοποθέτησης που βασίζονται σε προσομοιωμένη ανόπτηση παράγουν ενθαρρυντικά αποτελέσματα υπό την προϋπόθεση ότι έχουν επιλεγεί οι κατάλληλες συναρτήσεις κόστους αλλά και αρκούντως αργές διαδικασίες ψύξης.

Στο κεφάλαιο τούτο, παρουσιάζουμε ένα καινοτόμο αλγόριθμο για τη λύση του προβλήματος της τοποθέτησης, θεμελιωμένο στους γενετικούς αλγορίθμους. Οι βασικές συνεισφορές της εργασίας που παραδίδεται σε αυτό το κεφάλαιο μπορεί να συνοψιστεί ως ακολούθως:

- Προτείνεται ένας νέος παράλληλος γενετικός αλγόριθμος (GA) για τη λύση του προβλήματος της τοποθέτησης εφαρμογών σε συσκευές FPGA. Ο προτεινόμενος αλγόριθμος χρησιμοποιεί την έννοια της συμβατότητας, προκειμένου να βελτιωθεί η ποιότητα των λύσεων τοποθέτησης.
- Αναλύονται οι εσωτερικές δομές (επιλογή με βάση τον ελιτισμό, συμβατότητα με γνώμονα τη μετάλλαξη, διαδρομή με γνώμονα, αυτοπροσαρμοζόμενη μετάλλαξη), οι οποίες ορίζονται προς το σκοπό της εφαρμογής των γενετικών αλγορίθμων για το πρόβλημα της τοποθέτησης.
- Παρουσιάζεται μία αδρή, πλήν όμως επαρκής, παράλληλη υλοποίηση OpenMP του προτεινόμενου γενετικού αλγορίθμου, με τον εντοπισμό ανεξάρτητων

εργασιών στα στάδια του γενετικού αλγορίθμου και με την υλοποίηση δομών μνήμης προσαρμοσμένων σε πολυνηματικό περιβάλλον επεξεργασίας.

- Προτείνεται μία μεθοδολογία διερεύνησης λογισμικού για την παραμετροποίηση του γενετικού αλγορίθμου, συμφώνως προς τα χαρακτηριστικά της εφαρμογής εισόδου.
- Προβλέπεται μία εκτενής πειραματική αξιολόγηση σχετικά με την αποδοτικότητα και την επεκτασιμότητα του προτεινόμενου πλαισίου, εν συγκρίσει προς τα πλέον ερευνητικά εξελιγμένα εργαλεία για το πρόβλημα της τοποθέτησης.

Ειδικότερα, ο προτεινόμενος αλγόριθμος τοποθέτησης εκτελεί αποτελεσματικότερα διερεύνηση του χώρου λύσεων, ενώ ο ενδογενής παραλληλισμός του αξιοποιείται από τους επεξεργαστές πολλαπλών πυρήνων, οδηγώντας με αυτόν τον τρόπο σε μείωση του σχεδιαστικού χρόνου εκτέλεσης.

Η προτεινόμενη προσέγγιση υποστηρίζει είτε ταχύτατη διαδικασία τοποθέτησης της εφαρμογής (με σημαντικά χαμηλότερο χρόνο σχεδιαστικής εκτέλεσης), ή παρέχει ανώτερης ποιότητας τοποθέτηση (π.χ. επιτυγχάνοντας υψηλότερη συχνότητα λειτουργίας, μείωση της κατανάλωσης ενέργειας, κλπ) σε βάρος του σχεδιαστικού χρόνου εκτέλεσης. Συγκεκριμένα, στη πρώτη περίπτωση, η λύση μας οδηγεί σε 67x ταχύτερη εκτέλεση κατά μέσο όρο, ενώ η μέση αύξηση της μέγιστης συχνότητας λειτουργίας, στη δεύτερη προσέγγιση, είναι 16%. Ως σημείο αναφοράς για να την μελέτη μας, χρησιμοποιούμε το υφιστάμενο ποιοτικότερο ακαδημαϊκό εργαλείο για το πρόβλημα της τοποθέτησης FPGAs, VPR[15] [16], καθώς και την ερευνητική προσέγγιση που περιγράφεται σε προηγούμενη εργασία μας [18].

Το κεφάλαιο οργανώνεται ως εξής: η ενότητα 2 περιγράφει τη σχετική βιβλιογραφία σε αλγορίθμους τοποθέτησης, ενώ η ενότητα 3 εισάγει το προτεινόμενο μεθοδολογικό πλαίσιο για την υλοποίησης εφαρμογών σε επαναδιαμορφούμενες αρχιτεκτονικές με τη χρήση γενετικών αλγορίθμων. Δεδομένου ότι η αποτελεσματικότητα του αλγορίθμου εξαρτάται σε μεγάλο βαθμό από την προσαρμογή των παραμέτρων του στις ιδιότητες της εφαρμογής, η ενότητα 4 περιγράφει τη μεθοδολογία που προτείνουμε για τη ρύθμιση των παραμέτρων του εξελικτικού αλγόριθμου τοποθέτησης. Ποσοτική αξιολόγηση που αποδεικνύει την αποτελεσματικότητα του προτεινόμενου πλαισίου σε σχέση με τις υπάρχουσες ερευνητικές προσεγγίσεις παρουσιάζεται στην ενότητα 5, ενώ τα συμπεράσματα συνοψίζονται στην ενότητα 6.

4.2. Συναφής βιβλιογραφία

Την τελευταία δεκαετία, οι ερευνητές έχουν αναπτύξει μεθοδολογίες, αλγορίθμους και εργαλεία προς το σκοπό της επιτάχυνσης της διαδικασίας τοποθέτησης των λογικών πυλών, που έχουν προκύψει από τη διαδικασία της σύνθεσης, σε φυσικές πύλες στα FPGAs, με τη χρήση παραλλήλων αλγορίθμων.

Ο πίνακας 4.1 συνοψίζει ένα αντιπροσωπευτικό πλήθος των εν λόγω προσεγγίσεων. Η δεύτερη στήλη του πίνακα εξετάζει εάν οι επιλογές τους κατά την διαδικασία της τοποθέτησης εκτελούνται με πρωτεύοντα στόχο τη βελτίωση της μέγιστης συχνότητας λειτουργίας (Timing-Driven). Εν συνεχεία, οι δύο ακραίες δεξιά (τελευταίες) στήλες απεικονίζουν την υπολογιστική πλατφόρμα και την επιτευχθείσα επιτάχυνση αντιστοίχως. Ακολούθως αναλύονται με περισσότερες λεπτομέρειες οι περισσότερο διαδεδομένες ερευνητικά λύσεις στη σχετική βιβλιογραφία των αλγορίθμων τοποθέτησης για FPGAs.

Πίνακας 4.1 Υφιστάμενες λύσεις παράλληλων αλγορίθμων τοποθέτησης.

Αν/ρά	Timing-Driven	Υπολογιστική Πλατφόρμα	Επιτάχυνση
[7]	No	Sequent Balance 8,000 (8-proc.)	6.4× on 8 processors
[11]	No	VAX 11/784 (4-processors)	2.3× on 4 processors
[9]	No	6 Nat.Sem. 32,016 processors	4× with 5 processors
[13]	No	Hypercube multiprocessors	8× with 16 processors
[19]	No	Networks of machines	3.3× on 16 processors
[12]	No	FPGAs	500×-2,500× over CPUs
[10]	Yes	Multiprocessors	2.1× on 4 processors
[14]	No	MPPAs	1/256 less swaps - 1,024 cores
[8]	No	GPU	10× on NVIDIA GTX280
[20]	Yes	Multiprocessors	2.4× on 8 processors
[3]	Yes	Multiprocessors	161× using 25 processors

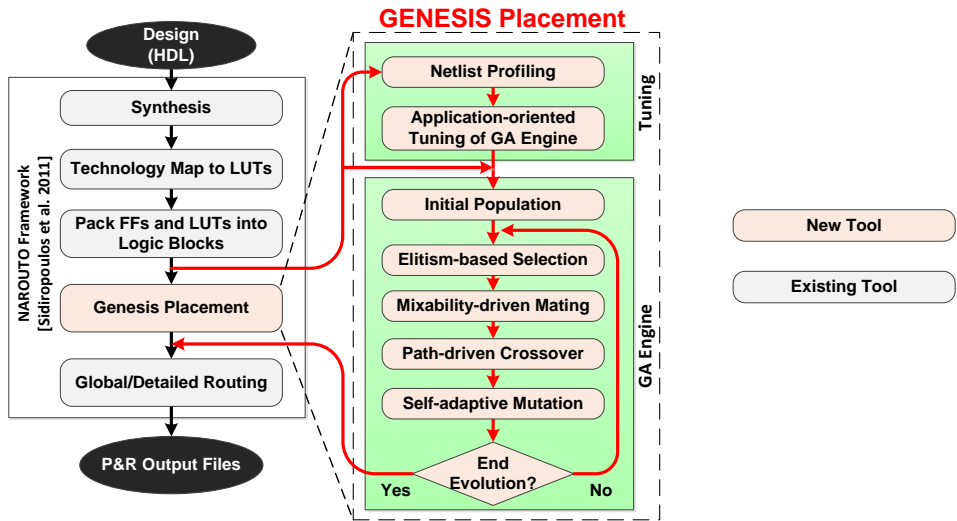
Στις εργασίες [10], [20] προτείνεται μία παράλληλη υλοποίηση του αλγόριθμου προσημειωμένης ανόπτωσης με κριτήριο τη βελτιστοποίηση της συχνότητας λειτουργίας, για το περιβάλλον ανάπτυξης Quartus II [21]. Παρά το γεγονός ότι η εν λόγω προσέγγιση αξιολογεί παράλληλα ένα σύνολο εναλλακτικών κινήσεων, οι κινήσεις αυτές περιορίζονται προκειμένου ο αλγόριθμος να επιτύχει μία ακολουθιακά ισότιμη τοποθέτηση. Οι ερευνητές υποστηρίζουν ότι αυτή η προσέγγιση επιτυγχάνει επιτάχυνση μεταξύ 2,1× και 2.4×, για 4 και 8 πυρήνες επεξεργασίας, αντίστοιχα, ενώ η ποιότητα των τοποθετήσεων είναι συγκρίσιμη προς την αντίστοιχη ακολουθιακή έκδοση. Στην εργασία [12] παρουσιάζεται ένας καταναεμημένος αλγόριθμος προσομοιωμένης ανόπτωσης, ο οποίος επιταχύνεται από μια συστολική αρχιτεκτονική. Η λειτουργικότητα αυτού του αλγορίθμου βασίζεται στον περιορισμό του εύρους ανταλλαγής του κάθε μπλοκ στα τέσσερα εγγύτερα αυτού. Ακόμα κι αν μια τέτοια επιλογή εύκολα παραλληλιζείται ευχερώς, η έλλειψη της δυνατότητας της να εφαρμόσει μια πιο δυναμική και ευρεία λύση για το πρόβλημα της τοποθέτησης, οδηγεί σε 36% υποβάθμιση της ποιότητας στο τελικό κύκλωμα. Μια επέκταση αυτής της εργασίας που παρουσιάζεται στην εργασία [12], έδειξε ότι υπάρχει δυνατότητα βελτίωσης της ποιότητας έως 5% σε σύγκριση με την προσέγγιση της προσημειωμένης ανόπτωσης, ενώ εξακολουθεί να προσφέρει αξιόλογους παράγοντες επιτάχυνσης. Στην εργασία [22] παρουσιάζεται μια μέθοδος ομαδοποίησης της λίστας πυλών (netlist), με στόχο να μειώσει το μέγεθος του προβλήματος της τοποθέτησης, έτσι ώστε να είναι ευκολότερα επιλύσιμο. Μια παρόμοια προσέγγιση ακολουθείται στην εργασία [23], όπου ένας min-cut αλγόριθμος χωρίζει την τοποθέτηση του συνολικού netlist σε μικρότερα προβλήματα τοποθέτησης, έτσι ώστε αυτά

να μπορούν να επιλυθούν παράλληλα. Ωστόσο, η ποιότητα τοποθετήσεων που παράγει αυτή η προσέγγιση είναι υποβαθμισμένη. Έτερη προσέγγιση βασιζόμενη σε έναν αλγόριθμο χωροθέτησης τμημάτων (macro-floorplanner) με στόχο βελτιστοποίησης τη συχνότητα επεξεργασίας, περιγράφεται στην εργασία [5]. Ακόμα κι αν ο συγγραφέας αναφέρει μια βελτίωση στην εκτέλεση του χρόνου εκτέλεσης σε σχέση με τις σχετικές προσεγγίσεις, τα προβλήματα επαλήθευσης που χρησιμοποιούνται είναι ιδιαίτερα μικρά σε μέγεθος. Ως εκ τούτου, τα παραγόμενα αποτελέσματα είναι αμφίβολης σημασίας αναφορικά προς τα μεγάλα προβλήματα του χώρου των σύγχρονων FPGAs. Οι συγγραφείς της εργασίας [24] προτείνουν έναν αλγόριθμο τοποθέτησης, ο οποίος παρέχει λύσεις που συμβιβάζουν την ποιότητα, υπό την έννοια του συνολικού μήκους καλωδίου και του χρόνου εκτέλεσης. Προσεγγίζεται ακόμα και μία λύση με βάση τον αλγόριθμο Tabu, προκειμένου να επιταχύνει το χρόνο εκτέλεσης της τοποθέτησης [6]. Οι συγγραφείς της εργασίας [3] παρουσίασαν ένα παράλληλο αλγόριθμο τοποθέτησης με κριτήριο βελτιστοποίησης τη συχνότητα λειτουργίας, ο οποίος επιτυγχάνει υψηλή και κλιμακούμενη επιτάχυνση σε σύγκριση με το ακαδημαϊκό εργαλείο VPR [15]. Ωστόσο τα κέρδη αυτά αντισταθμίζονται από τις χαμηλές επιδόσεις στην ποιότητα, τόσο της συχνότητας λειτουργίας όσο και του μήκους καλωδίου. Μια εναλλακτική προσέγγιση προϋποθέτει την ενσωμάτωση εξειδικευμένου υλικού για την επιτάχυνση της προσομοιωμένης ανόπτησης [12]. Η βασική ιδέα αυτής της εργασίας είναι η εισαγωγή παραλληλισμού στον αλγόριθμο, ωστόσο το γεγονός ότι απαιτεί ιδιαίτερα εξειδικευμένο υλικό, καθιστά ανεφάρμοστη πρακτικά αυτή την υλοποίηση. Τέλος, οι συγγραφείς της εργασίας [25] προτείνουν τη χρήση γενετικού αλγορίθμου στο εργαλείο VPR. Ωστόσο, τα πειραματικά αποτελέσματα που παρέχονται σε αυτή την εργασία δεν απέφεραν ποιοτικότερες λύσεις τοποθετήσεων σε σύγκριση με συμβατικούς αλγορίθμους, όπως η προσομοιωμένη ανόπτηση.

4.3. Το προτεινόμενο πλαίσιο σχεδίασης

Το πλήρες προτεινόμενο πλαίσιο σχεδίασης για τη ροή σύνθεσης σε επαναδιαμορφούμενες πλατφόρμες απεικονίζεται στην Εικόνα 4.2. Το πλαίσιο είναι παρόμοιο με υπάρχουσες ακαδημαϊκές και βιομηχανικές ροές σχεδίασης όπως οι [15] και [21], λαμβάνοντας υπόψη ότι η εγγενής αρθρωτή του ανάπτυξη του επιτρέπει να προσαρμόζει εύκολα εργαλεία με την ίδια λειτουργικότητα μεταξύ εναλλακτικών ροών. Πιο συγκεκριμένα, αυτή η εργασία επικεντρώνεται στο πρόβλημα της τοποθέτησης. Για αυτό το σκοπό η γενικότερη ροή σχεδίασης στηρίζεται στην ερευνητική εργασία [26] και συγκεκριμένα στο πλαίσιο *NAROUTO*, στο οποίο αντικαταστάθηκε ο υφιστάμενος αλγόριθμος τοποθέτησης με τον προτεινόμενο αλγόριθμο *GENESIS*.

Η σύνθεση είναι το πρώτο στάδιο της ροής εκτέλεσης του προτεινόμενου CAD εργαλείου. Κατά τη διάρκεια αυτού του σταδίου, η περιγραφή της εφαρμογής σε γλώσσα περιγραφής υλικού HDL (Hardware Description Language) μετατρέπεται σε μία λίστα λογικών συναρτήσεων σε επίπεδο πύλης (gate-level netlist) και flip-flops. Το επόμενο βήμα περιλαμβάνει την τεχνολογική απεικόνιση με σκοπό



Σχήμα 4.2 Προτεινόμενο μεθοδολογικό πλαίσιο για παράλληλη τοποθέτηση σε FPGAs.

τη μετάφραση της netlist που προέρχεται από το βήμα της σύνθεσης, σε μια netlist που αποτελείται από LUTs (Look-up Tables) K-εισόδων και flip-flops. Σε αυτή τη κατεύθυνση υπάρχουν δημοσιευμένοι αλγόριθμοι που βελτιστοποιούν συγκεκριμένες μετρικές συστήματος, όπως η επιφάνεια πυριτίου, η χρονική καθυστέρηση και η καταναλισκόμενη ενέργεια, είτε μεμονωμένα είτε ακόμα και ταυτόχρονα με συνδυασμό αυτών. Για αυτό το στάδιο της απεικόνισης, στα πλαίσια αυτής της έρευνας, επιλέγεται το διαδεδομένο ερευνητικό εργαλείο SIS [27].

Το επόμενο στάδιο περιλαμβάνει την ομαδοποίηση των LUTs και flip-flops, τα οποία προέκυψαν από τη διαδικασία της τεχνολογικής απεικόνισης, σε συστάδες (clusters), τηρώντας τον περιορισμό ότι κάθε λογική συστάδα δεν μπορεί να υπερβαίνει τα N-LUTs, ή N-flip-flops. Η έξοδος αυτής της διεργασίας ομαδοποίησης είναι ένα σύνολο από βασικές λογικές μονάδες (Basic Logic Elements). Αυτές οι λογικές μονάδες είναι δομικά παρόμοιες στο σύνολο των διαφορετικών κατασκευαστών FPGAs, ωστόσο κάθε ένας κατασκευαστής έχει υιοθετήσει τη δική του ορολογία για να περιγράψουν αυτή τη δομή. Για παράδειγμα, η εταιρεία Xilinx χρησιμοποιεί τον όρο CLB (Configurable Logic Block), ενώ η εταιρεία Altera χρησιμοποιεί τον όρο LAB (Logic Array Block). Κατά τη διαδικασία της ομαδοποίησης μπορούν να χρησιμοποιηθούν διαφορετικά κριτήρια βελτιστοποίησης.

Κατά την εργασία ομαδοποίησης, μπορούν να ληφθούν υπόψιν διαφορετικοί στόχοι βελτιστοποίησης. Τέτοια τυπικά παραδείγματα περιλαμβάνουν την ελαχιστοποίηση της καθυστέρησης [16] και της κατανάλωσης ενέργειας [28]. Οι εν λόγω βελτιστοποιήσεις βασίζονται κυρίως στην εσωτερική διασύνδεση των BLEs μεταξύ της ίδιας λογικής συστάδας, η οποία είναι ταχύτερη σε σχέση με τις αντίστοιχες συνδέσεις μεταξύ διαφορετικών λογικών συστάδων. Αμφότερες

οι συνδέσεις αποτελούν το δίκτυο δρομολόγησης του σχεδιασμού. Ως εκ τούτου, το μέγεθος της λογικής συστάδας N επηρεάζει σε μεγάλο βαθμό την ποιότητα της υλοποίησης. Επί παραδείγματι, μεγαλύτερη τιμή του N επιτρέπει σε περισσότερες λογικές συστάδες να μοιραστούν εσωτερικές συνδέσεις. Ωστόσο, αναλόγως προς την εφαρμογή, απαιτείται η προσεκτική ανάλυση συμβιβασμού, καθώς η καθυστέρηση στις εσωτερικές συνδέσεις των λογικών συστάδων αυξάνεται με το μέγεθος της λογικής συστάδας N . Για τους σκοπούς της παρούσας εργασίας, τίθεται η υπόθεση ότι η πλατφόρμα υλοποίησης FPGA αποτελείται από 4 BLEs ανά λογική συστάδα (δηλαδή $N=4$), ενώ παράλληλα, η ομαδοποίηση της netlist υλοποιείται με το εργαλείο Versatile Packing (T-VPACK) [15].

Το στάδιο της τοποθέτησης, περιλαμβάνει την εκχώρηση έκαστου CLB της netlist σε μια μοναδική φυσική θέση στο FPGA. Σε αυτήν την εργασία, γίνεται χρήση του γενετικού αλγορίθμου *GENESIS*. Εν αντιθέσει προς τις υφιστάμενους αλγορίθμους τοποθέτησης, ο προτεινόμενος αλγόριθμος χαρακτηρίζεται από εγγενή δυνατότητα παραλληλοποίησης και συνεπώς μπορεί επωφεληθεί από τους σύγχρονους πολυπύρηνους επεξεργαστές. Επιπλέον, η προτεινόμενη λύση υποστηρίζει αυτόματη ρύθμιση των παραμέτρων του αλγορίθμου, ανάλογα προς τα εγγενή χαρακτηριστικά της εφαρμογής, με σκοπό την περαιτέρω βελτίωση των επιδόσεων. Ο *GENESIS* αλγόριθμος παρέχει λύσης τοποθέτησης με στόχο την ελαχιστοποίηση αφενός της ημιπεριμέτρου του μήκους καλωδίου Half-perimeter Wirelength (HPWL), μια μετρική γνωστή ως κόστος οριοθέτησης και αφετέρου την καθυστέρηση διάδοσης των σημάτων της εφαρμογής και εκ τρίτου της κατανάλωσης ενέργειας. Οι αποφάσεις που λαμβάνονται από τον αλγόριθμο τοποθέτησης επηρεάζουν σημαντικά τη συνολική ποιότητα της λύσης τοποθέτησης, καθώς οι θέσεις των CLBs μένουν σταθερές για τις υπόλοιπες διεργασίες της ροής σχεδιασμού και επίσης κατά το χρόνο εκτέλεσης της εφαρμογής. Στις επόμενες ενότητες θα παρουσιαστούν πρόσθετες πληροφορίες σχετικά με τον προτεινόμενο αλγόριθμο.

Η δρομολόγηση είναι το τελευταίο στάδιο της προτεινόμενης ροής σχεδίασης, κατά τη διάρκεια της οποίας δρομολογούνται οι κατάλληλες διασυνδέσεις όλων των σημάτων του σχεδιασμού, με τη χρήση των προγραμματιζόμενων διακοπών δρομολόγησης του FPGA. Στην παρούσα εργασία, χρησιμοποιείται το εργαλείο δρομολόγησης του VPR, το οποίο βασίζεται στον αλγόριθμο PathFinder [16]. Τέλος, μετά την επιτυχή τοποθέτηση και δρομολόγηση της εφαρμογής στο FPGA, ακολουθεί το στάδιο της χρονικής επαλήθευσης και της αξιολόγησης της ποιότητας υλοποίησης, που περιλαμβάνει την εξαγωγή των μετρικών της μέγιστης επιτευχθείσας συχνότητας επεξεργασίας και της κατανάλωσης ισχύος.

4.3.1. Ο εξελικτικός αλγόριθμος δρομολόγησης GENESIS

Στην παρούσα ενότητα περιγράφεται λεπτομερώς ο εξελικτικός αλγόριθμος *GENESIS*. Η διαδικασία της τοποθέτησης οδηγεί στην φυσική εκχώρηση όλων των λογικών δομών και δομών εισόδου - εξόδου (I/O pads) μέσω της βελτιστοποίησης μίας ή περισσότερων μετρικών (π.χ. μήκος καλωδίου, ταχύτητα και κατανάλωση ισχύος). Δεδομένου ότι η διαδικασία της τοποθέτησης ανήκει στην κλάση των NP-hard προβλημάτων βελτιστοποίησης συνδυαστικής λογικής, δεν

μπορεί να υπάρξει αλγόριθμος που επιλύει βέλτιστα το πρόβλημα σε πολυωνυμικό χρόνο. Συνεπώς, πολλοί ευρετικοί κανόνες έχουν προταθεί, σε μια προσπάθεια επίτευξης των θεωρητικά βέλτιστων λύσεων σε εύλογο χρονικό διάστημα. Για το λόγο τούτο, υιοθετούνται οι γενετικοί αλγόριθμοι (GA), οι οποίοι έχουν ήδη εφαρμοστεί σε πλήθος προβλημάτων βελτιστοποίησης με αξιόλογες επιδόσεις. Οι γενετικοί αλγόριθμοι ενσωματώνουν μια τυχαία, αλλά κατευθυνόμενη αναζήτηση για τον εντοπισμό της ολικά βέλτιστης λύσης. Δεδομένου, ότι η αναζήτηση δεν προκαταλαμβάνεται έναντι της τοπικά βέλτιστης λύσης, οι εν λόγω αλγόριθμοι υπερέχουν ως προς τις συμβατικές βαθμωτές μεθόδους σύγκλισης. Σε αντίθεση προς τους αλγόριθμους, οι οποίοι επιχειρούν να ανεύρουν μία επαρκή λύση μέσω τυχαίων επιλογών, ο γενετικός αλγόριθμος μπορεί να κατευθύνει την εν λόγω έρευνα προς συγκεκριμένες περιοχές του χώρου αναζήτησης [29], οδηγώντας μεταξύ άλλων σε σημαντικά μειωμένο χρόνο εκτέλεσης, χωρίς να επηρεάζει σημαντικά την ποιότητα των παράγωγων λύσεων. Επιπλέον, όπως επιδεικνύεται στην ενότητα 4.3.2 ο γενετικός αλγόριθμος δύναται να παραλληλοποιηθεί αποτελεσματικά, έτσι ώστε να επωφεληθεί των σύγχρονων πολυπύρηνων επεξεργαστών και με τον τρόπο αυτό να βελτιωθεί η απόδοση σε σχέση με το εργαλείο VPR 4.30 [15].

Οι γενετικοί αλγόριθμοι εμπνέονται από τη θεωρία της εξέλιξης και έχουν σχεδιαστεί ως γενικές στρατηγικές αναζήτησης και μέθοδοι βελτιστοποίησης που εργάζονται επί πλήθους εφικτών λύσεων. Ένας γενετικός αλγόριθμος στοχεύει στην εύρεση της κατά το δυνατόν βέλτιστης λύσης μέσω της αναπαράστασης των υποψηφίων λύσεων με αλφαριθμητικά (strings)¹, και εκτελεί μία ακολουθία από μετασχηματισμούς, διεπόμενοι από ένα σύστημα επιλογής, το οποίο τίθεται υπέρ των λύσεων υψηλής ποιότητας. Κάθε χρωμόσωμα χαρακτηρίζεται από μία μετρική ποιότητας (fitness value) σύμφωνα με μια καθορισμένη συνάρτηση καταλληλότητας (fitness function). Συνήθως, η μετρική ποιότητας είναι η τιμή της αντικειμενικής συνάρτησης ή τυχόν κλιμακούμενης έκδοσής της. Κάθε νέα γενιά λύσεων/χρωμοσωμάτων εξελίσσεται με μία τεχνική επιλογής, κατά την οποία οι λύσεις με καλύτερη μετρική ποιότητας συγκεντρώνουν μεγαλύτερες πιθανότητες στην επιβίωση. Τα επιλεγμένα χρωμοσώματα χρησιμοποιούνται ως γονείς της επόμενης γενιάς, συνδυαζόμενα από κατάλληλους τελεστές αναπαραγωγής (reproduction operators).

Η λύση μας υιοθετεί την θεωρία της αναμιξιμότητας, σχετικά με το ρόλο της συνεύρεσης των δύο φύλων στην εξέλιξη, όπως προτείνεται στο [30]. Συμφώνως προς τη θεωρία αυτή, η εξελισσιμότητα² ενός πληθυσμού μπορεί να ενισχυθεί από την ικανότητα των αλληλόμορφων³ να αποδίδουν καλά χαρακτηριστικά στο χρωμόσωμα μέσω διαφορετικών συνδυασμών.

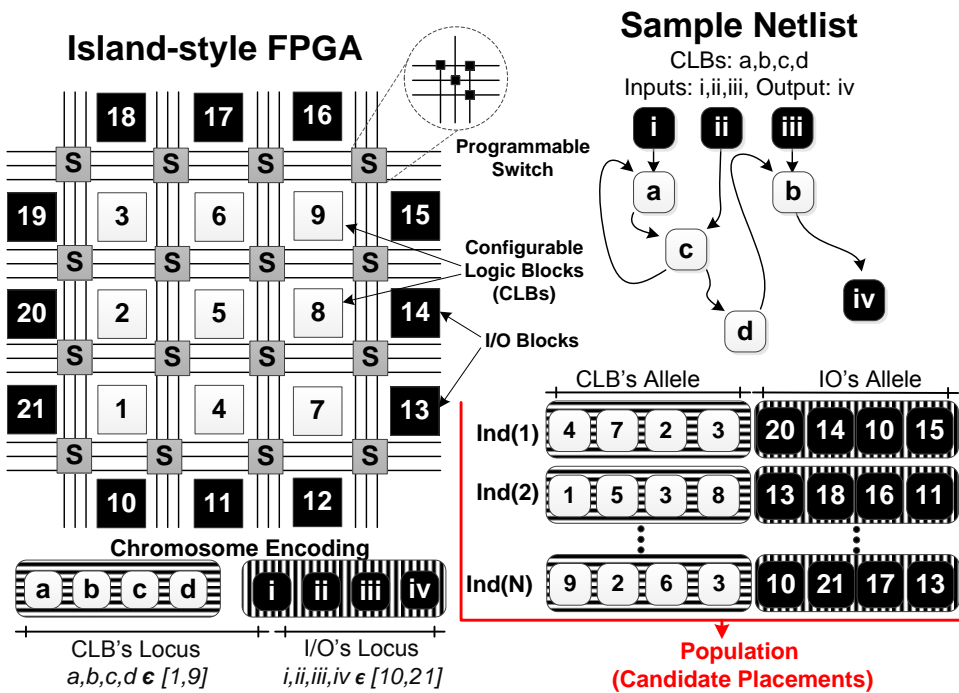
Οι συγγραφείς διερευνούν πώς είναι δυνατόν η συνεύρεση των δύο φύλων

¹Κάθε αλφαριθμητικό αντιστοιχεί σε μία υποψήφια λύση και αναπαρίσταται από ένα χρωμόσωμα.

²Ο όρος εξελισσιμότητα, αναφέρεται στην προσαρμογή των γενετικών τελεστών ενός πληθυσμού με την πάροδο του χρόνου. Πρόκειται περί της ικανότητας ενός πληθυσμού να παράγει προσαρμοστική γενετική ποικιλότητα και ως εκ τούτου να εξελιχθεί μέσω της φυσικής επιλογής.

³Κάθε αλληλόμορφο αντιπροσωπεύει ένα συγκεκριμένο σύνολο εναλλακτικών τιμών για κάθε γονίδιο. Κάθε χρωμόσωμα αποτελείται από ένα μεγάλο αριθμό γονιδίων, καθένα εκ των οποίων βρίσκεται σε μία μοναδική τοποθεσία επί του χρωμοσώματος.

να αυξάνει το μέσο όρο της μετρικής ποιότητας στο σύνολο του πληθυσμού, ενώ, ταυτόχρονα να διασπά εξαιρετικής ποιότητας συνδυασμούς γονιδίων, γεγονός που εμποδίζει την αύξηση της μετρικής ποιότητας. Όπως αποδεικνύεται, η ιδανική λύση είναι επίσης και ένας καλός συνδυασμός υπό την έννοια ότι εξελίσσεται ως σύνολο ανασυνδυασμών λύσεων της προηγούμενης γενιάς και επομένως είναι περισσότερο πιθανό να διατηρηθούν μεμονωμένα καλά γονίδια για συγκεκριμένα χαρακτηριστικά (“the ideal module is also a *good mixer* in the sense that it is transferred as a whole by recombination and is therefore more likely to maintain its individual contribution to fitness across different backgrounds”) [30]. Για τους σκοπούς της παρούσας μελέτης, παρουσιάζεται η εφαρμογή της θεωρίας της αναμιξιμότητας στην κατασκευή των τελεστών *crossover* (υποενότητα 4.3.1.3) and *mating* (υποενότητα 4.3.1.2) του αλγορίθμου GENESIS.



Σχήμα 4.3 Η προτεινόμενη αναπαράσταση λύσεων τοποθέτησης με τη χρήση χρωμοσωμάτων και ένα παράδειγμα κωδικοποίησης netlist με αλφαριθμητικά αλληλόμορφα.

Το παρακάτω παράδειγμα παρουσιάζει το βασικό κίνητρο της προσέγγισης που ακολουθείται σε αυτή την εργασία. Οι αρχιτεκτονικές παράμετροι του χώρου αναζήτησης του γενετικού αλγορίθμου κωδικοποιούνται με τη μορφή μιας δομής που μοιάζει με χρωμόσωμα, όπως απεικονίζεται στο Σχήμα 4.3. Πιο συγκεκριμένα, κάθε χρωμόσωμα αποτελείται από δύο μέρη για την κωδικοποίηση των λογικών δομών (CLBs) και δομών εσόδου/εξόδου (I/Os) της αρχιτεκτονικής, στα οποία αποδίδεται το όνομά ο αλληλότοπος *CLB* και αλληλότοπος

IO, αντίστοιχα, ενώ μία λύση ανάθεσης των CLB_s και I/Os σε ένα αλληλότοπο ορίζεται ως αλληλόμορφο. Οι φυσικές θέσεις των CLB_s και IOs στο FPGA ταξινομούνται με ακολουθίες αύξουσας σειράς αρίθμησης, όπως απεικονίζεται στο κάτω-αριστερό μέρος του Σχήματος 4.3.

Κάθε ξεχωριστή λύση του γενετικού αλγορίθμου αντιπροσωπεύει μία μοναδική λύση τοποθέτησης και εκπροσωπείται με ένα άτομο ($Ind(1), Ind(2), \dots, Ind(N)$), ενώ μια ομάδα ατόμων αποτελεί έναν πληθυσμό. Για να ποσοτικοποιηθεί η ποιότητα αυτών των ατόμων έχει ενσωματωθεί μια συνάρτηση καταλληλότητας. Για τους στόχους της παρούσας έρευνας, ο γενετικός αλγόριθμος είναι κατάλληλα σχεδιασμένος ώστε να αναζητά τοποθετήσεις που βελτιστοποιούν το κόστος χρόνου και καλωδίωσης. Το κόστος χρόνου σε αυτό το σημείο αναφέρεται στην συνολική καθυστέρηση των επί μέρους λογικών μονοπατιών του σχεδιασμού, λόγω των παρασιτικών χωρητικοτήτων του υλικού κατασκευής των FPGAs. Αντίστοιχα το κόστος καλωδίωσης αναφέρεται στο συνολικό μήκος καλωδίου που απαιτείται προκειμένου να δρομολογηθεί πλήρως (detail routing) ένας σχεδιασμός στο FPGA. Η εξίσωση 4.1 δίνει την τιμή καταλληλότητας για κάθε άτομο i του πληθυσμού.

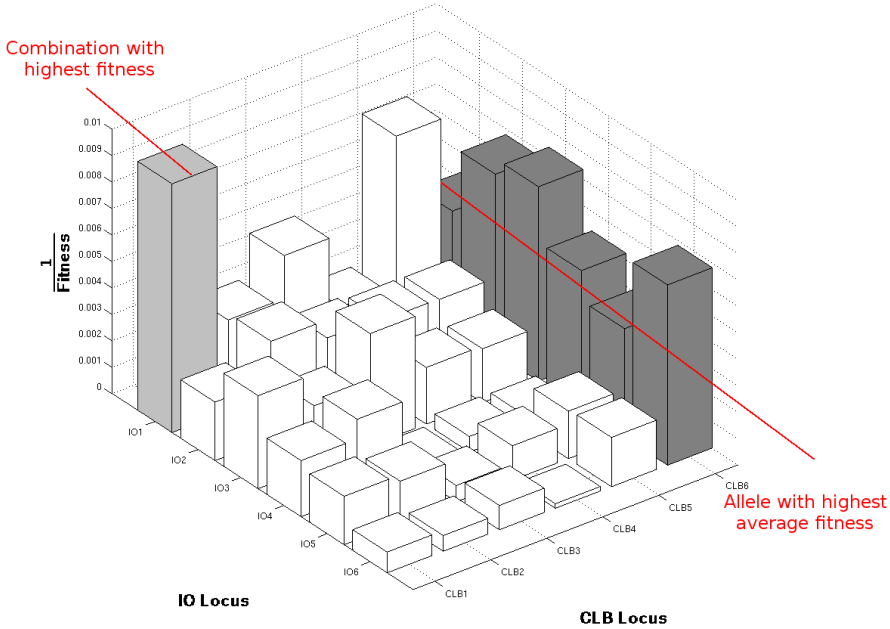
$$Fitness(i) = a \times T_{cost}(i) + (1 - a) \times W_{cost}(i) \quad (4.1)$$

όπου η μετρική T_{cost} υποδηλώνει το άθροισμα των καθυστερήσεων όλων των μονοπατιών της netlist και η μετρική W_{cost} αντιστοιχεί στο συνολικό μήκος καλωδίου όλων αυτών των μονοπατιών.

Όσον αφορά το T_{cost} , υπολογίζεται με βάση το μοντέλο καθυστέρησης Elmore, ενώ το W_{cost} υπολογίζεται από το μοντέλο ημιπεριμέτρου (HPWL) [15] [16]. Η παράμετρος a χρησιμοποιείται για τη στάθμιση της βελτιστοποίησης καθενός από τις δύο προαναφερθείσες μετρικές. Δεδομένου ότι η παράμετρος a επηρεάζει την ποιότητα των αποτελεσμάτων (QoR), γίνεται διερεύνηση της επίδρασής της επί των εφαρμογών αξιολόγησης μέσω μιας ευρετικής ανάλυσης εξερεύνησης (4.4.2). Η ποιοτικότερη λύση έχει τη χαμηλότερη τιμή καταλληλότητας, δεδομένου ότι οι μετρικές του κόστους καθυστέρησης και καλωδίωσης μειώνονται, όσο το συνολικό QoR βελτιώνεται.

Σύμφωνα με την προηγούμενη αναπαράσταση του γενετικού αλγορίθμου και του FPGA, δημιουργείται ένας πληθυσμός από FPGA τοποθετήσεις για την εφαρμογή αναφοράς *alu4*, του πακέτου εφαρμογών επαλήθευσης MCNC. Η εφαρμογή *alu4* αποτελείται από 14 εισόδους, 8 εξόδους και 1519 CLB_s. Το άτομο- i αναγνωρίζεται από το γονότυπο CLB_jIO_k , ο οποίος αντιπροσωπεύει τις φυσικές θέσεις των CLB_s και IOs στο FPGA, στις οποίες έχουν τοποθετηθεί τα αντίστοιχα CLB_s και IOs της εφαρμογής *alu4*. Ο συνδυασμός των έξι διαφορετικών τοποθετήσεων CLB ($j = 6$) και των έξι διαφορετικών τοποθετήσεων IO ($k = 6$) διατυπώνει ένα σύνολο $n=36$ διαφορετικών λύσεων, οι οποίες απεικονίζονται στο Σχήμα 4.4. Ενώ ο γονότυπος CLB_1IO_1 έχει τη μέγιστη μετρική καταλληλότητας, υπάρχει μια αίσθηση κατά την οποία το αλληλόμορφο CLB_6 αποδίδει συνολικά καλύτερα μεταξύ των αλληλόμορφων CLB σε διάφορα γενετικά πλαίσια, μια ιδιότητα που προδιαγράφει την μετρική της αναμιξιμότητας. Οι συγγραφείς της εργασίας [30] έχουν δείξει ότι η εξελισσιμότητα βελτιώνεται

όταν αλληλόμορφα του ίδιου γονιδίου ανταγωνίζονται μεταξύ τους με βάση το πόσο καλά αποδίδουν κατά μέσο όρο (δηλαδή CLB_6) και όχι βάση του πόσο καλά αποδίδουν σε οποιονδήποτε συγκεκριμένο συνδυασμό (δηλαδή CLB_1IO_1). Το συμπέρασμα αυτό οδήγησε την παρούσα έρευνα στην κατασκευή φορέων αναπαραγωγής που οδηγούν σε τοποθετήσεις, οι οποίες αποδίδουν ποιοτικό βαθμό QoR κατά μέσο όρο, για όλες τις λύσεις του πληθυσμού και όχι για συγκεκριμένο συνδυασμό.



Σχήμα 4.4 ο χώρος λύσεων της εφαρμογής *alu4* με έξι αλληλόμορφα ανά αλληλότοπο. Τα αλληλόμορφα $CLB_{1:6}$ στον αλληλότοπο-CLB και τα $IO_{1:6}$ στον αλληλότοπο-IO συνιστούν 36 γονότυπους CLB_jIO_k με μετρική καταλληλότητας $f_{j,k}$, που αντιπροσωπεύεται από το ύψος των ράβδων.

Ωστόσο, το πρόβλημα που προκύπτει σε αυτή την προσέγγιση, είναι το επίπεδο της λεπτομέρειας που χαρακτηρίζει αυτούς τους συνδυασμούς. Στην περίπτωση του αλγορίθμου τοποθέτησης με κριτήριο την ελαχιστοποίηση των χρονικών καθυστερήσεων, ο στόχος είναι η τοποθέτηση λογικών δομών CLBs/IOs της netlist σε φυσικές θέσεις CLBs/IOs του FPGA, με τέτοιο τρόπο ώστε το κρίσιμο μονοπάτι⁴ να ελαχιστοποιείται, οδηγώντας στη μέγιστη δυνατή συχνότητα

⁴Το κρίσιμο μονοπάτι ενός σχεδιασμού αναφέρεται σε αυτή τη διασύνδεση που επιφέρει τη μεγαλύτερη καθυστέρηση στο κύκλωμα και συνεπώς καθορίζει την περίοδο ρολογιού του κυκλώματος. Το κρίσιμο μονοπάτι καθορίζεται από τον αριθμό των επιπέδων λογικής, την εσωτερική καθυστέρηση των λογικών μονάδων, την καθυστέρηση των καλωδίων, τη χωρητικότητα εισόδου των λογικών μονάδων, το βαθμό εξόδου των λογικών μονάδων (fanout) και την ισχύ οδήγησης εξόδου των λογικών

λειτουργίας του σχεδιασμού. Κάθε μονοπάτι μιας netlist είναι μια σύνδεση των ακμών γράφο αναπαράστασης της netlist, έτσι ώστε η πληροφορία να ρέει από μια πηγή (source) σε έναν κόμβο προορισμού (sink), κατά τη διάρκεια ενός κύκλου ρολογιού. Στην περίπτωση του FPGA, πηγές είναι οι ακροδέκτες εισόδου και οι έξοδοι των flip-flops, ενώ κόμβοι προορισμού είναι οι ακροδέκτες εξόδου και οι εισόδου των flip-flops. Το Σχήμα 4.5 απεικονίζει ένα χαρακτηριστικό σχεδιασμό, προκειμένου να παρουσιάσει αυτούς τους ορισμούς.

Κατά τη σύνθεση της γλώσσας HDL, εφαρμόζονται διάφορες τεχνικές βελτιστοποίησης, έτσι ώστε η αφαιρετική HDL περιγραφή να μπορεί να απεικονισθεί αποτελεσματικά στους φυσικούς πόρους του FPGA. Ενώ η σύνθεση, στο στάδιο της γλώσσας HDL, οδηγεί σε μία λεπτομερή netlist αποτελούμενη από γενικής περιγραφής πύλες και μοντέλα flip-flop, όπως απεικονίζεται στο Σχήμα 4.5(β), τα τελευταία στάδια της σύνθεσης ασχολούνται με την απεικόνιση αυτών των λεπτομερών δομών με τα φυσικά δομικά κύτταρα του FPGA, δηλαδή τα slices. Όπως περιγράφεται στην υποενότητα 4.3, σε αυτή την εργασία χρησιμοποιείται το εργαλείο T-VPack προκειμένου να ομαδοποιήσει LUTs και καταχωρητές σε μία netlist από slices. Ο στόχος αυτής της διαδικασίας, είναι η μεγιστοποίηση της αξιοποίησης των CLBs, με την ταυτόχρονη ελαχιστοποίηση του αριθμού των συνδέσεων μεταξύ των λογικών συστάδων του κρίσιμου μονοπατιού. Αυτή η διαδικασία ομαδοποίησης δημιουργεί εξαρτήσεις μεταξύ των slices που ανήκουν σε διαφορετικά μονοπάτια, ωστόσο για αυτά τα slices, γίνεται προσπάθεια τοποθέτησης στο ίδιο CLB, προκειμένου να ικανοποιηθεί η απαίτηση χωρητικότητας/πληρότητας που θέτει το εργαλείο T-VPack. Αναφορικά προς το παράδειγμα που απεικονίζεται στο Σχήμα 4.5(β), στην περίπτωση που τα $FF1, FF2, LUT3$ και $LUT4$ συσχεύονται στο ίδιο CLB ($CLB2$), τότε η τοποθέτηση του $CLB2$ επηρεάζει τόσο το μήκος καλωδίου του μονοπατιού $Path2$ όσο και του $Path3$.

Ενώ οι υφιστάμενοι αλγόριθμοι τοποθέτησης προσπαθούν να βρουν μία βέλτιστη λύση τοποθέτησης, μέσα από την ελαχιστοποίηση του μήκους των κρίσιμων μονοπατιών, σε αυτήν την έρευνα προτείνεται η χρήση εξειδικευμένων τελεστών αναπαραγωγής για γενετικούς αλγορίθμους, οι οποίοι λαμβάνουν υπόψιν την κατανομή των CLBs και IOs στα μονοπάτια του σχεδιασμού, δεδομένης μιας ομαδοποίησης. Για την επίτευξη αυτού του σκοπού, χρησιμοποιούνται νέες μετρικές και ορισμοί που συσχετίζουν τέτοιου τύπου χαρακτηριστικά. Αρχικά ορίζεται η μετρική *dependability*, η οποία συμβολίζεται με το σύμβολο ξ , ως εξής:

Definition 1 Δοθέντος ενός σχεδιασμού με n -κόμβους (CLBs, IOs) στον γράφο εκτέλεσης (π.χ. Σχήμα 4.5(c)), k -ακμές (συνδέσεις μεταξύ δύο κόμβων) και m -μονοπάτια ($m \leq k$), έτσι ώστε για κάθε μονοπάτι $p_{i=1:m}$ να υπάρχει ένα σύνολο S_{p_i} που περιέχει όλους τους κόμβους που ανήκουν στο μονοπάτι p_i , τότε για κάθε κόμβο N_j η μετρική *dependability*, ξ_j , ορίζεται ως:

$$\xi_j = \sum_{path \ i=1}^m f(i,j), \quad f(i,j) = \begin{cases} 0, & \text{when node } j \in S_{p_i} \\ 1, & \text{when node } j \notin S_{p_i} \end{cases} \quad (4.2)$$

μονάδων.

Η μετρική ξ_j για κάθε κόμβο j αναπαριστά μία ποιοτική τιμή της βαρύτητας τοποθέτησης για αυτόν τον κόμβο. Ο προτεινόμενος αλγόριθμος τοποθέτησης λαμβάνει υπόψιν του τη μετρική *dependability*, προκειμένου να συγκλίνει γρήγορα στο ολικό βέλτιστο. Αναλύουμε τη χρήση αυτής της μετρικής κατά τη διάρκεια της κατασκευής των βημάτων μετασχηματισμού του γενετικού αλγορίθμου.

Η δεύτερη μετρική που χρησιμοποιείται στην κατασκευή των τελεστών εξέλιξης, είναι η επιμέρους μέση τιμή καταλληλότητας (average sub-fitness value) των μονοπατιών, \bar{f}_p . Ενώ η γενική τιμή καταλληλότητας f ποσοτικοποιεί την τοποθέτηση ολόκληρου του σχεδιασμού, η νέα μετρική f_{p_i} ποσοτικοποιεί την τοποθέτηση ενός μοναδικού μονοπατιού, ανάλογα με το άθροισμα των πλαισίων οριοθέτησης των συνδέσεων του.

Definition 2 Δεδομένης μιας αρχιτεκτονικής FPGA με l -φυσικές δομικές μονάδες CLBs και IOs, ενός σχεδιασμού με n -κόμβους (CLBs, IOs) στο γράφο εκτέλεσης (π.χ. Σχήμα 4.5(c)), k -ακμές (συνδέσεις μεταξύ δύο κόμβων) και m -μονοπάτια ($m \leq k$), έτσι ώστε για κάθε μονοπάτι $p_{i=1:m}$ να υπάρχει ένα σύνολο R_{p_i} που περιέχει όλες τις ακμές που ανήκουν στο μονοπάτι p_i , και μιας λύσης τοποθέτησης $\Pi(\hat{n} \Rightarrow \hat{l})$, τότε η μετρική της επιμέρους μέσης τιμής καταλληλότητας όλων των μονοπατιών, $\bar{f}_{p_{\Pi}}$, για τη λύση τοποθέτησης Π , ορίζεται ως:

$$\bar{f}_{p_{\Pi}} = \sum_{\text{path } p_{i=1}}^m f_{p_i\Pi}, \quad (4.3)$$

όπου για κάθε μονοπάτι p_i η μετρική της επιμέρους μέσης τιμής καταλληλότητας, $f_{p_i\Pi}$, για τη λύση τοποθέτησης Π , ορίζεται ως:

$$f_{p_i\Pi} = \sum_{\text{edge } e_{j=1}}^{R_{p_i}} [bb_x(j) + bb_y(j)], \quad (4.4)$$

, όπου για κάθε ακμή j , οι συμβολισμοί $bb_x(j)$ και $bb_y(j)$ δηλώνουν τις οριζόντιες και κάθετες αποστάσεις του πλαισίου οριοθέτησης της τοποθέτησης (bounding box), αντίστοιχα.

Επίσης ορίζεται η μετρική *diversity* προκειμένου να ποσοτικοποιηθεί η ομοιότητα μεταξύ δύο λύσεων τοποθέτησης. Αυτή η μετρική χρησιμοποιείται από τον τελεστή mutation (4.3.1.4) και ορίζεται ως εξής:

Definition 3 Δεδομένης μιας αρχιτεκτονικής FPGA με l -φυσικές δομικές μονάδες CLBs και IOs, ενός σχεδιασμού με n -κόμβους (CLBs, IOs) στο γράφο εκτέλεσης (π.χ. Σχήμα 4.5(c)) και j λύσεων τοποθέτησης $\Pi_{\{1,j\}}(\hat{n} \Rightarrow \hat{l})$, τότε

για κάθε κόμβο N_j η μετρική diversity, $\Delta_{\Pi_{(1,j)}}$, ορίζεται ως:

$$\Delta_{\Pi_{(1,j)}} = \sum_{node\ i=1}^n g(i,j), \quad g(i,j) = \begin{cases} 0, & \text{when } \Pi_1[node_i] = \Pi_2[node_i] = \dots = \Pi_j[node_i] \\ 1, & \text{when } \Pi_1[node_i] \neq \Pi_2[node_i] \neq \dots \neq \Pi_j[node_i] \end{cases} \quad (4.5)$$

Σχήμα 4.5 (a) Δομικό πρότυπο αρχιτεκτονικής FPGA. (b) Σύνοψη HDL-σε-CLBs. (c) Αναπαράσταση ενός σχεδιασμού με γράφο εκτέλεσης. (d) Φυσική τοποθέτηση σε FPGA.

Οι μετασχηματισμοί στα άτομα ενός πληθυσμού αποτελούν τα βήματα ανασυνδυασμού και εξέλιξης του γενετικού αλγορίθμου. Αυτοί οι μετασχηματισμοί εκτελούνται από τέσσερις τελεστές έτσι ώστε τα καλά χαρακτηριστικά του πληθυσμού προσδιορίζονται και συνδυάζονται στον πληθυσμό της επόμενης γενιάς. Αυτή η νέα γενιά πληθυσμού (που ονομάζεται επίσης και απόγονος γενεά - offspring) έχει αισίως την ιδιότητα κατά την οποία τα ποιοτικότερα άτομα του πληθυσμού (τα οποία αντιπροσωπεύονται από την καλύτερη λύση τοποθέτησης) καθώς και η μέση τιμή της καταλληλότητας των ατόμων είναι βελτιστοποιημένα σε σχέση με προηγούμενες γενεές. Η διαδικασία αυτή επαναλαμβάνεται έως ότου επιτευχθεί σύγκλιση του αλγορίθμου σε κάποιο κριτήριο διακοπής. Οι τέσσερις βασικοί τελεστές ενός γενετικού αλγορίθμου, κατά τη δημιουργία μίας νέας γενεάς πληθυσμού, είναι η επιλογή (selection), η σύζευξη (mating), η διασταύρωση (crossover) και η μετάλλαξη (mutation). Στα εδάφια 4.3.1.1, 4.3.1.2,

4.3.1.3 και 4.3.1.4 περιγράφουμε με περισσότερες λεπτομέρειες αυτούς τους τελεστές. Ο Πίνακας 4.2 συνοψίζει τα σύμβολα και τους ορισμούς που χρησιμοποιούνται στην υπόλοιπη ενότητα.

Πίνακας 4.2 Οι παράμετροι του προτεινόμενου γενετικού αλγορίθμου.

Σύμβολο	Περιγραφή
$P(t)$	Population: Το σύνολο των τοποθετήσεων κατά το χρόνο t της εξέλιξης.
n_p	Population size: Το σύνολο των λύσεων τοποθέτησης.
g_p	Population age: Ο συνολικός αριθμός των γενεών εξέλιξης.
$E(t)$	Elitism pool: Το υποσύνολο του $P(t)$ με τα ποιοτικότερα άτομα.
p_e	Elitism percentage: Το ποσοστό του πληθυσμού που συνιστά το υποσύνολο E , $p_e \times n_p = size(E)$.
$M(t)$	Mating pool: Το υποσύνολο του $P(t)$ με τα άτομα που σχηματίζουν ζευγάρια για την νέα γενεά.
p_{mat}	Mating percentage: Το ποσοστό του πληθυσμού που συνιστά το υποσύνολο M , $p_{mat} \times n_p = size(M)$.
$C(t)$	Crossover pool: Το υποσύνολο του $M(t)$ μετά την εφαρμογή του τελεστή διασταύρωσης, $size(M) = size(C)$
b_m	Mutation probability: Η πιθανότητα εφαρμογής μετάλλαξης στο άτομο i του πληθυσμού $P(t)$.

Η τιμή t αντιπροσωπεύει το βήμα χρόνου της εξέλιξης, π.χ. την i^{th} γενεά του πληθυσμού.

4.3.1.1. Επιλογή βάση ελιτισμού

Μέσω του τελεστή της επιλογής δημιουργείται ένας νέος προσωρινός πληθυσμός, όπου κάθε μέλος του είναι ένα αντίγραφο ενός ατόμου του πληθυσμού της προηγούμενης γενεάς. Τα άτομα του πληθυσμού ταξινομούνται με βάση την μετρική καταλληλότητάς τους και σύμφωνα με μια προκαθορισμένη πιθανότητα, ένα ποσοστό του πληθυσμού θα παραμείνει ανέπαφο στον επόμενο πληθυσμό. Αυτή η τεχνική αναφέρεται ως ελιτισμός. Το επιδιωκόμενο αποτέλεσμα αυτής της διαδικασίας είναι η βελτίωση της ποιότητας του πληθυσμού συνολικά. Ωστόσο, μέσα από αυτή τη διαδικασία δεν δημιουργούνται πραγματικά νέες λύσεις. Για αυτό το σκοπό, νέα πληροφορία στον πληθυσμό εισέρχεται μέσω του τελεστή διασταύρωσης, όπως αναλύεται στην υποενότητα 4.3.1.3. Το ποσοστό ελιτισμού p_e είναι μια σημαντική παράμετρος της ποιότητας των τοποθετήσεων, όπως τονίζεται και στο παράδειγμα του Σχήματος 4.6. Καθώς οι ακραίες τιμές ελιτισμού, είτε χαμηλές (0%), είτε υψηλές (75%), οδηγούν σε τοποθετήσεις μικρής ποιότητας, όπως αναμενόταν, οι ενδιάμεσες τιμές παίζουν καθοριστικό ρόλο στην τελική ποιότητα κατά τη σύγκλιση του αλγορίθμου. Για παράδειγμα, η χρήση ποσοστού ελιτισμού $p_e=25\%$ για το κύκλωμα *s38417* του πακέτου MCNC, οδηγεί στη σύγκλιση σε ένα τοπικό βέλτιστο στην γενεά 219, ενώ κατά την ίδια γενεά, η χρήση ποσοστού ελιτισμού $p_e=50\%$ οδηγεί σε 16% λιγότερο ποιοτική λύση τοποθέτησης. Ο διαδικασία αυτή περιγράφεται με ψευδο-κώδικα στον Αλγόριθμο 1.

4.3.1.2. Σύζευξη βάση αναμιξιμότητας

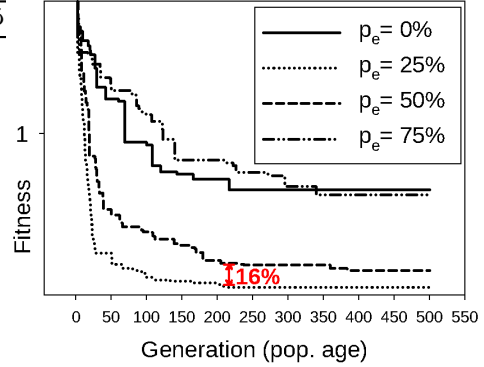
Κατά τι στάδιο του τελεστή σύζευξης, δύο άτομα της τρέχουσας γενεάς επιλέγονται προκειμένου να δημιουργήσουν ένα νέο άτομο. Αυτή η διαδικασία επα-

ΑΛΓΟΡΙΘΜΟΣ 1: Επιλογή βάση ελιτισμού**Input:** Population on time t : $P(t)$ **Input:** Elitism percentage p_e **Output:** 1) Fitness based Sorted Population on time t : $P_s(t)$ **Output:** 2) Elitism Pool on time $t + 1$: $E(t + 1)$

```

1 initialization;
2 for  $i = 1$  to  $n_p$  do
3   |  $f(i)$ =ComputeFitness(Individual $_i$ );
4 end
5  $P_s$ =SortPopulation( $P, *f$ );
6 for  $j = 1$  to  $(p_e \times n_p)$  do
7   |  $E_j(t + 1) \leftarrow P_s(\text{Individual}_j)$ ;
8 end

```



Σχήμα 4.6 Παράδειγμα επίδρασης του ποσοστού ελιτισμού για το κύκλωμα s38417 του πακέτου MCNC με 6406 CLBs, 29 εισόδους και 106 εξόδους, $n_p=100$, $g_p=500$, $mutation_rate=25$

να λαμβάνεται έως ότου όλα τα άτομα να συζευχθούν. Διαφορετικές προσεγγίσεις είναι δυνατόν να χρησιμοποιηθούν κατά τη διάρκεια αυτού του σταδίου. Ο συνήθης τρόπος ζευγαρώματος γονέων αποτελείται από τη λήψη δύο τυχαίων γονέων από το υποψήφιο υποσύνολο $M(t)$ του πληθυσμού. Οι γονείς ζευγαρώνουν και στη συνέχεια απομακρύνονται από το υποσύνολο $M(t)$. Για αυτή τη διαδικασία επιλογής, έχουν προταθεί στη βιβλιογραφία διάφορες τεχνικές σύζευξης, όπως η κοινή χρήση τιμής καταλληλότητας (fitness sharing) [31], οι πράκτορες εξέλιξης (evolving agents) [32], οι γενετικοί αλγόριθμοι με χρήση τεχνικών tabu search (tabu genetic algorithm) [33] και η αποπλάνηση (seduction) [34], οι οποίες δεν επιλέγουν ζευγάρια στην τύχη και έχει αποδειχθεί ότι βελτιώνουν αισθητά την απόδοση του γενετικού αλγορίθμου. Στον αλγόριθμο GENESIS, προτείνεται η ενσωμάτωση των προτιμήσεων σύζευξης με βάση τη θεωρία της αναμιξιμότητας, που έχει συζητηθεί ανωτέρω. Όπως φαίνεται στον Αλγόριθμο 2, τα ζεύγη επιλέγονται έτσι ώστε να εκθέτουν κατά το δυνατόν μικρή επιμέρους μέσης τιμής καταλληλότητας για όλα τα μονοπάτια, σε σύγκριση με οποιονδήποτε άλλο συνδυασμό ζευγαριών. Έτσι, σε τεκμηρίωση της γραμμής 15 του Αλγορίθμου 2, τα άτομα i και j συνδυάζονται μαζί στο υποσύνολο M , όταν οι συνολικές αντίστοιχες τοποθετήσεις τους έχουν μονοπάτια με ανάλογης ποιότητας τοποθετήσεις, δηλαδή τα άτομα ανταγωνίζονται μεταξύ τους με βάση το πόσο καλή ποιότητα παρέχουν κατά μέσο όρο στο σύνολο και όχι ανάλογα προς την παρεχόμενη ποιότητα για μεμονωμένους συνδυασμούς (“they compete with each other based on how well they perform on average rather than how well they perform in any one specific combination” [30]).

4.3.1.3. Διασταύρωση βάση ποιότητας μονοπατιών

Μετά την επιλογή των ζευγαριών αναπαραγωγής μέσω του τελεστή σύζευξης, ακολουθεί η ανταλλαγή πληροφορίας μεταξύ των χρωμοσωμάτων τους, για

ΑΛΓΟΡΙΘΜΟΣ 2: Σύζευξη βάση αναμιξιμότητας**Input:** Population on time t after Selection: $P(t)$ **Input:** Elitism pool on time $t + 1$: $E(t + 1)$ **Input:** Placement of every individual i on time t : $\Pi_i(t)$ **Input:** Mating percentage p_{mat} **Output:** Mating Pool on time $t + 1$: $M(t + 1)$

```

1 for  $i = 1$  to  $(p_{mat} \times n_p)$  do
2   for  $k = 1$  to  $m$  paths do
3      $Path\_Fitness\_Array[f_{p_k \Pi}]_i = \text{Compute\_path\_sub-}$ 
        $\text{fitness\_value}(Individual_i,$ 
          $path_k)$ ;
4   end
5 end
6 for  $i = 1$  to  $(p_{mat} \times n_p)$  do
7    $\min \leftarrow (+\infty)$ 
8   for  $j = i + 1$  to  $(p_{mat} \times n_p)$  do
9      $\text{Current\_path\_sub-fitness\_diff} = 0$ ;
10    for  $k = 1$  to  $m$  paths do
11       $\text{Current\_path\_sub-fitness\_diff} += \text{abs}(Path\_Fitness\_Array[f_{p_k \Pi}]_i -$ 
         $Path\_Fitness\_Array[f_{p_k \Pi}]_j)$ 
12    end
13    if  $(\text{Current\_path\_sub-fitness\_diff} < \min)$  then
14       $\min = \text{Current\_path\_sub-fitness\_diff}$ ;
15       $M(t + 1) \leftarrow P_s(Individual_i) \oplus P_s(Individual_j)$ ;
16    end
17  end
18 end

```

όλα τα άτομα στο υποσύνολο σύζευξης M . Συνήθως, στις υφιστάμενες προσεγγίσεις γενετικών αλγορίθμων, για κάθε ζεύγος, ο τελεστής διασταύρωσης εφαρμόζεται με μια προκαθορισμένη πιθανότητα, επιλέγοντας μία τυχαία θέση στο χρωμόσωμα και ανταλλάσσοντας τμήμα του χρωμοσώματος που αναφέρεται από αυτή τη θέση, χρησιμοποιώντας μια πολιτική αντικατάστασης. Για παράδειγμα, η απλούστερη εκδοχή του τελεστή διασταύρωσης εφαρμόζεται με την αντικατάσταση των τελευταίων ψηφίων των δύο αλφαριθμητικών. Η επίδραση του τελεστή διασταύρωσης είναι ότι ορισμένες ιδιότητες των ατόμων συνδυάζονται, δημιουργώντας εντελώς καινούριες ενώ άλλες ιδιότητες χάνονται εντελώς. Αναγνωρίζοντας το γεγονός ότι μιας υψηλής ποιότητας τοποθέτηση σε ένα FPGA αντιστοιχεί επίσης και σε υψηλή ποιότητα τοποθέτησης των επιμέρους μονοπατιών της, η προτεινόμενη προσέγγιση αποφεύγει την καταστροφή των μονοπατιών υψηλής ποιότητας, και ωθεί τον τελεστή διασταύρωσης να εφαρμοστεί σε μικρότερης εμβέλειας περιοχές, όπως είναι τα μονοπάτια, και όχι σε ολόκληρο το σχεδιασμό. Δηλαδή, αντί να ανταλλαχθούν τοποθετήσεις τυχαίων CLB_i ανάμεσα σε δύο άτομα, αντικαθίστανται όλα τα CLB και IO που ανήκουν στο τυχαίο μονοπάτι $path_j$.

Δεδομένου ότι ένας μεγάλος αριθμός επικαλύψεων μπορεί να συμβεί με τη χρήση μιας τέτοιας προσέγγισης, προτείνεται η εισήγηση της μετρικής *dependability* ξ , προκειμένου να αντιμετωπισθεί αυτό το πρόβλημα. Συγκεκριμένα, κάθε φορά που η ανταλλαγή ενός μονοπατιού μεταξύ δύο ατόμων του πληθυσμού αναγκάζει την τοποθέτηση ενός CLB/IO μία φυσική θέση του FPGA, στην υπάρχει ήδη τοποθετημένο άλλο CLB/IO, τότε η φυσική θέση ανατίθεται στο CLB/IO με την υψηλότερη τιμή της μετρικής *dependability*- ξ . Η πολιτική αυτή, πρακτικά, διασφαλίζει ότι το υψηλής τιμής ξ CLB/IO θα τοποθετηθεί σε μία φυσική θέση, έτσι ώστε να επηρεάζει την ποιότητα λιγότερων συνολικά διαδρομών σε σχέση με το χαμηλής τιμής ξ CLB/IO. Η διαδικασία αυτή εξηγείται επίσης μέσω του Αλγορίθμου 3.

4

ΑΛΓΟΡΙΘΜΟΣ 3: Διασταύρωση βάση ποιότητας μονοπατιών

Input: Population on time t after Mating: $P(t)$

Input: Mating pool on time $t + 1$: $M(t + 1)$

Input: Placement of every individual i on time t : $\Pi_i(t)$

Input: Dependability vector ξ for every node (CLB/IO) of netlist.

Output: Crossover Pool on time $t + 1$: $C(t + 1)$

```

1 for  $i = 1$  to  $size(M(t + 1) \text{ step } 2)$  do
2   mateA =  $M(t + 1)_i$ ; mateB =  $M(t + 1)_{i+1}$ ;
3   rpath =  $rand(1, n \text{ paths})$ ;
4   /* TryExchangePath(MateA, MateB, rpath); */
5   for  $k = 1$  to  $m \text{ nodes of rpath}$  do
6     foreach (MateA, MateB) do
7       if ( $\Pi[node_k]$  overlaps with  $\Pi[node_{k'}]$ ) /*  $node_{k'}$  has been mapped to
8          $\Pi[node_{k'}]$  at  $P(t)$  of the other mate*/ then
9         if ( $\xi[k]_i < \xi[j]_i$ ) then
10          | Assign  $node_{k'}$  to free node or exchange with low- $\xi$  node;
11         else
12          | Assign  $node_k$  to free node or exchange with low- $\xi$  node;
13        end
14      else
15      | ExchangePath(MateA, MateB, rpath); /* no overlap exists */
16    end
17  end
18 end
  
```

4.3.1.4. Αυτο-προσαρμοστική Μετάλλαξη

Οι προηγούμενοι τελεστές κατασκευάζονται χρησιμοποιώντας μια συντηρητική μεθοδολογία σχετικά με την ελευθερία των χρωμοσωμάτων στο χώρο αναζήτησης. Κάθε ένας τελεστής εκ της επιλογής βάση ελιτισμού, της σύζευξης βάση αναμιξιμότητας και της διασταύρωσης βάση ποιότητας μονοπατιών, ορίζει συσχετίσεις στις επιλογές εξέλιξης και συνεπώς μειώνει το βαθμό ελευθερίας στην τυχαία επιλογή, που αποτελεί το βασικό κριτήριο υπεροχής των γενετι-

κών αλγορίθμων. Συνεπώς υπάρχει μεγάλη πιθανότητα πρόωρης σύγκλισης του γενετικού αλγορίθμου σε υπο-βέλτιστη λύση. Για να αποφευχθεί αυτό, ο προτεινόμενος αλγόριθμος περιλαμβάνει ένα τελεστή μετάλλαξης. Κατά τη διάρκεια της εφαρμογής αυτού του τελεστή, ένα μονοπάτι επιλέγεται τυχαία και οι κόμβοι του ανατίθενται σε τυχαίες θέσεις στο FPGA. Ο σκοπός της μετάλλαξης είναι η αποφυγή εγκλεισμού του αλγορίθμου σε τοπικά ελάχιστα, αλλά και η εξασφάλιση διερεύνησης νέων λύσεων στο χώρο αναζήτησης. Ένα πρόβλημα που προκύπτει από αυτόν τον τελεστή, είναι ο καθορισμός της πιθανότητας με την οποία συμβαίνει μετάλλαξη σε άτομα του πληθυσμού. Από τη μία πλευρά, η πολύ μικρή πιθανότητα μετάλλαξης μπορεί να προκαλέσει τη σύγκλιση σε ένα τοπικό ελάχιστο, ενώ αντίθετα, οι πολλές συχνές μεταλλάξεις αποτρέπουν τη σύγκλιση του γενετικού αλγορίθμου, εξαναγκάζοντας τον αλγόριθμο σε ένα τυχαίο περίπατο.

Προς την εύρεση της κατάλληλης πιθανότητας για την εφαρμογή της μετάλλαξης, ο αλγόριθμος GENESIS παρέχει μια τεχνική αυτο-προσαρμοσιμότητας η οποία προσφέρει δυναμική πιθανότητα μετάλλαξης. Ενώ υπάρχουν υφιστάμενες εργασίες για τον έλεγχο της πιθανότητας μετάλλαξης κατά το χρόνο εκτέλεσης (π.χ. [35], [36]), οι περισσότερες από αυτές χρησιμοποιούν συχνά τη μετρική της καταλληλότητάς μεταξύ διαδοχικών γενεών, προκειμένου να προσαρμόσουν την πιθανότητα μετάλλαξης. Αυτή η προσέγγιση επεκτείνεται με την καθοδήγηση της επιλογής της πιθανότητας με βάση την ομοιότητα στις λύσεις τοποθετήσεων των ατόμων του πληθυσμού. Με βάση την μετρική *diversity*, η οποία αντιπροσωπεύει μια ποσοτική τιμή προς την ομοιότητα των τοποθετήσεων μεταξύ δύο ατόμων, η πιθανότητα μετάλλαξης αυξάνεται αναλογικά με τη τιμή *diversity* και αντίστροφα. Συνεπώς χρησιμοποιείται μια αντίστροφως ανάλογη συνάρτηση ενεργοποίησης f_d για να υποστηρίξει αυτή τη δυνατότητα, η λειτουργίας

της οποίας περιγράφεται από τον Αλγόριθμο 4.

ΑΛΓΟΡΙΘΜΟΣ 4: Αυτο-προσαρμοστική Μετάλλαξη

Input: Population on time t after Crossover: $P(t)$

Input: Crossover pool on time $t + 1$: $C(t + 1)$

Input: Placement of every individual i on time t : $\Pi_i(t)$

Input: Dependability vector ξ for every node of netlist.

Output: Population on time t : $P(t + 1)$

```

1 for  $i = 1$  to  $n_p$  do
2    $\Delta_p = \text{CalcDiversity}(P(t));$ 
3    $b_m = f_d(\Delta_p);$ 
4   if ( $\text{rand}(0,1) < b_m$ ) then
5      $rpath = \text{rand}(1, n \text{ paths});$ 
6     for  $k = 1$  to  $m$  nodes of  $rpath$  do
7       /* Random FPGA CLB for netlist CLB and FPGA IO for netlist IO */
8       /* Mutation of  $node_k$  */
9        $\Pi[node_k] = \text{rand}(1, m \text{ nodes});$ 
10      if ( $\Pi[node_k]$  overlaps with  $\Pi[node_{k'}]$ ) /* $node_{k'}$  has been mapped to  $\Pi[node_{k'}]$ 
        at  $P(t)$  of the other mate*/ then
11        if ( $\xi[k]_i < \xi[j]_i$ ) then
12          | Assign  $node_{k'}$  to free node or exchange with low- $\xi$  node;
13        else
14          | Assign  $node_k$  to free node or exchange with low- $\xi$  node;
15        end
16      end
17    end
18  end
19 end

```

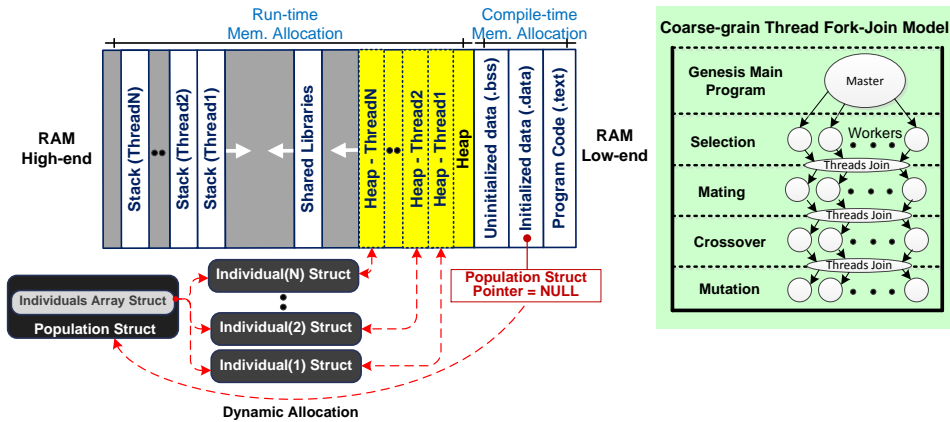
4.3.2. Η αδρομερής μηχανή παραλληλισμού του αλγορίθμου GENESIS

Το βασικό κίνητρο για την επίλυση του προβλήματος της τοποθέτησης με γενετικούς αλγορίθμους είναι η δυνατότητα διαίρεσης του συνολικού προβλήματος σε έναν αριθμό διακριτών υπο-προβλημάτων, τα οποία μπορούν να λυθούν ταυτόχρονα με τη χρήση συστημάτων πολλαπλών επεξεργαστών. Αυτή η προσέγγιση, που ακολουθεί το μοντέλο *διαίρει-και-βασίλευε*, εφαρμόζεται χρησιμοποιώντας μια αδρομερή προσέγγιση: οι τελεστές *επιλογή*, *σύζευξη*, *διασταύρωση* και *μετάλλαξη* ομαδοποιούνται σε διεργασίες μη επικαλυπτόμενων δεδομένων και εκτελούνται παράλληλα. Αυτή η τεχνική είναι εφικτή, διότι οι τελεστές που εφαρμόζονται σε ένα υποσύνολο του πληθυσμού είναι ανεξάρτητοι από εκείνους που εφαρμόζονται στον υπόλοιπο πληθυσμό, και συνεπώς δεν υπάρχει καμία απαίτηση για την επικοινωνία μεταξύ των νημάτων κατά τη διάρκεια αυτής της φάσης⁵. Συγκεκριμένα, κάθε τελεστής εργάζεται σε πολλά υποσύνολα των ατόμων του πληθυσμού. Έτσι, βάση μίας αδρομερούς περιγραφής, ο παραλληλισμός επιτυγχάνεται ως εξής: το έργο επεξεργασίας του κάθε υποσυνόλου εκχωρείται

⁵Μια εξαίρεση σε αυτή την απαίτηση επικοινωνίας αποτελεί ο τελεστής της επιλογής, ο οποίος χρειάζεται να ταξινομήσει όλα τα άτομα του πληθυσμού με βάση την μετρική της καταλληλότητάς τους.

σε ένα διαφορετικό παράλληλο νήμα. Έτσι οι εργασίες των τελεστών εκτελούνται παράλληλα κατά την διάρκεια μίας γενεάς του γενετικού αλγορίθμου.

Αυτή η προσέγγιση αδρομερούς παραλληλισμού στο πρόβλημα της τοποθέτησης είναι αποτελεσματική λόγω της τοπικότητας των αναφορών, τόσο σε επίπεδο εντολών, όσο και σε επίπεδο δεδομένων, για κάθε επανάληψη των τελεστών του γενετικού αλγορίθμου. Επίσης, η προτεινόμενη τεχνική αναδεικνύεται λόγω της δυνατότητας παραλληλισμού σε επίπεδο διεργασίας, καθώς όλα τα κοινά δεδομένα του πληθυσμού μπορούν να ανταλλάσσονται σε ένα ντετερμινιστικό μοτίβο παραγωγού/καταναλωτή, μειώνοντας τις πιθανότητες ταυτόχρονης επεξεργασίας των δεδομένων (data races) και μειώνοντας έτσι το ακολουθιακό ποσοστό του γενετικού αλγορίθμου.



Σχήμα 4.7 Η πολυ-νηματική δομή του υποσυστήματος μνήμης και το αντίστοιχο παράλληλο μοντέλο επεξεργασίας fork-join.

Η συνολική προσέγγιση για την παραλληλοποίηση του γενετικού αλγορίθμου, απεικονίζεται στο Σχήμα 4.7. Ως γλώσσα προγραμματισμού επιλέχθηκε η γλώσσα C [37] και ως προγραμματιστικό περιβάλλον για την παραλληλία του αλγορίθμου επιλέχθηκε το OpenMP API [38]. Προκειμένου να παρέχεται αποτελεσματική πρόσβαση στην κοινή μνήμη από κάθε νήμα επεξεργασίας, αναπτύχθηκε μία οργάνωση μνήμης προσανατολισμένη στις ανάγκες αποθήκευσης δεδομένων ενός ατόμου. Ο συνολικός πληθυσμός οργανώνεται σαν μία δομή (struct) που αποθηκεύεται στην κοινή μνήμη και αποτελείται από μια σειρά από δομές που αντιστοιχούν στα άτομα του πληθυσμού. Κάθε επιμέρους ατομική δομή αναφέρεται περαιτέρω σε μια σειρά από δομές που αποθηκεύουν σύνθετα δεδομένα του ατόμου, όπως η δομή τοποθέτησης, η δομή του γράφου εκτέλεσης και η δομή των μετρικών ποιότητας της τοποθέτησης. Αυτές οι δομές αρχικοποιούνται στα πρώτα στάδια του χρόνου εκτέλεσης, μέσω εξατομικευμένων κλήσεων δυναμικής διαχείρισης μνήμης (malloc). Δεδομένου ότι οι δεσμεύσεις/αποδεσμεύσεις της μνήμης δεν εμφανίζουν κάποια σχέση παραγωγού/καταναλωτή, η αρχιτεκτονική της σωρού (heap) έχει προσαρμοστεί στην δομή των *pure private heaps*, που περιγράφεται στην εργασία [39]. Συγκεκριμένα, μια τέτοια προσαρμογή προϋ-

ποθέτει ότι κάθε νήμα έχει τη δική του ιδιωτική σωρό, που χρησιμοποιείται για τις ανάγκες μνήμης αποκλειστικά της δικής του λειτουργίας. Όσον αφορά τις επιδόσεις και την επεκτασιμότητα, τα *pure private heaps* αποτελούν μια πολύ αποτελεσματική οργάνωση της μνήμης σε πολυ-νηματικά περιβάλλοντα, δεδομένου ότι κάθε νήμα έχει το δικό του ιδιωτικό χώρο σωρού και συνεπώς επιτυγχάνεται ο λιγότερο δυνατός ρυθμός διακοπών προσπέλασης της μνήμης (lock contention).

Η αποτελεσματικότητα του αλγορίθμου GENESIS βασίζεται στην παράλληλη εκτέλεση των επί μέρους διεργασιών του. Ωστόσο, ο παραλληλισμός πραγματοποιείται εντός και όχι μεταξύ, των βασικών τελεστών εξέλιξης του αλγορίθμου, δηλαδή την επιλογή, το σύζευξη, τη διασταύρωση και τη μετάλλαξη. Κάθε τελεστής εφαρμόζεται σε πολλά υποσύνολα των ατόμων του πληθυσμού. Έτσι, βάση μιας αδρομερούς περιγραφής, ο παραλληλισμός επιτυγχάνεται ως εξής: η επεξεργασία του κάθε υποσυνόλου εκχωρείται σε ένα κύριο νήμα (master thread), το οποίο διαχωρίζει τις εργασίες που πρέπει να εκτελεστούν παράλληλα σε κάθε εποχή της εξέλιξης και τις αναθέτει περαιτέρω σε παράλληλα νήματα επεξεργασίας. Με τη διερεύνηση πολλών λύσεων ταυτόχρονα, είναι περισσότερο πιθανό να βρεθεί μια λύση υψηλής ποιότητας, σε σύγκριση με τους αλγορίθμους τοποθέτησης οι οποίοι αρχίζουν από μια τυχαία κατάσταση και σταδιακά τη βελτιώνουν, όπως για παράδειγμα ο αλγόριθμος της προσημειωμένης ανόπτησης, που χρησιμοποιεί το εργαλείο VPR.

Σε γενικές γραμμές, η επιβάρυνση της διαχείρισης παραλληλισμού είναι ένα πρόβλημα για τα μοντέλα προγραμματισμού κοινής μνήμης. Ωστόσο, έχει αποδειχθεί ότι τα προγράμματα που παραλληλοποιούνται για αρχιτεκτονικές κοινής μνήμης μπορούν να επιτύχουν ικανοποιητική κλιμάκωση στους επεξεργαστικούς πόρους, ακόμα μέχρι και για εκατοντάδες επεξεργαστών [40], [41]. Ωστόσο αυτό επιτυγχάνεται με λογική κλιμάκωση του μεγέθους του προβλήματος, προκειμένου να αυξηθεί το ποσοστό εργασίας του κάθε νήματος και να ελαχιστοποιηθεί η συχνότητα συγχρονισμού μεταξύ των νημάτων. Συνεπώς, προκειμένου να εντοπιστεί ο σωστός συνδυασμός παραλληλισμού τόσο σε επίπεδο δεδομένων, όσο και εντολών, αποφεύγοντας μεγάλο ποσοστό συγχρονισμού των κοινών δεδομένων, κατά τη διάρκεια ανάπτυξης του προτεινόμενου αλγορίθμου, εφαρμόστηκε χαρακτηρισμός αυτού κατά τη φάση εκτέλεσης. Αυτή η διαδικασία επιτρέπει την καταγραφή των συναλλαγών των δεδομένων σε δια-νηματικό και ενδο-νηματικό επίπεδο. Επομένως αποτρέπεται η επικάλυψη της επιτάχυνσης που προέρχεται από την παραλληλοποίηση της εφαρμογής, από την καθυστέρηση που επιφέρει η επικοινωνία και ο συγχρονισμός των νημάτων. Με βάση αυτή την ανάλυση, διαπιστώθηκε ότι η μέγιστη τοπικότητα των αναφορών στην κύρια μνήμη επιτεύχθη στον επίπεδο του κάθε τελεστή. Αυτό είναι αναμενόμενο καθώς ο καθένας από αυτούς τους τελεστές συγκεντρώνει τους ίδιους τύπους δεδομένων επί του πληθυσμού. Δεδομένου, επίσης, ότι κάθε τελεστής λειτουργεί σε έως δύο λύσεις ανά επανάληψη, επελέγη ένα σύστημα αδρομερούς παραλληλισμού, όπου ο κύριος βρόχος του κάθε τελεστή διαιρείται σε πολλαπλά νήματα, ανάλογα προς τον αριθμό των διαθέσιμων πυρήνων επεξεργασίας του εκάστοτε υπολογιστικού συστήματος και τις επιλογές του σχεδιαστή.

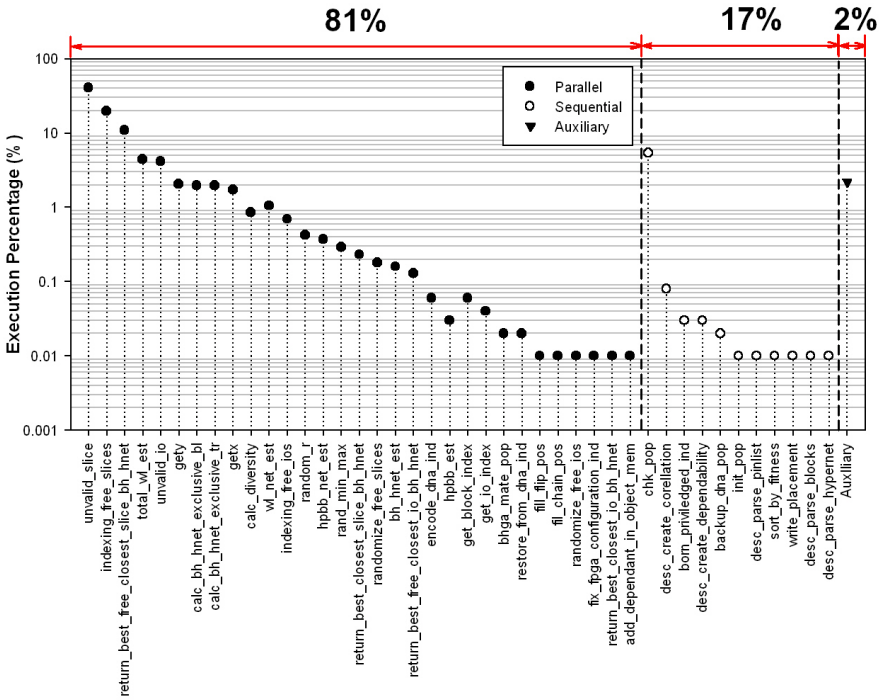
Προκειμένου να διευκρινιστεί καλύτερα η προαναφερθείσα διαδικασία, το

Σχήμα 4.8 απεικονίζει το μέσο όρο των ποσοστών εκτέλεσης επί του χρόνου επεξεργασίας για τις συναρτήσεις που συνθέτουν τον αλγόριθμο GENESIS, σε ένα δείγμα από τα είκοσι μεγαλύτερα προβλήματα τοποθέτησης του πακέτου MCNC. Τα αποτελέσματα αυτά ανακτήθηκαν με τη χρήση του εργαλείου Callgrind, του πακέτου Valgrind [42]. Για την ανάλυση αυτή, οι συναρτήσεις του αλγορίθμου συγκεντρώθηκαν σε δύο ομάδες ανάλογα με το αν έχουν την ικανότητα να εκτελεστούν παράλληλα με χρήση του OpenMP API, ή όχι. Όπως είναι απόλυτα λογικό, τα αποτελέσματα αυτά εξαρτώνται από το μέγεθος της εφαρμογής (netlist), την αρχιτεκτονική του FPGA, τον αριθμός των παράλληλων φυσικών πυρήνων του συστήματος επεξεργασίας, τη χρήση της τεχνικής HyperThreading, καθώς και από πολλές παραμέτρους του αλγορίθμου GENESIS, όπως είναι το μέγεθος του πληθυσμού, τα βήματα εξέλιξης, το ποσοστό ελιτισμού, το ποσοστό σύζευξης και η πιθανότητα μετάλλαξης. Ωστόσο οι αποκλίσεις στα ποσοστά εκτέλεσης χρόνου των επί μέρους συναρτήσεων του αλγορίθμου είναι αρκετά μικρές. Με βάση το Σχήμα 4.8, είναι εμφανές ότι, κατά μέσο όρο, το 81% του συνολικού χρόνου εκτέλεσης αντιστοιχεί σε πηγαίο κώδικα που παραλληλοποιείται με τη χρήση του προγραμματιστικού πλαισίου OpenMP. Αυτό το αποτέλεσμα θέτει ένα άνω όριο στη θεωρητική επιτάχυνση που μπορεί να επιτευχθεί σε έναν πολυ-νηματικό επεξεργαστή κοινής μνήμης, όπως αναλύεται περαιτέρω στην υποενότητα 4.5.

4.4. Προσαρμογή του εργαλείου GENESIS στα χαρακτηριστικά της εφαρμογής

Ο αλγόριθμος τοποθέτησης GENESIS, περιλαμβάνει μια σειρά από παραμέτρους που ελέγχουν τη διαδικασία βελτιστοποίησης. Σε ένα συμβατικό σενάριο υλοποίησης, αυτές οι παράμετροι είναι προ-ρυθμισμένες από τους χρήστες του εργαλείου, συνεπώς κάθε εφαρμογή βελτιστοποιείται σύμφωνα με ένα σταθερό διάλυμα παραμέτρων που αντιπροσωπεύουν ένα μοναδικό στιγμιότυπο εκτέλεσης του αλγορίθμου. Ωστόσο, υπάρχει μεγάλη διαφοροποίηση αναφορικά προς τις δομές του κυκλώματος και την οργάνωσή τους, ανάλογα με την εκάστοτε εφαρμογή, και συνεπώς είναι εξαιρετικά δύσκολο να βρεθεί μία ενιαία διαμόρφωση, η οποία βελτιστοποιεί τον αλγόριθμο τοποθέτηση για όλες τις εφαρμογές και όλες τις επιθυμητές προδιαγραφές λειτουργίας. Για να αντιμετωπιστεί αυτή η διαφοροποίηση, επεκτείνεται το εργαλείο GENESIS με ένα μεθοδολογικό πλαίσιο που επιτρέπει τη ρύθμιση των παραμέτρων του, ανάλογα με τα ιδιαίτερα χαρακτηριστικά της εφαρμογής.

Η προτεινόμενη μεθοδολογία προσαρμογής αποτελείται από δύο φάσεις. Λαμβάνοντας υπόψη ένα σύνολο αντιπροσωπευτικών εφαρμογών, κατά τη Φάση 1 πραγματοποιείται μία διαδικασία ομαδοποίησης των εφαρμογών που παρουσίασαν παρεμφερή χαρακτηριστικά. Οι εφαρμογές οι οποίες συγκεντρώνονται στην ίδια ομάδα, διερευνώνται με ένα συνδυασμένο τρόπο, προκειμένου να χαρακτηρίσουν το χώρο λύσεων της συγκεκριμένης ομάδας. Γενικά, ακολουθείται η προσέγγιση της πολυεπίπεδης βελτιστοποίησης, έτσι από τη διαδικασία



Σχήμα 4.8 Κατανομή των ποσοστών του χρόνου εκτέλεσης για τις συναρτήσεις πηγαίου κώδικα του αλγορίθμου GENESIS.

χαρακτηρισμού κάθε ομάδας, εξάγεται μια σειρά από προσεγγισμένες βέλτιστες διαμορφώσεις Pareto, έτσι ώστε να δίνεται η δυνατότητα στο σχεδιαστή να επιλέγει τα εκάστοτε επιθυμητά κριτήρια συμβιβασμού ανάλογα με τις προδιαγραφές σχεδίασης, π.χ. Ενέργεια×Καθυστέρηση, Ενέργεια×Επιφάνεια πυριτίου, κοκ.

Κατά τη Φάση 2, η επιλεγμένη διαμόρφωση ανά ομάδα, η οποία έχει υπολογισθεί κατά τη Φάση 1, θα χρησιμοποιηθεί στην εκτέλεση του εργαλείου GENESIS, προκειμένου να διαμορφώσει τις παραμέτρους του εξελικτικού αλγόριθμου βελτιστοποίησης ανάλογα με τα χαρακτηριστικά της εφαρμογής. Η ανάθεση κάθε νέα εφαρμογής σε μία συγκεκριμένη ομάδα - επομένως και σε ένα συγκεκριμένο διάλυσμα διαμόρφωσης - βασίζεται στην απόσταση συγκεκριμένων μετρικών χαρακτηρισμού της, από τις αντίστοιχες μετρικές της ομάδας. Όσο πιο μικρή είναι αυτή η απόσταση τόσο παρόμοια είναι τα χαρακτηριστικά αυτής της εφαρμογής με το σύνολο των εφαρμογών που συνθέτουν την εκάστοτε ομάδα. (In case of a tie, i.e. the distance between two or more centroids is the same, the application is assigned to the cluster with the higher cardinality.)

4.4.1. Προσαρμοσμένη Ομαδοποίηση σε Επίπεδο Εφαρμογής

Ο στόχος της διαδικασίας ομαδοποίησης σε επίπεδο εφαρμογής είναι η ομαδοποίηση των εφαρμογών που εμφανίζουν παρόμοια χαρακτηριστικά ως προς τη δομή τους κυκλώματός τους, βάση ενός συνόλου αντιπροσωπευτικών εφαρμογών. Χωρίς βλάβη της γενικότητας, στην εργασία αυτή εξετάστηκαν οι είκοσι μεγαλύτερες εφαρμογές της σουίτας αξιολόγησης MCNC και οι οποίες παρατίθενται στον Πίνακα 4.3. Ο στόχος της ομαδοποίησης ανά εφαρμογή είναι η συγκέντρωση των εφαρμογών που εμφανίζουν παρόμοια χαρακτηριστικά. Για κάθε μία από τις υπό-σύσταση ομάδες, διερευνάται μια βελτιστοποιημένη ρύθμιση, προκειμένου να χρησιμοποιηθεί κατά τη διάρκεια εκτέλεσης του εργαλείου GENESIS. Συνεπώς, κατά την έναρξη της διαδικασίας τοποθέτησης, κάθε εφαρμογή κατατάσσεται σε μία από αυτές τις ομάδες έτσι ώστε ο αλγόριθμος βελτιστοποίησης να είναι ρυθμισμένος κατάλληλα. Για αυτό το λόγο, η ομαδοποίηση βασίζεται σε χαρακτηριστικά που είναι διαθέσιμα μετά από το στάδιο της λογικής σύνθεσης και της τεχνολογικής ομαδοποίησης (technology packing) και όχι κατά το στάδιο της τοποθέτησης της εφαρμογής. Με αυτό το τρόπο εξασφαλίζεται ο χαρακτηρισμός των εφαρμογών βάση αποκλειστικά των ιδιαιτεροτήτων που εμφανίζουν στην είσοδο του προβλήματος τοποθέτησης.

Κατά το στάδιο της λογικής σύνθεσης και της τεχνολογικής ομαδοποίησης, η περιγραφή του κυκλώματος της εκάστοτε εφαρμογής αναλύεται ποσοτικά. Η ανάλυση αυτή ποικίλλει ανάλογα με τα εργαλεία σύνθεσης και τεχνολογικής ομαδοποίησης. Προκειμένου η προτεινόμενη ανάλυση να είναι ανεξάρτητη από συγκεκριμένα εργαλεία σύνθεσης, στα πλαίσια αυτής της έρευνας, εξετάζεται ένα ελάχιστο αντιπροσωπευτικό σύνολο των διαθέσιμων δεδομένων ανάλυσης, το οποίο αναφέρεται σε ένα διάνυσμα μέτρησης, π.χ. $\#CLBs, \#IOs, \#Nets$. Οι τέσσερις πρώτες στήλες του Πίνακα 4.3 δείχνουν τα αποτελέσματα της λογικής σύνθεσης και της τεχνολογικής ομαδοποίησης για ένα δεδομένο σύνολο αντιπροσωπευτικών εφαρμογών.

Μετά το στάδιο της σύνθεσης, μια εφαρμογή αξιολογείται ως προς τον αριθμό των δεσμευμένων πόρων. Ωστόσο, αυτή η προσέγγιση αδυνατεί να χαρακτηρίσει την εσωτερική δομή του κυκλώματος, που είναι ιδιαίτερα σημαντική για τη διαδικασία της τοποθέτησης. Το προτεινόμενο πλαίσιο αντιμετωπίζει αυτή την αναποτελεσματικότητα του διανύσματος μετρήσεων αξιολόγησης, ορίζοντας ένα νέο σύνολο χαρακτηριστικών που αντικατοπτρίζει με ουσιαστικότερο τρόπο τη δομή του κυκλώματος, για το στάδιο της τοποθέτησης. Συνεπώς ορίζονται δύο νέα χαρακτηριστικά και επαναχρησιμοποιείται ο αριθμός των συνδέσεων ($\#Nets$), ο οποίος είναι ένα χαρακτηριστικό που βρίσκεται και στο αρχικό διάνυσμα μετρικών αξιολόγησης. Συγκεκριμένα, χρησιμοποιείται το παρακάτω διάνυσμα χαρακτηριστικών:

$$f_1 = \#Nets \quad (4.6)$$

$$f_2 = \frac{\#LUTs}{\frac{\#IOs + \#DFFs}{2}} \quad (4.7)$$

$$f_3 = \begin{cases} \sqrt{\#CLBs}, & \text{if } \sqrt{\#CLBs} > \frac{\#IOs}{4}. \\ \frac{\#IOs}{4}, & \text{if } \sqrt{\#CLBs} \leq \frac{\#IOs}{4}. \end{cases} \quad (4.8)$$

Το χαρακτηριστικό f_1 αναφέρει την ένταση κάθε εφαρμογής όσον αφορά το πλήθος των συνδέσεων μεταξύ των κόμβων του γράφου εκτέλεσης. Αυτοί οι κόμβοι αναφέρονται στις λεπτομερείς μονάδες που συστήνουν την αρχιτεκτονική ενός FPGA, όπως τα flip-flops, τα LUTs, αλλά και στις περισσότερες αδρομερείς μονάδες, όπως τα DSPs. Το χαρακτηριστικό f_2 εκτιμά το μέγεθος καθυστέρησης των μονοπατιών του σχεδιασμού, δηλαδή τον αριθμό των LUTs που ανήκουν σε μία διαδρομή που αποτελείται από δύο DFF ή από δύο I/O κόμβους, δεδομένης μίας ισοροπημένης κατανομή των κόμβων στα μονοπάτια του σχεδιασμού. Τέλος, το χαρακτηριστικό f_3 υπολογίζει το ελάχιστο μέγεθος FPGA στο οποίο μπορεί να τοποθετηθεί πλήρως μία εφαρμογή. Σε αυτό το σημείο γίνονται δύο υποθέσεις. Η πρώτη αναφέρεται στη γεωμετρική διάταξη του FPGA. Για την συγκεκριμένη εργασία, έχει θεωρηθεί τετραγωνική διάταξη. Η δεύτερη υπόθεση αναφέρεται στο είδος της εφαρμογής, αναφορικά προς τη δραστηριότητα των κόμβων της. Διακρίνονται δύο κατηγορίες, οι εφαρμογές στις οποίες το μεγαλύτερο ποσοστό της δραστηριότητας αναφέρεται στους λογικούς πόρους (compute-intensive) και οι εφαρμογές στις οποίες το μεγαλύτερο ποσοστό της δραστηριότητας αναφέρεται στους πόρους εισόδου/εξόδου (I/O-intensive).

Μετά τον ορισμό του νέου χώρου χαρακτηριστικών, F , μέσω των εξισώσεων 4.6 - 4.8, οι επιλεγθείσες εφαρμογές χαρακτηρίζονται με το νέο διάνυσμα αξιολόγησης και εκτελείται ομαδοποίηση, προκειμένου να τοποθετηθούν σε ομάδες οι εφαρμογές, οι οποίες μοιράζονται τα ίδια κυκλωματικά χαρακτηριστικά. Χρησιμοποιήθηκε ο αλγόριθμος ομαδοποίησης k -means [43] για την προαναφερόμενη διαδικασία, με παραμετρικό αριθμό των ομάδων, k . Η διαδικασία ομαδοποίησης κατακεραματίζει το χώρο χαρακτηριστικών F σε k περιοχές ενδιαφέροντος, F_i $i \in \{0 \dots k - 1\}$. Ο αριθμός των περιοχών, k συσχετίζεται σε μεγάλο βαθμό με την αποτελεσματικότητα της διαδικασίας ομαδοποίησης. Η αδρομερής ομαδοποίηση, δηλαδή η χρήση μικρών τιμών του k , αυξάνει το εύρος των χαρακτηριστικών που προσδιορίζουν κάθε ομάδα εφαρμογής, ενώ λεπτομερής ομαδοποίηση, όπως για παράδειγμα η αντιστοίχιση μίας εφαρμογής σε μία μεμονωμένη ομάδα, μειώνει το εύρος των χαρακτηριστικών, αλλά μειώνει επίσης και τη ευκολία συσχέτισης μίας εφαρμογής σε κάποια ομάδα. Στα πλαίσια αυτής της εργασίας, χρησιμοποιήθηκαν οκτώ ομάδες ($k=8$) για το σύνολο των είκοσι εφαρμογών χαρακτηρισμού, ως μία συμβιβαστική λύση μεταξύ αδρομερούς και λεπτομερούς ομαδοποίησης, όπως φαίνεται και στο Σχήμα 4.9.

4.4.2. Εξερεύνηση Χώρου Ομαδοποίησης για τη Βελτιστοποίηση του Χαρακτηρισμού Αξιολόγησης

Ο στόχος της διερεύνησης της διαδικασίας ομαδοποίησης είναι η εύρεση των παραμέτρων διαμόρφωσης του εργαλείου GENESIS, οι οποίες παρέχουν βέλτιστες λύσεις τοποθετήσεων για κάθε σύνολο εφαρμογών που εμφανίζει παρόμοια χαρακτηριστικά με αυτά των ομάδων F_i . Επί του παρόντος, ο αλγόριθμος

Πίνακας 4.3 Χαρακτηρισμός μετρικών για τις επιλεγμένες εφαρμογές αξιολόγησης.

Εφαρμογή	Χώρος Μετρικών Σύνθεσης				Ελ/το μέγεθος FPGA
	#CLBs	#IOs	#Nets	#DFFs	
alu4	1,519	22	1,523	0	39×39
apex2	1,878	42	1,899	0	44×44
apex4	1,262	27	1,264	0	36×36
bigkey	1,707	460	2,046	224	107×107
clma	8,071	33	8,135	33	90×90
des	1,591	501	1,837	0	126×126
diffeq	1,497	103	1,539	377	39×39
dsip	1,370	426	1,599	224	107×107
elliptic	3,604	245	3,731	1122	62×62
ex1010	4,598	20	4,555	0	68×68
ex5p	1,064	71	1,046	0	33×33
frisc	3,556	136	3,552	886	60×60
misex3	1,397	28	1,390	0	38×38
pdc	4,575	56	4,522	0	68×68
s298	1,931	10	1,935	8	44×44
s38417	6,235	135	6,237	1463	79×79
s38584	6,447	343	6,331	1260	81×81
seq	1,750	76	1,774	0	42×42
spla	3,690	62	3,647	0	61×61
tseng	1,047	174	1,096	385	44×44
Μέσος όρος	2,679.42	154.58	2982.9	299.1	64×64

GENESIS εκθέτει τρεις βασικές παραμέτρους που ρυθμίζουν την εξελικτική διαδικασία βελτιστοποίησης. Συγκεκριμένα, αυτές είναι (i) *population age* (pa), (ii) *population size* (ps) και (iii) *aggressiveness of timing over wirelength optimization*, η οποία αντιπροσωπεύεται από την παράμετρο a της συνάρτησης καταλληλότητας. Συγκεκριμένα, όσον αφορά την τελευταία παράμετρο, η ευαισθησία της στην Εξίσωση 4.1 επηρεάζει ιδιαίτερα τις μετρικές QoR του χώρου λύσεων. Παρόλο που υπάρχει συσχέτιση μεταξύ του μήκους των συνδέσεων και της καθυστέρησης, εφόσον το μοντέλο Elmore υπολογίζει το μήκος των συνδέσεων, ωστόσο, με βάση την προκείμενη ανάλυση, η οποία απεικονίζεται στον Πίνακα 4.4, προκύπτει ότι ο υπολογισμός και των δύο αυτών παραμέτρων κατά την τοποθέτηση οδηγεί σε βέλτιστη συχνότητα λειτουργίας. Προκειμένου να προσδιοριστεί η τιμή της παραμέτρου a προτείνεται η ακόλουθη ευρετική ανάλυση.

Ο χώρος διαμόρφωσης Ω ορίζεται ως το σύνολο όλων των πιθανών διανυσμάτων διαμόρφωσης, $U < ps_k, pa_k, a_k > \in \Omega$. Κατά τη φάση παραμετροποίησης του εργαλείου GENESIS, για κάθε μία από τις σχηματισμένες ομάδες, F_i , δημιουργείται ένα διάνυσμα διαμόρφωσης, το οποίο βελτιστοποιεί μία ή περισσότερες μετρικές του σχεδιασμού, π.χ. καθυστέρηση, κατανάλωση ενέργειας, επιφάνεια πυριτίου, χρόνος εκτέλεσης τοποθέτησης κλπ. Αυτά τα βελτιστοποιημένα δια-

Σχήμα 4.9 Ομαδοποίηση στο χώρο χαρακτηριστικών F .

νύσματα χρησιμοποιούνται κατά την φάση εκτέλεσης του εργαλείου GENESIS, για τη ρύθμιση του αλγορίθμου τοποθέτησης, ανάλογα με τα χαρακτηριστικά της εκάστοτε εφαρμογής εισόδου.

Το πρόβλημα της εύρεσης/προσέγγισης βελτιστοποιημένων ομάδων ανά διαμόρφωση διανύσματος χαρακτηριστικών ανάγεται, σε γενικευμένη μορφή, ως ένα πολυ-κριτηριακό πρόβλημα βελτιστοποίησης. Η βελτιστοποίηση ενός μόνο κριτηρίου αποτελεί υποσύνολο του προκειμένου προβλήματος και συνεπώς μπορούν να εφαρμοστούν οι ίδιες τεχνικές με την πολυ-κριτηριακή ανάλυση. Το διάνυσμα των κριτηρίων βελτιστοποίησης ορίζεται ως $\langle LO = \{o_1, \dots, o_N\}$, π.χ. $\langle o_1, o_2, o_3 \rangle = \langle Delay, Power, Runtime \rangle$.

Δεδομένου ότι κάθε ομάδα, F_i , περιλαμβάνει ένα πλήθος εφαρμογών, δηλαδή $app_{i,j} \in F_i$, τότε πολλά στιγμιότυπα των διανυσμάτων κριτηρίων βελτιστοποίησης, $LO^{i,j}$ υπάρχουν μέσα σε κάθε ομάδα F_i , ένα για κάθε $app_{i,j} \in F_i$. Συνεπώς, η βελτιστοποίηση μπορεί να επιτευχθεί σε όλες τις εφαρμογές, $app_{i,j}$, οι οποίες βρίσκονται μέσα σε κάθε ομάδα, προκειμένου να είναι δυνατή η κάλυψη των μετρικών για την επικράτεια των ομάδων F_i . Το πρόβλημα της εύρεσης της βέλτιστης διαμόρφωσης ανά ομάδα, μπορεί να οριστεί ως εξής για την ομάδα F_i :

Πίνακας 4.4 Αξιολόγηση εναλλακτικών ρυθμίσεων του αλγορίθμου GENESIS.

Εφαρμογή	GENESIS	Wirelength-driven (a=0)		Timing-driven (a=1)	
	Καθυστέρηση (Λύση 1)	Καθυστέρηση	Αναλογία (vs GENESIS)	Καθυστέρηση	Αναλογία (vs GENESIS)
alu4	7.38E-008	1.07E-007	1.45×	1.09E-007	1.47×
apex2	8.81E-008	1.39E-007	1.58×	1.14E-007	1.29×
apex4	7.72E-008	1.16E-007	1.50×	9.16E-008	1.19×
bigkey	6.63E-008	2.05E-007	3.09×	9.79E-008	1.48×
clma	1.89E-007	2.57E-007	1.36×	2.11E-007	1.12×
des	1.05E-007	2.86E-007	2.73×	2.60E-007	2.48×
diffeq	6.09E-008	9.73E-008	1.60×	7.42E-008	1.22×
dsip	6.38E-008	1.92E-007	3.00×	1.02E-007	1.60×
elliptic	1.10E-007	2.68E-007	2.43×	1.36E-007	1.24×
ex1010	1.16E-007	1.97E-007	1.69×	1.66E-007	1.43×
ex5p	6.64E-008	1.20E-007	1.80×	9.57E-008	1.44×
frisc	1.27E-007	2.35E-007	1.85×	1.78E-007	1.40×
misex3	7.11E-008	1.10E-007	1.54×	8.58E-008	1.21×
pdc	1.08E-007	2.74E-007	2.54×	1.74E-007	1.61×
s298	1.32E-007	2.32E-007	1.75×	1.41E-007	1.07×
s38417	1.01E-007	1.74E-007	1.72×	1.14E-007	1.13×
s38584	8.7E-008	1.33E-007	1.53×	1.15E-007	1.32×
seq	7.79E-008	1.25E-007	1.60×	1.07E-007	1.37×
spla	8.46E-008	1.90E-007	2.25×	1.51E-007	1.79×
tseng	5.29E-008	8.25E-008	1.56×	6.17E-008	1.17×
Μέσος όρος	9.29E-008	1.77E-007	1.93×	1.29E-007	1.40×

$$\min_{x \in \Omega} \begin{bmatrix} gm(o_1^{i,0}, \dots, o_1^{i,j}, o_1^{i,|F_i|-1})(x) \\ gm(o_2^{i,0}, \dots, o_2^{i,j}, o_2^{i,|F_i|-1})(x) \\ \dots \\ gm(o_N^{i,0}, \dots, o_N^{i,j}, o_N^{i,|F_i|-1})(x) \end{bmatrix} \quad (4.9)$$

, όπου $gm(x_1, \dots, x_i, \dots, x_N) = \sqrt[N]{x_1 \times \dots \times x_i \times \dots \times x_N}$ παριστάνει το γεωμετρικό μέσο όρο του κριτηρίου που είναι συσχετισμένο με την εφαρμογή $app_{i,j}$.

Δεδομένου ότι το πρόβλημα είναι πολυ-κριτηριακό, αναζητείται η εύρεση ενός συνόλου βέλτιστων (Pareto) παραμέτρων, $X_i^{Popt} \in \Omega$. Η βασική ιδέα για αυτή τη προσέγγιση, είναι ότι ο σχεδιαστής μπορεί να επιλέξει/αποφασίσει το διάλυμα παραμέτρων ρύθμισης του εργαλείου GENESIS $x \in X_i^{Popt}$, το οποίο αποδίδει τις βέλτιστες λύσεις τοποθέτησης για διάφορες εφαρμογές που ανήκουν στις ομάδες F_i . Ωστόσο, η εύρεση του ακριβούς χώρου Pareto, για το πρόβλημα βελτιστοποίησης 4.9, είναι μία υπολογιστικά ανέφικτη διαδικασία, δεδομένου ότι απαιτεί την πλήρη αξιολόγηση (συμπεριλαμβανομένων των σταδίων της τοποθέτησης και δρομολόγησης) σε όλες τις διαμορφώσεις που ανήκουν στο χώρο Ω όλων των εφαρμογών που ανήκουν σε κάθε ομάδα F_i .

Για την επίλυση του παραπάνω προβλήματος, αναπτύχθηκε μια στρατηγική εξερεύνησης, η οποία παρουσιάζεται στον Αλγόριθμο 5, προκειμένου να υπολογίσει ένα χώρο Pareto, \hat{X}_i^{Popt} για κάθε ομάδα F_i . Η προτεινόμενη στρατηγική επαναλαμβάνεται επί του συνόλου των καθορισμένων ομάδων. Για κάθε ομάδα,

δειγματοληπτείται ο χώρος διαμόρφωσης Ω ακολουθώντας μια ομοιόμορφη κατανομή πιθανοτήτων, $SampleSpace(F_i) \subseteq \Omega$. Κάθε εφαρμογή, $app_{i,j} \in F_i$ διερευνάται και αξιολογείται, σύμφωνα με τη κοινό δείγμα διαμόρφωσης του χώρου λύσεων, $SampleSpace(F_i)$. Ο παραγόμενος διανυσματικός χώρος, D_j^i , αποτελεί μία απεικόνιση που συνδέει κάθε φορέα διαμόρφωσης ο οποίος βρίσκεται στο $SampleSpace(F_i)$ με τις μετρικές που ανήκουν στο διάνυσμα βελτιστοποίησης, LO . Ο γεωμετρικός μέσος όρος εφαρμόζεται σε κάθε D_j^i , δημιουργώντας το χώρο D^i , ο οποίος αντιπροσωπεύει το πρόβλημα που ορίζεται στην Εξίσωση 4.9.

Ο παραγόμενος διανυσματικός χώρος, D^i παρέχει μια συγκεντρωτική προβολή του χώρου λύσεων, δεδομένου ότι περιλαμβάνει χαρακτηρισμένες λύσεις (όσον αφορά τους στόχους βελτιστοποίησης) μόνο για τα “γνωστά σημεία”, δηλαδή τις διαμορφώσεις, $SampleSpace(F_i) \subseteq \Omega$, που έχουν πραγματικά αξιολογηθεί μέσω του εργαλείου GENESIS. Τα υπόλοιπα διανύσματα διαμόρφωσης του χώρου Ω , τα οποία δεν έχουν αξιολογηθεί, αποτελούν τα “άγνωστο σημεία” του χώρου λύσεων. Αυτά τα “άγνωστη σημεία” του χώρου Ω , προσεγγίζονται με τη χρήση της τεχνικής της παρεμβολής⁶, ενώ τα “γνωστά σημεία” στο χώρο D^i αποτελούν τη βάση εκπαίδευσης των υπολοίπων. Ο χώρος \hat{D}_{total}^i είναι η προβλεπόμενη προσέγγιση του χώρου λύσεων επί όλων των διανυσμάτων διαμόρφωσης του χώρου $|\Omega$, που δημιουργούνται μετά την παρεμβολή. Συγκεκριμένα, υιοθετήθηκε η τεχνική παρεμβολής *inverse distance weighting* (IDW) [44]. Η τιμή ενός αγνώστου σημείου, $x \in F_i$ υπολογίζεται με τη χρήση N γνωστών παρατηρήσεων $y_k \in D^i$ μέσω της ακόλουθης εξίσωσης 4.10 :

$$r(x) = \frac{\sum_{k=1}^N w_k(x)y_k}{\sum_{k=1}^N w_k(x)}, w_k(x) = \frac{1}{dist(x, x_k)^p} \quad (4.10)$$

, όπου $w_k(x)$ είναι η συνάρτηση στάθμισης, $dist$ είναι η απόσταση μεταξύ του γνωστού σημείου x_k και του άγνωστου σημείου x και p είναι ένας πραγματικός αριθμός που ονομάζεται η παράμετρος ισχύος του μοντέλου. Όσο μεγαλύτερη είναι η τιμή του p τόσο μεγαλύτερη είναι η επίδραση των σημείων που βρίσκονται πλησιέστερα προς το σημείο παρεμβολής. Για διαμόρφωση χώρου N -διαστάσεων, υπάρχει ένα θεωρητικό κατώτερο όριο των παραμέτρων ισχύος, $p \geq \ln N$, προκειμένου να περιοριστεί το φαινόμενο οι τιμές παρεμβολής να επηρεάζονται από απομακρυσμένα σημεία. Έτσι, για τον τρισδιάστατο χώρο διαμόρφωσης, $\langle ps_k, pa_k, a_k \rangle \in \Omega$, θεωρείται μια σειρά από $p \in [3, 6]$ με βήμα 0.5. Μέσω επαναληπτικής αξιολόγησης, η τιμή της παραμέτρου p είναι αυτή που μεγιστοποιεί την ακρίβεια του μοντέλου IDW. Τέλος, ο προσεγγιστικός χώρος λύσης, \hat{D}_{total}^i , φιλτράρεται σύμφωνα με τη το σύνολο των κριτηρίων βελτιστοποίησης, LO προκειμένου δημιουργηθεί ο αντίστοιχος χώρος Pareto.

Το Σχήμα 4.10 δείχνει τις καμπύλες Pareto χώρων για κάθε ομάδα F_i που δημιουργείται από την προτεινόμενη προσέγγιση εξερεύνησης. Στη συγκεκριμένη εργασία, έγινε διερεύνηση των παραμέτρων του εργαλείου GENESIS σύμφωνα με δύο στόχους βελτιστοποίησης, την χρονική καθυστέρηση σε επίπεδο τοποθε-

⁶Σε αντίθεση με τις τεχνικές παλινδρόμησης (regression), οι τεχνικές παρεμβολής δεν παράγουν σφάλματα πρόβλεψης για τα γνωστά δεδομένα.

τημένους και δρομολογημένους κυκλώματος και το χρόνο εκτέλεσης του εργαλείου για τη διεξαγωγή των διαδικασιών τοποθέτησης και δρομολόγησης. Ο σχεδιαστής είναι υπεύθυνος για την επιλογή ενός συγκεκριμένου Pareto χώρου διαμόρφωσης ανά ομάδα, ο οποίος πρέπει να χρησιμοποιείται κατά τη διάρκεια της εκτέλεσης του εργαλείου GENESIS, με γνώμονα τη συσχέτιση των ειδικών χαρακτηριστικών της υπό-σχεδίασης εφαρμογής με αυτή τη συγκεκριμένη ομάδα.

ΑΛΓΟΡΙΘΜΟΣ 5: Εξερεύνηση Προσέγγισης Βέλτιστης Διαμόρφωσης ανά Ομάδα

Input: Parameter space of GENESIS optimizer: Ω

Input: Cluster set: $F_i, i \in \{0, \dots, k-1\}$

Input: Max. number of per cluster evaluations: M_E

Input: List of optimization objectives: $LO = \{o_1, \dots, o_N\}$

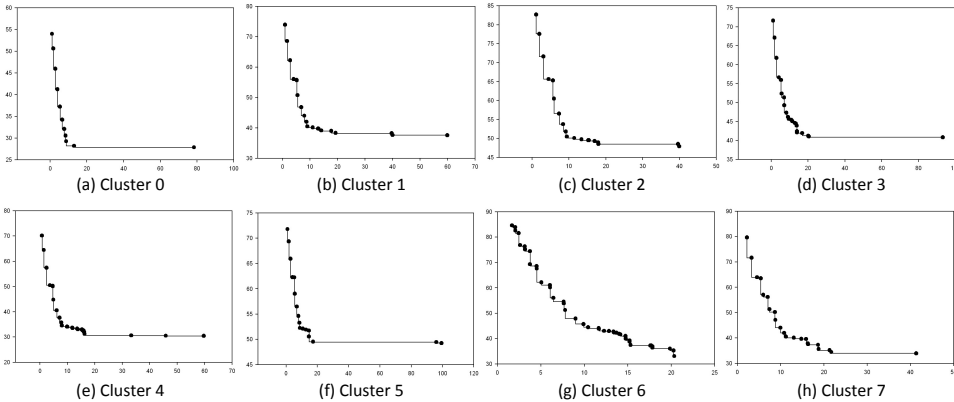
Output: Optimal per cluster parameter configuration vector: $\hat{X}_{F_i}^{Opt} \in \Omega$

```

foreach  $F_i, i \in \{0, \dots, k-1\}$  do
   $SampleSpace(F_i) \leftarrow rand(\Omega, M_E)$ 
  foreach  $app_{i,j} \in F_i$  do
     $D_j^i \leftarrow explore(app_{i,j}, SampleSpace(F_i))$ 
  end
   $D^i \leftarrow gmean(D_0^i, \dots, D_j^i, \dots, D_{|F_i|-1}^i)$ 
   $\hat{D}_{total}^i \leftarrow interpolate(D^i, \Omega)$ 
   $\hat{X}_{F_i}^{Opt} \leftarrow ParetoFrontGeneration(\hat{D}_{total}^i, LO)$ 
end

```

end



Σχήμα 4.10 Ο χώρος Pareto για την χρονική καθυστέρηση του κυκλώματος σε σχέση με το χρόνο εκτέλεσης του εργαλείου για κάθε ομάδα F_i . Τα δεδομένα είναι κανονικοποιημένα αναφορικά προς την λιγότερο αποτελεσματική λύση ανά ομάδα.

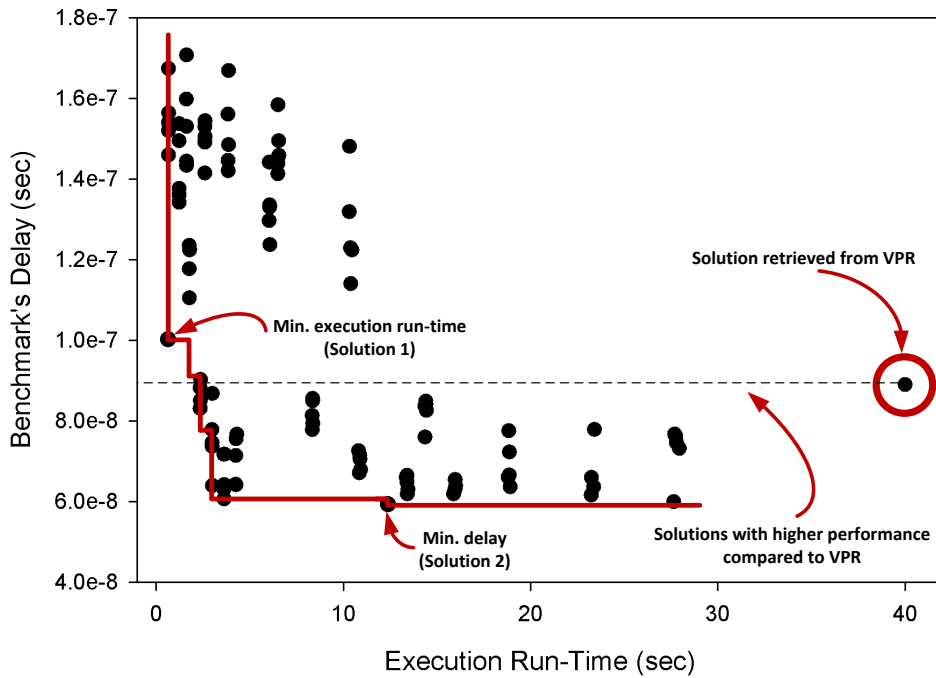
4.5. Πειραματικά Αποτελέσματα

Αυτή η ενότητα παρέχει μια σειρά από ποσοτικά αποτελέσματα που αποδεικνύουν την αποτελεσματικότητα της προτεινόμενης λύσης, σε σύγκριση με τον αλγόριθμο τοποθέτησης του εργαλείου VPR 4.30 [15]. Για το σκοπό αυτό, χρησι-

μοποιούνται οι είκοσι μεγαλύτερες εφαρμογές του πακέτου αξιολόγησης MCNC, για το πρόβλημα της τοποθέτησης, ενώ η φυσική υλοποίηση προορίζεται για ένα FPGA με αρχιτεκτονική island-style [45]. Σημειώνεται ότι αυτού του τύπου αρχιτεκτονική είναι η διαδεδομένη εμπορικά και συναντάται σε FPGAs διαφόρων οικογενειών/κατασκευαστών, όπως τα Xilinx Virtex και τα Altera Stratix FPGAs. Επίσης, η στρατηγική εξερεύνησης του χαρακτηρισμού ομαδοποίησης έχει εφαρμοστεί με το πλαίσιο Multicube [46]. Όσον αφορά το υπολογιστικό σύστημα λήψης μετρήσεων, χρησιμοποιήθηκε ένας επεξεργαστή Intel Xeon, οκτώ πυρήνων, με συχνότητα επεξεργασίας 2.33GHz και 4GB φυσικής μνήμης RAM. Ο Πίνακας 4.3 συνοψίζει τις ιδιότητες των εφαρμογών αξιολόγησης, καθώς και το ελάχιστο μέγεθος του FPGA, στο οποίο μπορεί να τοποθετηθεί η εκάστοτε εφαρμογή. Στο σύνολο των πειραματικών αποτελεσμάτων, αναφέρονται δύο διακριτές διαμορφώσεις του γενετικού αλγορίθμου. Συγκεκριμένα, η πρώτη διαμόρφωση, η οποία αναφέρεται ως “Λύση 1” (Solution 1), αντιστοιχεί στη ρύθμιση του αλγορίθμου προκειμένου να επιτύχει γρήγορη εκτέλεση, ενώ η δεύτερη περίπτωση, η οποία αναφέρεται ως “Λύση 2” (Solution 2), αντιστοιχεί στη ρύθμιση του αλγορίθμου προκειμένου να επιτύχει ποιοτικότερο αποτέλεσμα τοποθέτησης, που για αυτό το στάδιο αξιολόγησης, αντιστοιχεί σε επίτευξη μέγιστης συχνότητα λειτουργίας.

Δεδομένου ότι η ποιότητα των λύσεων που παρέχει ο προτεινόμενος αλγόριθμος εξαρτάται σε μεγάλο βαθμό από τις παραμέτρους του, κατά τη διάρκεια της αξιολόγησης ελέγχθησαν εναλλακτικές ρυθμίσεις του αλγορίθμου. Για παράδειγμα, το Σχήμα 4.11 απεικονίζει την επίπτωση του μεγέθους του πληθυσμού και των γενεών εξέλιξης στην χρονική καθυστέρηση της εφαρμογής και στον χρόνο εκτέλεσης, για την τοποθέτηση της εφαρμογής αξιολόγησης *bigkey*. Σε αυτό το σχήμα επιπλέον απεικονίζεται η αντίστοιχη λύση που ανακτάται με τη χρήση του εργαλείου VPR[15]. Με βάση αυτό το σχήμα συμπεραίνεται ότι σχεδόν οι μισές από τις διαθέσιμες λύσεων μπορούν να λειτουργήσουν σε υψηλότερες συχνότητες (δηλαδή με μικρότερη καθυστέρηση στο κρίσιμο μονοπάτι) σε σχέση με τις αντίστοιχες τοποθετήσεις που ανακτούνται με το VPR, ενώ η μέγιστη βελτίωση της απόδοσης φτάνει το ποσοστό του 40%.

Ένα από τα πιο σημαντικά χαρακτηριστικά του προτεινόμενου αλγορίθμου, σε σύγκριση με τους υφιστάμενους, είναι η εγγενής κλιμάκωσή του στα πολυ-επεξεργαστικά συστήματα. Όπως αναφέρθηκε στην υποενότητα 4.3.2, ο αλγόριθμος GENESIS αναπτύχθηκε με τη χρήση του προγραμματιστικού πλαισίου OpenMP, προκειμένου να επιτύχει μεγάλους βαθμούς επιτάχυνσης. Ο Πίνακας 4.5 ποσοτικοποιεί το χρόνο εκτέλεσης, τη μέγιστη επιτευχθείσα συχνότητα λειτουργίας, καθώς και την κατανάλωση ενέργειας για το πρόβλημα της τοποθέτησης, αναφορικά προς το εργαλείο VPR και χρησιμοποιώντας και τις δύο διαμορφώσεις του προτεινόμενου αλγορίθμου, Λύση 1, Λύση 2 (Solution 1, Solution 2).



Σχήμα 4.11 Αξιολόγηση για διαφορετικούς συνδυασμούς των παραμέτρων *population size* και *population age* για την εφαρμογή *bigkey*.

Πίνακας 4.5 Σύγκριση του χρόνου εκτέλεσης, της μέγιστη επιτευχθείσας συχνότητα λειτουργίας και της κατανάλωσης ενέργειας για τους αλγόριθμους τοποθέτησης GENESIS και VPR.

Εφαρμογή	Χρόνος Εκτέλεσης (second)			Περίοδος (second)			Κατανάλωση Ενέργειας (Watt)		
	VPR	GENESIS		VPR	GENESIS		VPR	GENESIS	
		Solution 1	Solution 2		Solution 1	Solution 2		Solution 1	Solution 2
alu4	25.00	1.587	0.386	7.79E-08	7.38E-08	1.30E-07	0.1841	0.1889	0.1704
apex2	34.00	2.288	0.562	8.63E-08	8.81E-08	1.81E-07	0.2175	0.2170	0.1927
apex4	21.00	1.276	0.317	7.33E-08	7.72E-08	1.41E-07	0.1478	0.1455	0.1289
bigkey	56.00	2.162	0.714	8.96E-08	6.63E-08	1.54E-07	1.1096	1.1879	1.1004
clma	372.00	23.121	4.938	2.06E-07	1.89E-07	4.65E-07	0.7684	0.7800	0.7055
des	50.00	2.446	0.910	1.47E-07	1.05E-07	3.09E-07	1.4125	1.4293	1.3928
diffeq	25.00	1.711	0.427	6.48E-08	6.09E-08	1.79E-07	0.1695	0.1738	0.1460
dsip	45.00	1.703	0.639	9.52E-08	6.38E-08	1.51E-07	1.0896	1.1670	1.0790
elliptic	98.00	6.527	1.483	1.13E-07	1.10E-07	2.61E-07	0.3891	0.3891	0.3520
ex1010	127.00	7.724	1.918	1.98E-07	1.16E-07	2.70E-07	0.4214	0.4351	0.3856
ex5p	18.00	1.071	0.265	6.60E-08	6.64E-08	1.30E-07	0.1381	0.1371	0.1180
frisc	94.00	6.240	1.422	1.44E-07	1.27E-07	3.24E-07	0.3314	0.3391	0.3004
misex3	22.00	1.441	0.356	7.71E-08	7.11E-08	1.14E-07	0.1728	0.1777	0.1594
pdcc	134.00	8.844	2.020	1.42E-07	1.08E-07	2.27E-07	0.4374	0.4460	0.3910
s298	33.00	2.078	0.495	1.31E-07	1.32E-07	2.78E-07	0.1989	0.1992	0.1798
s38417	189.00	16.385	3.212	1.02E-07	1.01E-07	2.00E-07	0.6902	0.4873	0.6600
s38584	204.00	16.697	3.223	8.95E-08	8.70E-08	1.34E-07	0.7070	0.5169	0.7241
seq	32.00	2.078	0.507	6.98E-08	7.79E-08	1.08E-07	0.2135	0.2064	0.1897
spla	94.00	5.861	1.219	1.24E-07	8.46E-08	1.33E-07	0.3529	0.3633	0.3238
tseng	17.00	0.983	0.124	5.53E-08	5.29E-08	1.60E-07	0.2141	0.2182	0.1862
Μέσος Όρος	84.50	5.611	1.257	1.08E-07	9.28E-08	2.03E-07	0.4683	0.4602	0.4443
Αναλογία	1.00	15.06×	67.24×	1.00	1.16×	0.53×	1.00	1.02×	1.05×

Όπως φαίνεται στον Πίνακα 4.5, η προτεινόμενη προσέγγιση επιτυγχάνει κατά μέσο όρο 15x και 67x ταχύτερες λύσεις, σε σύγκριση με το VPR, ενώ για συγκεκριμένες εφαρμογές (π.χ. *tseng*) η αναλογία επιτάχυνσης μπορεί να είναι έως και 137x. Επιπλέον, η παρούσα υλοποίηση του προτεινόμενου αλγόριθμου υπερτερεί έναντι της προηγούμενης έκδοσής του, η οποία παρουσιάζεται στην εργασία [18], δεδομένου ότι η προηγούμενη έκδοση είχε επιτύχει μέση επιτάχυνση της εκτέλεσης σε σχέση με το VPR της τάξης του 6,6x. Στο σημείο αυτό τονίζεται ότι η επιτάχυνση της εκτέλεσης δεν είναι σταθερή μεταξύ των εφαρμογών αξιολόγησης, διότι εξαρτάται από τις εγγενείς ιδιότητες των εφαρμογών (π.χ. ο αριθμός των εισόδων/εξόδων και των λογικών μονάδων, τα λογικά στάδια του κρίσιμου μονοπατιού, το fanout κ.α.). Το εργαλείο σχεδίασης GENESIS παρέχει στον σχεδιαστή την ελευθερία ρύθμισης των παραμέτρων με χειροκίνητο τρόπο, έτσι ώστε να είναι δυνατή η λεπτομερής ρύθμιση του αλγορίθμου βελτιστοποίησης αναφορικά προς την εφαρμογή και τις απαιτήσεις του περιβάλλοντος υλοποίησης (π.χ. χαμηλή κατανάλωση ισχύος, χαμηλό κόστος σχεδιασμού κτλ.). Ωστόσο, αναφορικά με τα αποτελέσματα που συγκεντρώνονται στον Πίνακα 4.5, οι παράμετροι του αλγορίθμου GENESIS έχουν επιλεγθεί ύστερα από διερεύνηση του χώρου λύσεων ανά εφαρμογή, με τη μεθοδολογία που περιγράφεται στην υποενότητα 4.4.

Μια δεύτερη σημαντική παρατήρηση για τα αποτελέσματα του Πίνακα 4.5 είναι ότι ταυτόχρονα με την επιτευχθείσα επιτάχυνση του χρόνου εκτέλεσης για τη λύση του προβλήματος της τοποθέτησης, ο αλγόριθμος οδηγεί επίσης και σε αποτελεσματικότερη διερεύνηση του χώρου λύσεων και συνεπώς καταφέρνει να ανακαλύψει λύσεις με κέρδη στη μέγιστη επιτευχθείσα συχνότητα λειτουργίας και στην κατανάλωση ενέργειας. Αναλυτικότερα, η αλγοριθμική διαμόρφωση που αντιστοιχεί στη “Λύση 1” επιτυγχάνει τη βελτίωση της μέγιστης συχνότητας λειτουργίας (δηλαδή μείωση της καθυστέρησης του κρίσιμου μονοπατιού) κατά ένα παράγοντα 1.16x, ενώ η “Λύση 2” παρουσιάζει μείωση της μέσης μέγιστης συχνότητας λειτουργίας κατά 0,53x, σε σύγκριση με τις αντίστοιχες λύσεις του εργαλείου VPR. Για την περίπτωση της “Λύσης 1”, διευκρινίζεται ότι τα κέρδη είναι συμπληρωματικά προς τους παράγοντες επιτάχυνσης που αναφέρθηκαν προηγουμένως αναφορικά με το χρόνο εκτέλεση του αλγορίθμου GENESIS. Επίσης σχετικά με την “Λύση 2”, είναι φανερό το εύρος συμβιβασμού (trade-off) που μπορεί να επιτευχθεί με μία υπερ-ταχεία εκτέλεση του αλγορίθμου (Πίνακας 4.5, στήλη 4) σε σχέση με την ποιότητα της προσφερόμενης λύσης (Πίνακας 4.5, στήλη 7).

Επίσης, όπως απεικονίζεται και στον Πίνακα 4.5, το προτεινόμενο πλαίσιο δεν επιβαρύνει, σε μεγάλο βαθμό, τους τελικούς σχεδιασμούς με επιπλέον κατανάλωση ενέργειας, συγκριτικά με τις λύσεις του VPR. Συγκεκριμένα, το προτεινόμενο πλαίσιο οδηγεί, κατά μέσο όρο, σε πολύ μικρή αύξηση της κατανάλωσης ισχύος κατά 1.05x. Σημειώνεται ότι η τιμή της κατανάλωσης υπολογίζεται βάση της μέγιστης επιτευχθείσας συχνότητας λειτουργίας και άρα είναι απολύτως λογική αυτή η αύξηση, καθώς η σχέση μεταξύ των δύο μετρικών είναι γραμμικά ανάλογη. Προκειμένου να εξασφαλιστεί μία δίκαιη σύγκριση ανάμεσα στα δύο εργαλεία, έγινε ρύθμιση του εργαλείου GENESI προκειμένου να επιτύχει μέγι-

στη συχνότητα λειτουργίας ίση με την συχνότητα των λύσεων του VPR. Σε αυτή τη περίπτωση, η αντίστοιχη μέση εξοικονόμηση ενέργειας (“Λύση 2”) είναι κατά μέσο όρο 6%.

Η προηγούμενη ανάλυση τονίζει ότι η προτεινόμενη προσέγγιση παρουσιάζει ταχύτερη εκτέλεση του προβλήματος της τοποθέτησης, σε σύγκριση με τα υφιστάμενα ανταγωνιστικά εργαλεία σχεδίασης (π.χ. VPR), ενώ ταυτόχρονα οδηγεί και σε βελτιστοποιημένες λύσεις τοποθέτησης αναφορικά προς τις μετρικές QoR. Η επιτυχία αυτή αποδίδεται στο γεγονός ότι ο προτεινόμενος αλγόριθμος εξερευνά πιο αποτελεσματικά το χώρο λύσεων και επομένως μειώνει την πιθανότητα παγίδευσης σε τοπικά ελάχιστα, σε σύγκριση με βαθμωτές λύσεις βελτιστοποίησης, όπως η προσομοιωμένη απόπτηση. Προκειμένου να ποσοτικοποιηθεί αυτή η διαφοροποίηση, το Σχήμα 4.12 απεικονίζει τον αριθμό των διαφορετικών ενδιάμεσων λύσεων που διερευνώνται κατά τη διάρκεια εκτέλεσης των εργαλείων GENESIS και VPR. Αναφέρεται ότι ο κατακόρυφος άξονας σε αυτή την εικόνα είναι συνεπτυγμένος σε λογαριθμική κλίμακα. Τα αποτελέσματα που συνοψίζονται σε αυτό το σχήμα δείχνουν ότι οι λύσεις “Λύση 1” και “Λύση 2”, του προτεινόμενου αλγορίθμου, διερευνούν και αξιολογούν κατά μέσο όρο $157\times$ και $35\times$ περισσότερες λύσεις αντίστοιχα, σε σύγκριση με το εργαλείο VPR.

Ακολούθως, ποσοτικοποιείται η απόδοση της προτεινόμενης προσέγγισης αναφορικά με την επιτάχυνση του χρόνου εκτέλεσης, αξιοποιώντας τους σύγχρονους πολυπύρηνους επεξεργαστές. Το Σχήμα 4.13 απεικονίζει την επιτευχθείσα επιτάχυνση από την εκτέλεση του αλγορίθμου σε επεξεργαστές τεσσάρων και οκτώ πυρήνων, για διαφορετικούς αριθμούς νημάτων εκτέλεσης. Για λόγους ορθής απεικόνισης, όλα τα αποτελέσματα είναι κανονικοποιημένο επί των αντίστοιχων μετρήσεων για συστήματα ενός επεξεργαστικού πυρήνα και ενός νήματος εκτέλεσης. Επιπλέον, για λόγους πληρότητας απεικονίζεται επίσης η αντίστοιχη θεωρητική καμπύλη επιτάχυνσης, όπως ανακτάται από το μοντέλο επιτάχυνσης του νόμου του Amdahl [47]. Η καμπύλη αυτή αντιστοιχεί στο θεωρητικό όριο επιτάχυνσης, δεδομένου ότι το παραλληλοποιημένο ποσοστό του πηγαιού κώδικα του εργαλείου GENESIS είναι 81%, με βάση τα αποτελέσματα ανάλυσης της κατανομής εκτέλεσης (profiling).

Με βάση το παραπάνω διάγραμμα, συμπεραίνεται ότι τα πειραματικά αποτελέσματα και για τις δύο καμπύλες (τετραπύρηνα και οκταπύρηνα συστήματα) συμβαδίζουν με τα αναμενόμενα θεωρητικά όρια. Αυτό σημαίνει ότι ο προτεινόμενος αλγόριθμος εκμεταλλεύεται τη μέγιστη δυνατή αύξηση παραλληλίας του επεξεργαστή. Επίσης, επιβεβαιώνεται ότι η μέγιστη επιτάχυνση του αλγορίθμου επιτυγχάνεται όταν ο αριθμός των νημάτων είναι ίσος με τον αριθμό των φυσικών πυρήνων επεξεργασίας, ενώ επαληθεύεται ότι μετά από αυτό το σημείο, επέρχεται κορεσμός στην επεξεργασία. Τέλος, παρατηρείται ότι η απόκλιση μεταξύ της πραγματικής και της θεωρητικής κλιμάκωσης για το οκταπύρηνο σύστημα είναι υψηλότερη από την αντίστοιχη για το τετραπύρηνο. Μία ανάλυση για αυτή την παρατήρηση εδράζει στο γεγονός ότι το οκταπύρηνο σύστημα στο οποίο πάρθηκαν οι μετρήσεις, αποτελείτο από δύο διακριτούς επεξεργαστές των τεσσάρων πυρήνων ο καθένας, κατανομημένοι σε ξεχωριστά

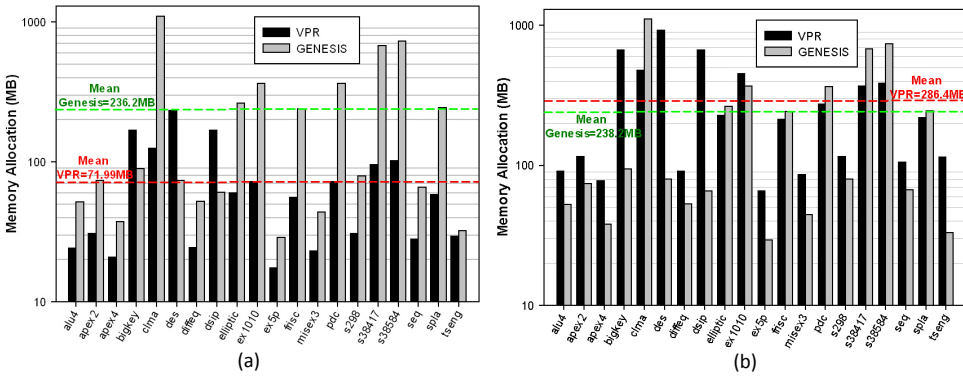
Σχήμα 4.12 Υποψήφιες λύσεις που αξιολογούνται κατά τα στάδια εκτέλεσης των εργαλείων GENESIS και VPR.

ολοκληρωμένα κυκλώματα. Συνεπώς το περιβάλλον προγραμματισμό και δυναμικής διαχείρισης των νημάτων (OpenMP) είναι ρυθμισμένο για οκτώ εικονικούς πυρήνες, ισάξιου επεξεργαστικού και επικοινωνιακού χειριστικού κόστους. Ωστόσο στο φυσικό επίπεδο, υπάρχει μεγάλη διαφορά στην καθυστέρηση δια-νηματικής επικοινωνίας, στην περίπτωση που τα νήματα έχουν ανατεθεί σε πυρήνες του ίδιου ή διαφορετικού επεξεργαστή, καθώς η επικοινωνία στην τελευταία περίπτωση διέρχεται εκτός του ολοκληρωμένου (π.χ. είτε μέσω του διαύλου FSB/HyperTransport/QPI, είτε μέσω του northbridge, του συστήματος) και είναι αρκετά περιορισμένης απόδοσης σε σχέση με την ενδοεπικοινωνία στο ολοκληρωμένο κύκλωμα.

Ως επιπλέον μετρική αξιολόγησης, χρησιμοποιείται το αποτύπωμα μνήμης (memory footprint) κατά τη διάρκεια εκτέλεσης του αλγορίθμου. Καθώς κλιμακώνονται συνεχώς, οι εφαρμογές σε μέγεθος και αντίστοιχα τα FPGA σε λογικούς πόρους, το μέγεθος της μνήμης που απαιτείται για την εκτέλεση των σχεδιαστικών εργαλείων πρέπει να ακολουθεί μια αναλογική κλιμάκωση, η οποία όμως είναι διαχειρίσιμη δεδομένων των διαθέσιμων μεγεθών μνήμης της εκάστοτε αγοράς και του αντίστοιχου κόστους εξοπλισμού. Από τις σχετικές μετρήσεις για το εργαλείο VPR, διαπιστώθηκε ότι δεσμεύει πόρους μνήμης από το σύστημα ανάλογα με το μέγεθος της υποκείμενης συσκευής FPGA. Ωστόσο,

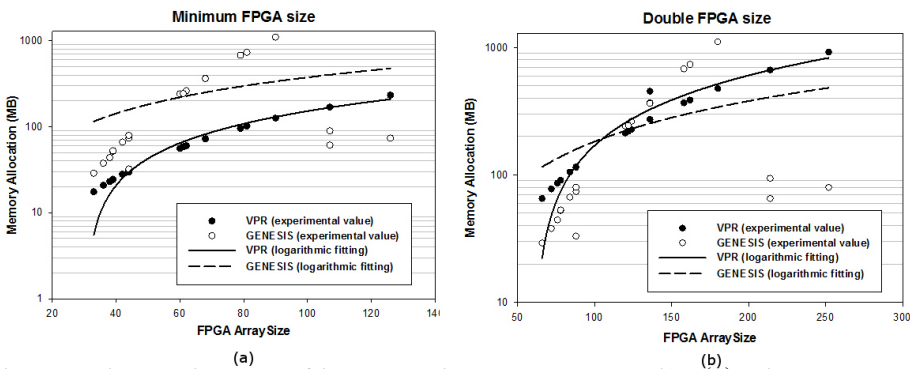
Σχήμα 4.13 Κλιμάκωση της επιτάχυνσης εκτέλεσης του εργαλείου GENESIS, προς τον αριθμό των φυσικών πυρήνων και των νημάτων επεξεργασίας.

αυτή η προσέγγιση οδηγεί στη δέσμευση του ίδιου μεγέθους μνήμης, για διαφορετικές εφαρμογές, όταν επιλέγεται το ίδιο FPGA, ανεξάρτητα του μεγέθους του προβλήματος επίλυσης. Αντίθετα το προτεινόμενο εργαλείο GENESIS δεσμεύει μνήμη ανάλογη του μεγέθους της εφαρμογής και πιο συγκεκριμένα, ανάλογη του αριθμού των κόμβων του γράφου εκτέλεσης της εφαρμογής και επίσης ανεξάρτητη του μεγέθους του FPGA. Γενικά μελετήθηκαν δύο διαφορετικά σενάρια προκειμένου να αξιολογηθεί η επίδραση του μεγέθους δέσμευσης μνήμης από τα δύο εργαλεία, GENESIS και VPR. Συγκεκριμένα, το σενάριο 1, το οποίο απεικονίζεται στο Σχήμα 4.14(a) αντιστοιχεί στην τοποθέτηση της εφαρμογής στο FPGA, με το ελάχιστο μέγεθος πόρων στο οποίο μπορεί η εκάστοτε εφαρμογή να τοποθετηθεί (απεικονίζεται στην τελευταία στήλη του Πίνακα 4.3). Ως δεύτερο σενάριο εξέτασης (σενάριο 2) επιλέχθηκε η τοποθέτηση σε FPGAs με διπλάσιο μέγεθος από αυτο που χρησιμοποιείται στο σενάριο 1 και τα αποτελέσματα απεικονίζονται στο Σχήμα 4.14(b). Το σενάριο 2 επιλέχθηκε προκειμένου να ποσοτικοποιήσει την κλιμακούμενη συμπεριφορά του εργαλείου GENESIS σε σχέση με το μέγεθος του FPGA. Για το σενάριο 1, στην περίπτωση των FPGA μικρού μεγέθους, ο μέσος όρος δέσμευσης μνήμης του εργαλείου GENESIS είναι περίπου 236MB, και είναι υψηλότερος σε σχέση με τη μνήμη που απαιτείται από το εργαλείο VPR με μέσο όρο περίπου 72MB. Ωστόσο, στην περίπτωση των



Σχήμα 4.14 Δέσμευση φυσικής μνήμης για (a) ελάχιστο μέγεθος FPGA και (b) διπλάσιο μέγεθος FPGA, αναφορικά προς τις απαιτήσεις της εφαρμογής.

FPGAs μεγαλύτερου μεγέθους, η δέσμευση μνήμης του εργαλείου GENESIS παραμένει σχεδόν σταθερή, δηλαδή κατά μέσο όρο 238 MB, ενώ για το εργαλείο VPR η μέση δέσμευση μνήμης αυξάνεται στα 286MB. Συνεπώς, διπλασιάζοντας το μέγεθος του FPGA, η προτεινόμενη λύση επιβάλλει μία επιβάρυνση της τάξεως του 8%, ενώ το εργαλείο VPR μία αντίστοιχη αύξηση 398%. Το Σχήμα 4.15 δείχνει την τάση κλιμάκωσης της δέσμευσης μνήμης αναφορικά με το μέγεθος FPGA, τόσο για το σενάριο του ελάχιστου μεγέθους FPGA (Fig. 4.15(a)), όσο και του διπλάσιου μεγέθους αυτού (Fig. 4.15(b)). Είναι προφανές ότι το προτεινόμενο εργαλείο GENESIS παρουσιάζει σταθερή κλιμάκωση, σε σύγκριση με το εργαλείο VPR, ειδικά στην περίπτωση που οι ελεύθεροι πόροι του αυξάνονται.



Σχήμα 4.15 Τάση κλιμάκωσης της δέσμευσης μνήμης για FPGA με μέγεθος (α) ελάχιστο αναφορικά με το μέγεθος της εφαρμογής και (β) διπλάσιο αναφορικά με το μέγεθος της εφαρμογής.

4.6. Γενικά Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάστηκε ο αλγόριθμος GENESIS, μία αποτελεσματική και ταχεία λύση στο πρόβλημα τοποθέτησης για επαναδιαμορφούμενες αρχιτεκτονικές FPGA. Ο προτεινόμενος αλγόριθμος χρησιμοποιεί τους γενετικούς αλγόριθμους, επεκτείνοντας τους με την έννοια της αναμιξιμότητας, προκειμένου να βελτιωθεί η ποιότητα των παραγομένων λύσεων τοποθέτησης. Έγινε ανάλυση στις εσωτερικές δομές και στους τελεστές εξέλιξης που αναπτύχθηκαν στο εργαλείο GENESIS για την προσαρμογή του γενετικού αλγορίθμου στο πρόβλημα της τοποθέτησης των FPGAs. Επιπλέον, χρησιμοποιήθηκε μία τεχνική βασισμένη στο μοντέλο διαίρει-και-βασίλευε, για την αποτελεσματική αδρομερή υποστήριξη παραλληλισμού στο προτεινόμενο εργαλείο. Ο παραλληλισμός μπορεί να ωφελήσει σημαντικά την επιτάχυνση του προβλήματος τοποθέτησης, δεδομένων των σύγχρονων επεξεργαστών πολλαπλών πυρήνων, επιτρέποντας έτσι στους σχεδιαστές τη μελέτη συμβιβασμών μεταξύ της βελτίωσης της ποιότητας σε σχέση με το χρόνο εκτέλεσης του εργαλείου. Για τη βελτίωση της ποιότητας λύσεων, το προτεινόμενο μεθοδολογικό πλαίσιο χρησιμοποιεί ένα στάδιο διερεύνησης των παραμέτρων του αλγορίθμου GENESIS, προσαρμοσμένες στην εκάστοτε εφαρμογή. Η ποιότητα των παραγόμενων σχεδιαστικών λύσεων, καθώς και η αποτελεσματικότητα του προτεινόμενου πλαισίου έχουν αξιολογηθεί μέσα από εκτενή πειραματισμό αξιολόγησης. Τα πειραματικά αποτελέσματα επαλήθευσαν την υπεροχή του εργαλείου GENESIS σε σχέση με το αποδοτικότερο υφιστάμενο εργαλείο τοποθέτησης VPR, προσφέροντας λύσεις οι οποίες, είτε διερευνώνται ταχύτερα κατά ένα παράγοντα έως 67x, είτε είναι ποιοτικότερες αναφορικά με την επιτευχθείσα συχνότητα λειτουργίας κατά 1,16x, ενώ ταυτόχρονα διατηρούν σταθερή κλιμάκωση όσον αφορά τη δέσμευση μνήμης και τις υπολογιστικές απαιτήσεις πόρων.

Βιβλιογραφία

- [1] X. X. Journal, *Xilinx xcell journal issue 86, may 2014*, .
- [2] ITRS, *International technology roadmap for semiconductors*, (2012).
- [3] C. C. Wang and G. G. Lemieux, *Scalable and deterministic timing-driven parallel placement for fpgas*, in *Proceedings of the 19th ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, FPGA '11 (ACM, New York, NY, USA, 2011) pp. 153–162.
- [4] H. Sidiropoulos, K. Siozios, P. Figuli, D. Soudris, and M. Hubner, *On supporting efficient partial reconfiguration with just-in-time compilation*, in *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2012 IEEE 26th International* (2012) pp. 328–335.
- [5] R. Tessier, *Fast placement approaches for fpgas*, *ACM Trans. Des. Autom. Electron. Syst.* **7**, 284 (2002).
- [6] J. M. Emmert and D. Bhatia, *Tabu search: Ultra-fast placement for fpgas*, in *Proceedings of the 9th International Workshop on Field-Programmable Logic and Applications*, FPL '99 (Springer-Verlag, London, UK, UK, 1999) pp. 81–90.
- [7] A. Casotto, F. Romeo, and A. Sangiovanni-Vincentelli, *A parallel simulated annealing algorithm for the placement of macro-cells*, *Computer-Aided Design of Integrated Circuits and Systems*, *IEEE Transactions on* **6**, 838 (1987).
- [8] A. Choong, R. Beidas, and J. Zhu, *Parallelizing simulated annealing-based placement using gpgpu*, in *Field Programmable Logic and Applications (FPL), 2010 International Conference on* (2010) pp. 31–34.
- [9] J. Rose, W. Snelgrove, and Z. Vranesic, *Parallel standard cell placement algorithms with quality equivalent to simulated annealing*, *Computer-Aided Design of Integrated Circuits and Systems*, *IEEE Transactions on* **7**, 387 (1988).
- [10] A. Ludwin, V. Betz, and K. Padalia, *High-quality, deterministic parallel placement for fpgas on commodity hardware*, in *Proceedings of the 16th International ACM/SIGDA Symposium on Field Programmable Gate Arrays*, FPGA '08 (ACM, New York, NY, USA, 2008) pp. 14–23.
- [11] S. Kravitz and R. Rutenbar, *Placement by simulated annealing on a multiprocessor*, *Computer-Aided Design of Integrated Circuits and Systems*, *IEEE Transactions on* **6**, 534 (1987).
- [12] M. G. Wrighton and A. M. DeHon, *Hardware-assisted simulated annealing with application for fast fpga placement*, in *Proceedings of the 2003 ACM/SIGDA eleventh international symposium on Field programmable gate arrays* (ACM, 2003) pp. 33–42.

- [13] E. Witte, R. Chamberlain, and M. Franklin, *Parallel simulated annealing using speculative computation*, *Parallel and Distributed Systems, IEEE Transactions on* **2**, 483 (1991).
- [14] G. Smecher, S. Wilton, and G. Lemieux, *Self-hosted placement for massively parallel processor arrays*, in *Field-Programmable Technology, 2009. FPT 2009. International Conference on* (2009) pp. 159–166.
- [15] V. Betz and J. Rose, *Vpr: A new packing, placement and routing tool for fpga research*, in *Proceedings of the 7th International Workshop on Field-Programmable Logic and Applications, FPL '97* (Springer-Verlag, London, UK, UK, 1997) pp. 213–222.
- [16] V. Betz, J. Rose, and A. Marquardt, eds., *Architecture and CAD for Deep-Submicron FPGAs* (Kluwer Academic Publishers, Norwell, MA, USA, 1999).
- [17] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, *Convergence and finite-time behavior of simulated annealing*, in *Decision and Control, 1985 24th IEEE Conference on* (1985) pp. 761–767.
- [18] D. Diamantopoulos, K. Siozios, S. Xydis, and D. Soudris, *A framework for supporting parallel application placement onto reconfigurable platforms*, in *Proceedings of the PARMA Workshop, HiPEAC Conference, Berlin, Germany* (2013).
- [19] W.-J. Sun and C. Sechen, *A loosely coupled parallel algorithm for standard cell placement*, in *Computer-Aided Design, 1994., IEEE/ACM International Conference on* (1994) pp. 137–144.
- [20] A. Ludwin and V. Betz, *Efficient and deterministic parallel placement for fpgas*, *ACM Trans. Des. Autom. Electron. Syst.* **16**, 22:1 (2011).
- [21] Altera, *Altera quartus ii design software*, (2015).
- [22] N. Selvakkumaran and G. Karypis, *Multiobjective hypergraph-partitioning algorithms for cut and maximum subdomain-degree minimization*, *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* **25**, 504 (2006).
- [23] C. Ababei, *Speeding up fpga placement via partitioning and multithreading*, *Int. J. Reconfig. Comput.* **2009**, 6:1 (2009).
- [24] Y. Sankar and J. Rose, *Trading quality for compile time: Ultra-fast placement for fpgas*, in *Proceedings of the 1999 ACM/SIGDA Seventh International Symposium on Field Programmable Gate Arrays, FPGA '99* (ACM, New York, NY, USA, 1999) pp. 157–166.
- [25] P. Jamieson, *Revisiting genetic algorithms for the fpga placement problem*. in *GEM* (Citeseer, 2010) pp. 16–22.

- [26] H. Sidiropoulos, K. Siozios, and D. Soudris, *A methodology and tool framework for supporting rapid exploration of memory hierarchies in fpgas*, in *Field Programmable Logic and Applications (FPL), 2011 International Conference on* (2011) pp. 238–243.
- [27] E. Sentovich, K. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. Stephan, R. K. Brayton, and A. L. Sangiovanni-Vincentelli, *SIS: A System for Sequential Circuit Synthesis*, Tech. Rep. UCB/ERL M92/41 (EECS Department, University of California, Berkeley, 1992).
- [28] J. Lamoureux and S. Wilton, *On the interaction between power-aware fpga cad algorithms*, in *Computer Aided Design, 2003. ICCAD-2003. International Conference on* (2003) pp. 701–708.
- [29] M. Srinivas and L. Patnaik, *Adaptive probabilities of crossover and mutation in genetic algorithms*, *Systems, Man and Cybernetics, IEEE Transactions on* **24**, 656 (1994).
- [30] A. Livnat, C. Papadimitriou, J. Dushoff, and M. W. Feldman, *A mixability theory for the role of sex in evolution*, *Proceedings of the National Academy of Sciences* **105**, 19803 (2008), <http://www.pnas.org/content/105/50/19803.full.pdf+html> .
- [31] P. Darwen and X. Yao, *A dilemma for fitness sharing with a scaling function*, in *Evolutionary Computation, 1995., IEEE International Conference on*, Vol. 1 (1995) pp. 166–.
- [32] R. E. Smith and C. Bonacina, *Mating restriction and niching pressure: results from agents and implications for general evolutionary computation*, in *In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2003)* (Springer, 2003) pp. 1382–1393.
- [33] C.-K. Ting, S.-T. Li, and C. Lee, *On the harmonious mating strategy through tabu search*, *Information Sciences* **156**, 189 (2003), evolutionary Computation.
- [34] E. M. A. Ronald, *When selection meets seduction*, in *Proceedings of the 6th International Conference on Genetic Algorithms* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995) pp. 167–173.
- [35] T.-P. Hong and H.-S. Wang, *A dynamic mutation genetic algorithm*, in *Systems, Man, and Cybernetics, 1996., IEEE International Conference on*, Vol. 3 (1996) pp. 2000–2005 vol.3.
- [36] S. Hartmann, *A self-adapting genetic algorithm for project scheduling under resource constraints*, *Naval Research Logistics (NRL)* **49**, 433 (2002).
- [37] D. M. Ritchie, *The c programming language*, (1988).
- [38] L. Dagum and R. Menon, *Openmp: an industry standard api for shared-memory programming*, *Computational Science & Engineering, IEEE* **5**, 46 (1998).

- [39] S. Xydis, A. Bartzas, I. Anagnostopoulos, D. Soudris, and K. Pekmestzi, *Custom multi-threaded dynamic memory management for multiprocessor system-on-chip platforms*, in *Embedded Computer Systems (SAMOS), 2010 International Conference on* (2010) pp. 102–109.
- [40] J. Singh, J. Hennessy, and C. Holt, *Application and architectural bottlenecks in large scale distributed shared memory machines*, in *Computer Architecture, 1996 23rd Annual International Symposium on* (1996) pp. 134–134.
- [41] D. Jiang and J. Singh, *Scaling application performance on a cache-coherent multiprocessors*, in *Computer Architecture, 1999. Proceedings of the 26th International Symposium on* (1999) pp. 305–316.
- [42] N. Nethercote and J. Seward, *Valgrind: A framework for heavyweight dynamic binary instrumentation*, in *Proceedings of the 2007 ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI '07* (ACM, New York, NY, USA, 2007) pp. 89–100.
- [43] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, *An efficient k-means clustering algorithm: analysis and implementation*, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**, 881 (2002).
- [44] D. Shepard, *A two-dimensional interpolation function for irregularly-spaced data*, in *Proceedings of the 1968 23rd ACM National Conference*, ACM '68 (ACM, New York, NY, USA, 1968) pp. 517–524.
- [45] V. Kalenteridis, H. Pournara, K. Siozos, K. Tatas, N. Vassiliadis, I. Pappas, G. Koutroumpetis, S. Nikolaidis, S. Siskos, D. Soudris, and A. Thanailakis, *A complete platform and toolset for system implementation on fine-grain reconfigurable hardware*, *Microprocessors and Microsystems* **29**, 247 (2005).
- [46] V. Zaccaria, G. Palermo, F. Castro, C. Silvano, and G. Mariani, *Multicube explorer: An open source framework for design space exploration of chip multiprocessors*, in *Architecture of Computing Systems (ARCS), 2010 23rd International Conference on* (2010) pp. 1–7.
- [47] G. M. Amdahl, *Validity of the single processor approach to achieving large scale computing capabilities*, in *Proceedings of the April 18-20, 1967, Spring Joint Computer Conference*, AFIPS '67 (Spring) (ACM, New York, NY, USA, 1967) pp. 483–485.

5

Σύνθεση Επαναδιαμορφούμενων Συστημάτων Πολλαπλών Επιταχυντών Υλικού

Οι πρόσφατες απαιτήσεις των εφαρμογών έχουν καθιερώσει τη σύσταση νέων τεχνολογικών όρων, προκειμένου να περιγράψουν τις ιδιαιτερότητες τους. Μεταξύ αυτών, οι όροι “Big Data” και “Internet-of-Things” έχουν επικρατήσει και περιγράφουν τις συνεχώς αυξανόμενες απαιτήσεις των εφαρμογών σε μέγεθος δεδομένων επεξεργασίας και επικοινωνίας. Αυτές οι απαιτήσεις αυξάνουν την επεξεργαστική πολυπλοκότητα και συνεπώς τις ενεργειακές ανάγκες. Ταυτόχρονα, η τεχνολογική πρόοδος στη βιομηχανία ημιαγωγών προσφέρει, ανά τεχνολογικό κόμβο ολοκλήρωσης, μεγαλύτερη πυκνότητα επεξεργαστικών πόρων ανά επιφάνεια πυριτίου, χαρακτηρίζοντας τη συνέχιση του Νόμου του Moore [1–3] και μικρότερη ικανότητα ενεργειακής διαχείρισης ανά επιφάνεια πυριτίου, χαρακτηρίζοντας τη λήξη του Νόμου του Dennard [4, 5]. Το αποτέλεσμα από την αντίθεση των δύο νόμων έχει χαρακτηριστεί από την ερευνητική και βιομηχανική κοινότητα ως η εποχή της “Σκοτεινής Σιλικόνης” [5–8]. Στα πλαίσια αυτής της διδακτορικής διατριβής, προτείνονται σε αυτό το κεφάλαιο νέα αρχιτεκτονικά πρότυπα για κλιμακούμενες υπολογιστικές λύσεις. Στόχευση αυτών των αρχιτεκτονικών είναι η αύξηση της επεξεργαστικής ισχύος ανά μονάδα κατανάλωσης ενέργειας. Παράλληλα με τα αρχιτεκτονικά πρότυπα, προτείνονται τα αντίστοιχα περιβάλλοντα προγραμματισμού αυτών. Το κεφάλαιο 5 συνδέεται με τη γενικότερη μεθοδολογία της διατριβής που παρουσιάζεται στην Ενότητα 1.3, ως προς τη συνεισφορά του στη “Σύνθεση Μονοπατιού Επιτάχυνσης” και τη “Βελτιστοποίηση Ροής Σύνθεσης”.

5.1. Αρχιτεκτονικό πρότυπο και προγραμματιστική διεπαφή για συστήματα Π.Ε.

5.1.1. Εισαγωγή - Κίνητρο έρευνας σε συστήματα Π.Ε. για FPGAs

Η μεγάλη πρόκληση για τα επερχόμενα υπολογιστικά συστήματα, είναι η κλιμάκωση τους σε επίπεδο διαχείρισης δεδομένων μεγέθους Exabytes¹. Πρόσφατες ερευνητικές μελέτες έχουν δείξει ότι η επίτευξη αυτού του στόχου απαιτεί εναλλακτικές σχεδιαστικές λύσεις συν-σχεδίασης υλικού/λογισμικού, τόσο σε επίπεδο αρχιτεκτονικής, όσο και σε επίπεδο τεχνολογίας [9], [10]. Ανάμεσα στις πιο πρόσφατες και υποσχόμενες προσεγγίσεις, ανήκουν οι ετερογενείς αρχιτεκτονικές πολλαπλών επεξεργαστών, οι οποίες έχουν προταθεί ως λύσεις στο πρόβλημα της υψηλής κατανάλωσης ενέργειας των σύγχρονων υπολογιστικών συστημάτων (power wall problem) [11–13]. Για παράδειγμα, η εταιρεία Microsoft έδειξε ότι οι συγκεκριμένες αρχιτεκτονικές μπορούν να επιταχύνουν υπολογιστικούς πυρήνες σε λογισμικό μεγάλης κλίμακας, αποδίδοντας βελτιστοποίηση σε ποσοστό έως 95% αναφορικά προς την απόδοση των διακομιστών της [14].

Αναφορικά προς την αυτοματοποιημένη ηλεκτρονική σχεδίαση (EDA), τα εργαλεία σύνθεσης υψηλού επιπέδου (HLS) αναμένεται να διαδραματίσουν κεντρικό ρόλο [15] στη σχεδίαση πλατφορμών πολλαπλών επιταχυντών. Αυξάνοντας την σχεδιαστική αφαίρεση, οι σχεδιαστές μπορούν να αξιολογήσουν την απόδοση, τη κατανάλωση ενέργειας, τους επεξεργαστικούς πόρους και τις απαιτήσεις κόστους σε ένα πλήθος διαφορετικών διαμορφώσεων επιταχυντών, παρέχοντας με αυτό το τρόπο ελεγχόμενες συστημικές προδιαγραφές με σημαντικά μικρότερη προγραμματιστική δυσκολία σε σχέση με την παραδοσιακή ροή σχεδίασης, με γλώσσες περιγραφής υλικού HDL.

Αναφορικά προς τη τεχνολογία και την αρχιτεκτονική, οι ετερογενείς επαναδιαμορφούμενες πλατφόρμες αποτελούν μια υποσχόμενη προσέγγιση για συστήματα πολλαπλών επιταχυντών. Η εγγενής ευελιξία προγραμματισμού τους και η συνεχής κλιμάκωση τους σε επεξεργαστικούς πόρους επιτρέπει την ανάπτυξη ανταγωνιστικών πλατφορμών σε σύγκριση με τα συμβατικά ολοκληρωμένα κυκλώματα ASIC. Στην εργασία [16], έχει αποδειχθεί ότι οι συσκευές FPGAs μπορούν να προσφέρουν δυναμικά επιτάχυνση πολλαπλών βαθμών σε σχέση με τους συμβατικούς επεξεργαστές, για αλγόριθμους με μεγάλο επεξεργαστικό φορτίο, διατηρώντας ταυτόχρονα τη κατανάλωση ενέργειας σε χαμηλά επίπεδα. Αντίστοιχα με τη πολυ-πύρηνη τάση των επεξεργαστών γενικού σκοπού, τα συστήματα πολλαπλών επιταχυντών υλικού μπορούν να συνδυάσουν πολλαπλά παράλληλα μονοπάτια δεδομένων επεξεργασίας (datapaths) προκειμένου να αυξήσουν περαιτέρω την απόδοσή τους.

Γενικά, ως απάντηση στην εποχή της “Σκοτεινής Σιλικόνης”, η ερευνητική

¹Ο όρος “Exabyte” αποτελεί ένα πολλαπλάσιο της βασικής μονάδας πληροφορίας byte, το οποίο αποτελείται από οκτώ δυαδικά ψηφία. Ένα Exabyte αποτελεί πληροφορία από bytes πλήθους έκτης δύναμης της χιλιάδας, δηλαδή $1000^6 = 10^{18}$. Ο όρος έχει καθιερωθεί ως ορόσημο στη βιομηχανία πληροφορικής σήμερα, καθώς αντικατοπτρίζει τυπικά μεγέθη μετακίνησης και επεξεργασίας δεδομένων, βάση κλιμακούμενων στατιστικών και προβλέψεων για τις ανάγκες των υπηρεσιών πληροφορικής στα επόμενα χρόνια.

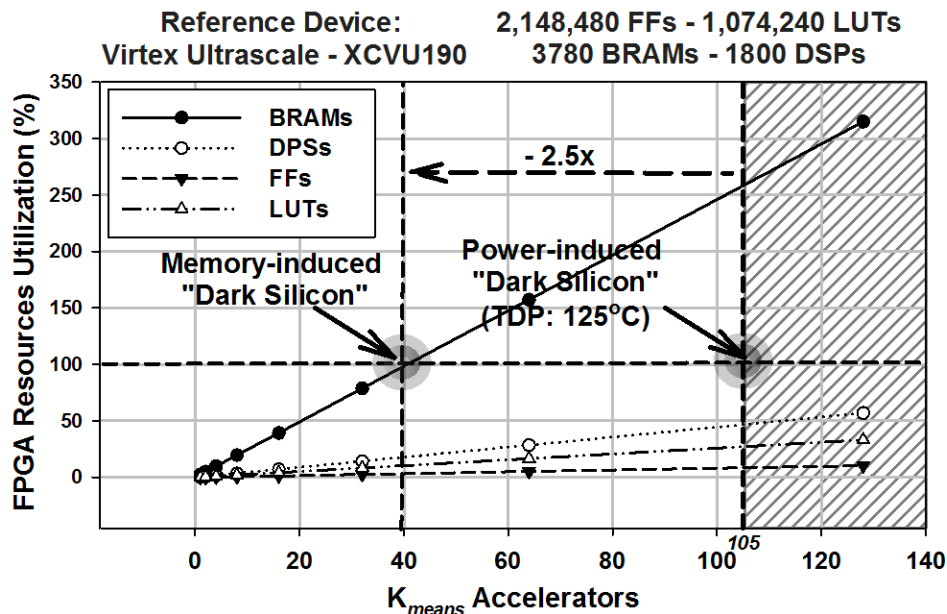
κοινότητα έχει μελετήσει τα συστήματα εξειδικευμένου υλικού ως μέσο επεξεργαστικής επιτάχυνσης, ακόμη και τα συστήματα πολλαπλών επιταχυντών. Στην εργασία [17], οι συγγραφείς προτείνουν οργάνωση της μνήμης για συστήματα πολλαπλών επιταχυντών, οι οποίοι μοιράζεται στατικά το χώρο διευθύνσεων μεταξύ τους. Ομοίως, στην εργασία [18] προτείνεται ένα αρχιτεκτονικό πρότυπο πολλαπλών επιταχυντών που επιτρέπει την επαναχρησιμοποίηση των πόρων μνήμης, υιοθετώντας μία μη ομοιόμορφη αρχιτεκτονική κρυφής μνήμης (NUCA). Ωστόσο, οι προαναφερθείσες εργασίες [17, 18] στοχεύουν σε υλοποιήσεις ASIC και εστιάζουν κυρίως στη βελτιστοποίηση επιδόσεων του υποσυστήματος μνήμης. Επίσης η ΔΔΜ για εργαλεία HLS έχει μελετηθεί από τις εργασίες [49], [20] και [21]. Ωστόσο, αυτές οι εργασίες δεν στοχεύουν συστήματα πολλαπλών επιταχυντών, και επιπλέον δε παρέχουν μηχανισμούς στο σύγχρονο πρόβλημα της υπό-αξιοποίησης των συσκευών FPGAs για αυτές τις αρχιτεκτονικές.

Σε αυτό το κεφάλαιο μελετάται το πρόβλημα κλιμάκωσης των αρχιτεκτονικών πολλαπλών επεξεργαστών, αναφορικά προς το πλήθος, το επίπεδο ανάλυσης, καθώς και τον τρόπο προγραμματισμού των επεξεργαστικών τους στοιχείων. Η έρευνα επικεντρώνεται στο σύστημα διαχείρισης μνήμης αυτών των αρχιτεκτονικών. Μια πρόσφατη έρευνα από έντεκα ελεύθερα διαθέσιμους επιταχυντές υλικού έδειξε ότι κατά μέσο όρο η μνήμη των επιταχυντών αποτελεί περίπου το 69% [22]. Συνεπώς, η υλοποίηση αρχιτεκτονικών πολλαπλών επιταχυντών σε FPGAs οδηγεί αναπόφευκτα σε ταχεία δέσμευση της διαθέσιμης on-chip μνήμης. Συγκεκριμένα η δέσμευση μνήμης σε σχέση με τους υπόλοιπους επεξεργαστικούς πόρους του FPGA είναι δυσανάλογη, με την πρώτη να υπερτερεί. Αυτή η συσχέτιση οδηγεί σε σοβαρή υπό-αξιοποίηση του FPGA, παρόμοια με την έννοια της “Σκοτεινής Σιλικόνης”, η οποία έχει αποδοθεί στο αντίξοο φαινόμενο, ότι παρόλο που σήμερα υπάρχει η κατασκευαστική και τεχνολογική δυνατότητα να προσθέτονται, σε κάθε τεχνολογική γενεά, περισσότερα τρανζίστορ σε ένα ολοκληρωμένο κύκλωμα, εντούτοις, δεν μπορούν να λειτουργήσουν ταυτόχρονα όλα μαζί, καθώς η συσσωρευμένη ενέργεια δεν μπορεί να απαχθεί με συμβατικά μέσα ψύξης. Υπό την ίδια οπτική, η δυσανάλογη δέσμευση πόρων στα FPGAs, οδηγεί σε κατάσταση μη-ολοκληρωτικής χρήσης τους, καθώς υφίστανται αχρησιμοποίητοι πόροι υλικού, π.χ. flip-flops, LUTs και DSPs, εφόσον δεν μπορούν να αξιοποιηθούν σε επιταχυντές λόγω της έλλειψης πόρων μνήμης.

Στην πραγματικότητα, τα σύγχρονα εργαλεία CAD για FPGAs, τόσο σε επίπεδο RTL, όσο και σε επίπεδο HLS, επιτρέπουν μόνο τη στατική δέσμευση μνήμης, η οποία υπαγορεύει την αποκλειστική χρήση της μέγιστης μνήμης που χρειάζεται ένα επιταχυντής, σε όλο το χρόνο λειτουργίας του. Ενώ η στατική δέσμευση δεν αποτελεί ιδιαίτερο σχεδιαστικό πρόβλημα για περιορισμένο αριθμό επιταχυντών, εντούτοις, σε ένα σύστημα πολλαπλών επιταχυντών, οι πόροι μνήμης αποτελούν αποτρεπτικό παράγοντα στη κλιμάκωση επεξεργαστικής ισχύος. Το Σχήμα 5.1 απεικονίζει ένα παράδειγμα μελέτης του προαναφερθέντος προβλήματος υπο-αξιοποίησης του FPGA, για τον αλγόριθμο ομαδοποίησης K_{means} ²

²Η ίδια συμπεριφορά παρατηρείται για ένα πλήθος διαφορετικών εφαρμογών οι οποίες παρουσιάζουν παρόμοιες αναλογίες αξιοποίησης πόρων.

στη σύγχρονη συσκευή FPGA Virtex UltraScale XVCU190³, κατά την κλιμάκωση του αριθμού των παράλληλων επιταχυντών.



Σχήμα 5.1 Ανάλυση κλιμάκωσης επιταχυντών του αλγορίθμου ομαδοποίησης K_{means} : A_i -Επιταχυντές = [1 : 128], N_p -Δεδομένα = 2×10^4 , P_k -Ομάδες = 3

Συγκεκριμένα, στο Σχήμα 5.1 απεικονίζονται δύο τιμές κατωφλίου, αναφορικά προς τη μέγιστη χρήση επεξεργαστικών πόρων και το μέγιστο περιθώριο κατανάλωσης ενέργειας, αντίστοιχα. Η επιλεγμένη συσκευή έχει χαρακτηριστική τιμή μέγιστης επιτρεπόμενης θερμοκρασίας $TDP=125^\circ C$. Θεωρώντας θερμοκρασία περιβάλλοντος $50^\circ C$, η μέγιστη επεξεργαστική ισχύς που μπορεί να ενσωματώσει η συγκεκριμένη συσκευή αποτιμάτε σε 105 επιταχυντές υλικού, οι οποίοι καταναλώνουν περίπου 20 Watts. Ωστόσο, όπως φαίνεται, η μνήμη θέτει έναν αυστηρότερο περιορισμό στο πλήθος των επιταχυντών, δηλαδή έως και 2.5x λιγότερους επιταχυντές. Αυτή η μέτρηση δείχνει ότι οι πόροι μνήμης αποτελούν ισχυρότερο αποτρεπτικό παράγοντα σε σχέση με τη κατανάλωση ενέργειας. Οι πόροι μνήμης δεσμεύονται ταχύτερα από τους υπόλοιπους επεξεργαστικούς πόρους του FPGA (FFs, LUTs και DSPs), αποτελώντας τον κύριο περιοριστικό παράγοντα στη κλιμάκωση της επεξεργαστικής ισχύος. Επίσης αποτελούν τον κύριο παράγοντα χαμηλής και δυσανάλογης αξιοποίησης των πόρων του FPGA.

Στην εργασία αυτή, αναγνωρίζονται τα δύο βασικά προβλήματα προς την ανάπτυξη συστημάτων πολλαπλών επιταχυντών ως εξής:

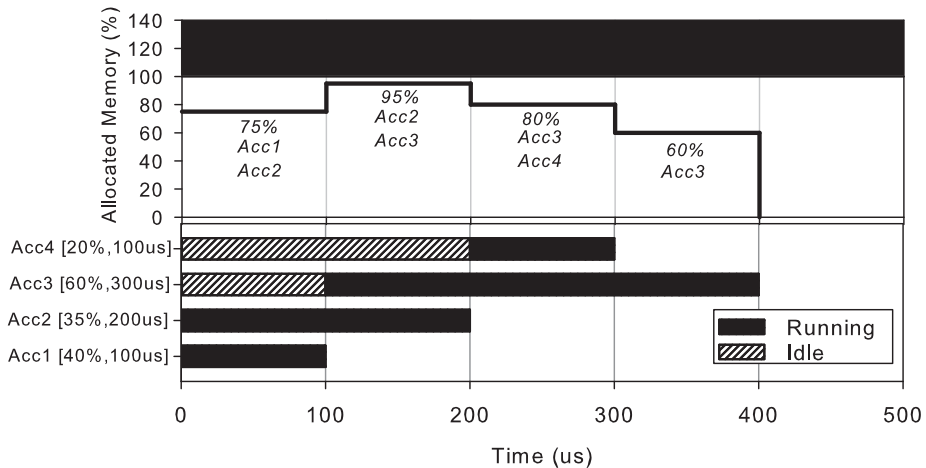
³Η συσκευή XVCU190 αποτελεί το FPGA με τη μεγαλύτερη on-chip μνήμη στην αγορά σήμερα, με συνολικό μέγεθος 132.9Mb.

- Το πρώτο πρόβλημα προκύπτει από τη δυσκολία προγραμματισμού των επιταχυντών, καθώς και τη δυσκολία δημιουργίας των μονοπατιών δεδομένων και ελέγχου αυτών. Η υφιστάμενη προσέγγιση ανάπτυξης του υποσυστήματος υλικού με γλώσσες περιγραφής υλικού HDL αποτελεί ήδη χρονόβόρα διαδικασία ανάπτυξης για ακολουθιακά μονοπάτια μονο-επεξεργαστικής ροής. Συνεπώς η προσέγγιση μίας αποκεντροποιημένης και κατανεμημένης επεξεργαστικής ροής δυσκολεύει ακόμα περισσότερο την ανάπτυξη μέσω γλωσσών HDL. Επίσης σε ένα τόσο πολύπλοκο σύστημα η διερεύνηση του χώρου λύσεων με γλώσσες HDL είναι μια σχεδόν ακατόρθωτη διαδικασία.
- Το δεύτερο πρόβλημα προκύπτει από τη δυσανάλογη αξιοποίηση των επεξεργαστικών πόρων σε σχέση με τους πόρους μνήμης των συσκευών FPGA, ως περιγράφηκε ανωτέρω. Το πρόβλημα αυτό σχετίζεται με την υποαξιοποίηση πόρων του FPGA, που για συγκεκριμένες εφαρμογές και αρχιτεκτονικές, μπορεί να οδηγήσει σε μειωμένη επεξεργαστική ισχύ, ακόμα μικρότερη και από αυτήν που επιβάλλουν οι τεχνολογικοί περιορισμοί, όπως π.χ. η ενέργεια κατανάλωσης με το αντίστοιχο πρόβλημα “Dark-Silicon”. Οι αρχιτεκτονικές πολλαπλών επιταχυντών υλικού εμπίπτουν σε αυτή τη κατηγορία.

Η στόχευση της παρούσας εργασίας είναι αφενός ο αποδοτικός προγραμματισμός συστημάτων πολλαπλών επιταχυντών υλικού και αφετέρου η βελτιστοποίηση της δέσμευσης μνήμης σε συσκευές FPGA. Η κύρια συνεισφορά της εργασίας είναι η εισαγωγή ενός νέου μεθοδολογικού πλαισίου, βασισμένο σε εργαλεία HLS, για τη σχεδίαση πλατφορμών πολλαπλών επιταχυντών, το οποίο υιοθετεί τεχνικές Δυναμικής Διαχείρισης Μνήμης (ΔΔΜ), κατά το χρόνο εκτέλεσης. Κατά την τυπική ροή σχεδιασμού με εργαλεία HLS, τα οποία επιτρέπουν μόνο στατική δέσμευση μνήμης, στη περίπτωση όπου οι απαιτήσεις μνήμης των επιταχυντών υλικού υπερβαίνουν τους διαθέσιμους πόρους μνήμης, τότε ο σχεδιασμός δεν είναι συνθέσιμος και επομένως τα εργαλεία αποτυγχάνουν να εξάγουν κώδικα προγραμματισμού του FPGA. Σε αυτή τη περίπτωση οι σχεδιαστές υποβαθμίζουν τα χαρακτηριστικά του συστήματος, έτσι ώστε αυτό να ανταποκρίνεται στους διαθέσιμους πόρους. Η προτεινόμενη λύση επιτρέπει πολλαπλούς επιταχυντές υλικού προκειμένου να αντιμετωπίσει την υπό-αξιοποίηση των πόρων του FPGA, η οποία προκαλείται κυρίως από τη στατική στρατηγική δέσμευσης μνήμης που χρησιμοποιούνται από τα σύγχρονα εργαλεία HLS. Παράλληλα, δείχνεται ότι οι εφαρμογές που έχουν ανάγκες δέσμευσης μνήμης, τόσο δυναμικές όσο και στατικές, μπορούν να επωφεληθούν από τις προτεινόμενες τεχνικές. Σε κάθε περίπτωση, απαιτείται μία απλή τροποποίηση του πηγαίου κώδικα.

Προτείνεται η υιοθέτηση τεχνικών Δυναμικής Διαχείρισης Μνήμης (ΔΔΜ) για την αντιμετώπιση των αυστηρών περιορισμών που επιβάλλονται από τη στατική δέσμευση μνήμης στα εργαλεία σύνθεσης για FPGAs. Στην στατική κατανομή μνήμης, όλες οι αποφάσεις δέσμευσης πόρων μνήμης λαμβάνονται κατά το χρόνο σχεδίασης. Εάν οι απαιτήσεις μνήμης των επιταχυντών υλικού υπερβούν τους διαθέσιμους πόρους μνήμης, τότε ο σχεδιασμός κρίνεται μη-συνθέσιμος. Ωστόσο, στη δυναμική δέσμευση μνήμης, η δέσμευση πόρων μνήμης γίνεται

βάση των αιτημάτων μνήμης των επιταχυντών κατά το χρόνο εκτέλεσης. Στη περίπτωση που δεν υπάρχουν διαθέσιμοι πόροι μνήμης η προτεινόμενη τεχνική ΔΔΜ καθυστερεί την εκτέλεση του επιταχυντή, έως ότου ελευθερωθούν οι απαιτούμενοι πόροι. Το Σχήμα 5.2 απεικονίζει ένα παραδειγματικό σενάριο, όπου τέσσερις επιταχυντές υλικού, οι οποίοι συγκεντρωτικά έχουν υψηλότερες απαιτήσεις μνήμης από τους διαθέσιμους πόρους του FPGA, μπορούν να συντεθούν τελικά και να προγραμματιστούν στο FPGA, με την υποστήριξη της δυναμικής διαχείρισης μνήμης. Το βασικό στοιχείο είναι η ενσωμάτωση ενός επιπέδου διαχείρισης μνήμης, κατά τη σύνθεση, σε επίπεδο συστήματος. Το επίπεδο αυτό είναι υπεύθυνο για τη δέσμευση και κατανομή μνήμης σύμφωνα με τα αιτήματα των επιταχυντών, κατά το χρόνο εκτέλεσης, καθώς και την παύση εκτέλεσης των επιταχυντών, στη περίπτωση όπου οι αιτήσεις δέσμευσης μνήμης υπερβαίνουν τους διαθέσιμους πόρους.

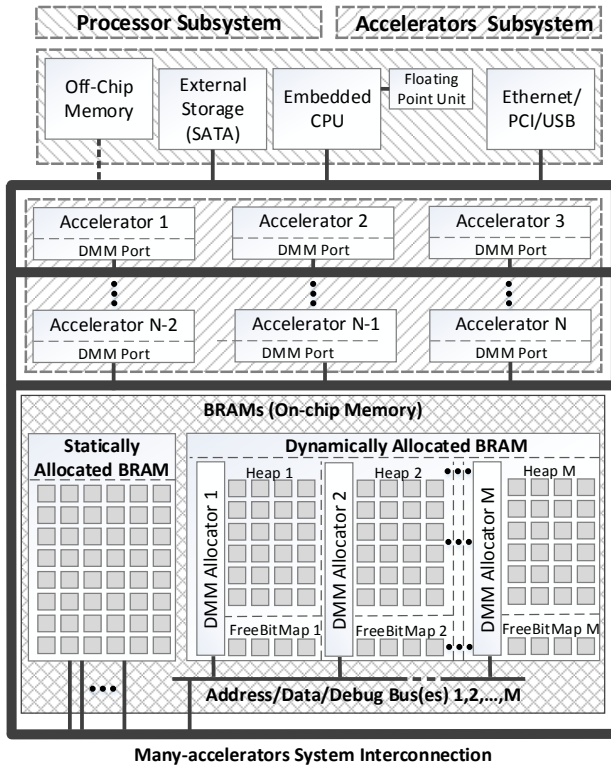


Σχήμα 5.2 Παραδειγματικό σενάριο εκτέλεσης τεσσάρων επιταχυντών υλικού. *Acc-*i**[Στατική Χρήση Μνήμης%, Καθυστέρηση]. Όλοι οι επιταχυντές αναμένεται να εκκινηθούν τη χρονική στιγμή 0. Ο σχεδιασμός κρίνεται μη-συνθέσιμος με στατική δέσμευση μνήμης. Το κατώτερο τμήμα του σχήματος απεικονίζει τον τελικό χρονοπρογραμματισμό με χρήση τεχνικών ΔΔΜ. Το ανώτερο τμήμα του σχήματος απεικονίζει την αντίστοιχη δέσμευση της συνολικής μνήμης. Εξαιτίας της κλιμάκωσης του χρόνου, στο συγκεκριμένο σχήμα δεν απεικονίζεται το αποτύπωμα μνήμης κατά τις στιγμιαίες μεταβάσεις νέων δεσμεύσεων/αποδεσμεύσεων μνήμης.

Σε αυτή την εργασία προτείνεται το πλαίσιο DMM-HLS, το οποίο (i) επεκτείνει τυπικά εργαλεία HLS με μηχανισμούς ΔΔΜ και (ii) παρέχει κατάλληλη προγραμματιστική διεπαφή HLS malloc/free API, η οποία επιτρέπει τη μετατροπή στατικών δεσμεύσεων μνήμης σε δυναμικές. Η αποτελεσματικότητα του προτεινόμενου πλαισίου DMM-HLS αξιολογείται σε αρχιτεκτονικές πολλαπλών-επιταχυντών υλικού για ένα σύνολο αντιπροσωπευτικών εφαρμογών. Δείχνεται ότι το πλαίσιο DMM-HLS προσφέρει περισσότερες δυνατότητες κλιμάκωσης αρχιτεκτονικών λύσεων, με μέσο όρο αύξησης των επιταχυντών έως 3.8x σε

σύγκριση με τα συστήματα πολλαπλών επιταχυντών που σχεδιάζονται με σύγχρονα εργαλεία HLS. Αυτή η αύξηση της κλιμάκωσης οδηγεί επίσης σε σημαντικά κέρδη ρυθμαπόδοσης (throughput), σημειώνοντας, κατά μέσο όρο, κέρδη έως 24.1x και 3.8x, για ρυθμίσεις ιδιωτικής και διαμοιραζόμενης μνήμης, αντίστοιχα.

5.1.2. Το πλαίσιο DMM-HLS για συστήματα πολλαπλών επιταχυντών σε FPGAs



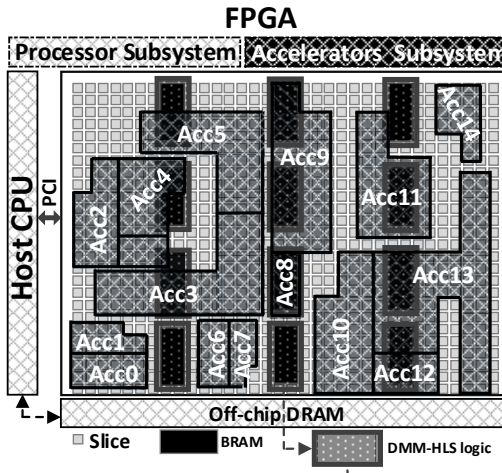
Σχήμα 5.3 Προτεινόμενο λογικό αρχιτεκτονικό πρότυπο για συστήματα πολλαπλών επιταχυντών υλικού σε FPGAs.

Το Σχήμα 5.3 απεικονίζει το προτεινόμενο αρχιτεκτονικό πρότυπο για συστήματα πολλαπλών επιταχυντών υλικού σε FPGAs. Περιλαμβάνει i) το υποσύστημα επεξεργαστή, το οποίο εκτελεί τη ροή ελέγχου της εφαρμογής και ii) το υποσύστημα επιταχυντών υλικού, το οποίο αποτελείται από τους υπολογιστικούς πυρήνες επιτάχυνσης επεξεργασίας. Οι επιταχυντές σχεδιάζονται και συνθέτονται σε υλικό μέσω του εργαλείου Xilinx Vivado-HLS. Οι πότοι μνήμης του FPGA (BRAMs) διαχειρίζονται μέσω του πλαισίου DMM-HLS. Το πλαίσιο DMM-HLS υποστηρίζει επιταχυντές υλικού τόσο με στατική όσο και με δυνα-

μική δέσμευση μνήμης. Παρέχει δε ένα προγραμματιστικό πλαίσιο, το οποίο αποτελείται από δύο κύριες κλήσεις, παρόμοιες με τις αντίστοιχες του *glibc* `malloc/free` API, για την δέσμευση και αποδέσμευση μνήμης.

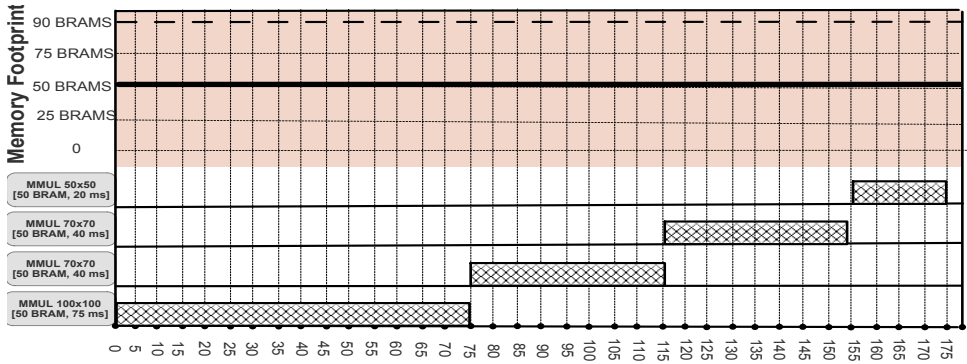
- `void* HlsMalloc(size_t size, uint heap_id)`
- `void HlsFree(void *ptr, uint heap_id)`

, όπου η μεταβλητή `size` είναι το μέγεθος της αίτησης δέσμευσης μνήμης σε bytes, η μεταβλητή `heap_id` είναι ο μοναδικός αριθμός αναγνώρισης του σωρού στον οποίο επρόκειτο να γίνει η δέσμευση και η μεταβλητή `*ptr` είναι ο δείκτης διεύθυνσης στην οποία υπάρχει αίτηση αποδέσμευσης μνήμης.

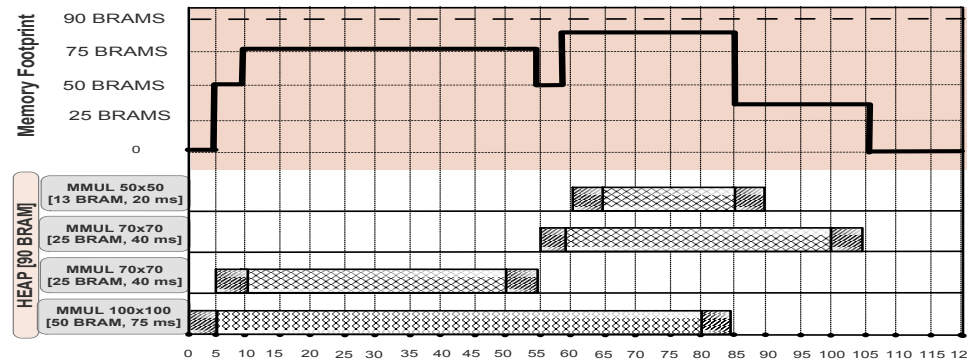


Σχήμα 5.4 Προτεινόμενο αρχιτεκτονικό πρότυπο για συστήματα πολλαπλών επιταχυντών υλικού σε FPGAs, σε επίπεδο υλικού.

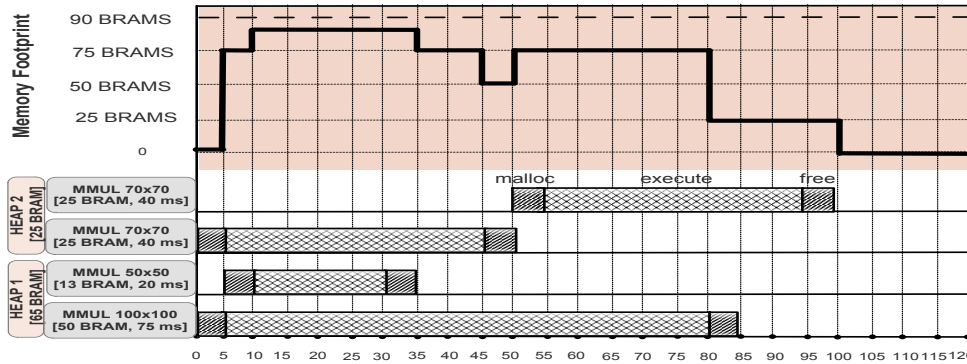
Στο προτεινόμενο πλαίσιο, τα δεδομένα μίας εφαρμογής μπορούν να χωριστούν σε δύο κατηγορίες, στατικής και δυναμικής δέσμευσης. Για την αξιολόγηση του πλαισίου, ακολουθήθηκε μία απλή κατάταξη των μεταβλητών δεδομένων στις δύο αυτές κατηγορίες. Τα δεδομένα σωρού ή/και δεδομένα γενικής εμβέλειας (*global scope data structures*), δηλαδή τα δεδομένα των επιταχυντών που επεξεργάζονται κατ' επίκληση, δεσμεύονται δυναμικά κατά το χρόνο εκτέλεσης. Αντίθετα τα τοπικά δεδομένα, π.χ. τοπικά αρχεία καταχωρητών κτλ., δεσμεύονται στατικά, κατά το χρόνο σχεδίασης/σύνθεσης. Η φυσική υλοποίηση του αρχιτεκτονικού προτύπου απεικονίζεται στο Σχήμα 5.4. Οι επιταχυντές δεσμεύουν επεξεργαστικούς πόρους μέσω των εργαλείων φυσικού σχεδιασμού για FPGs (*synthesis, mapping, place & route*). Οι μνήμες BRAMs στο FPGA δεσμεύονται είτε απευθείας από τους επιταχυντές, όταν πρόκειται για στατική δέσμευση, είτε μέσω ενός επιπέδου διαχείρισης ΔΔΜ. Το επίπεδο αυτό υλοποιείται μέσω μίας συνθέσιμης βιβλιοθήκης λογισμικού για εργαλεία HLS.



(α)



(β)



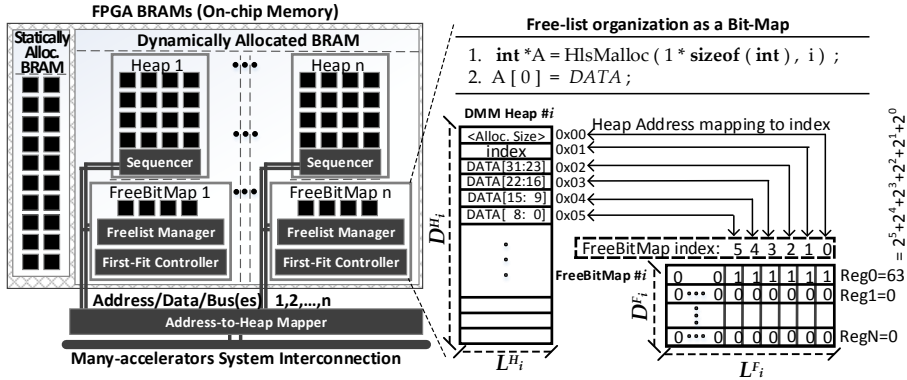
(γ)

Σχήμα 5.5 Κέρδος απόδοσης λόγω παραλληλίας και επικάλυψης διεργασιών εκτέλεσης, μέσω του πλαισίου DMM-HLS. Χρονοπρογραμματισμός των επιταχυντών και αποτύπωμα μνήμης για α) συμβατικά εργαλεία HLS με στατική δέσμευση μνήμης, β) πλαίσιο DMM-HLS με ένα μοναδικό σωρό, γ) πλαίσιο DMM-HLS με δύο σωρούς. Συσκευή FPGA: 90 BRAMs. Συνολική ζήτηση μνήμης: 113 BRAMs.

Το πλαίσιο DMM-HLS εφαρμόζει τεχνικές από το ερευνητικό πεδίο της δυναμικής διαχείρισης μνήμης πολλαπλών νημάτων για πολυπύρηνους επεξεργαστές [23], [24]. Υποστηρίζει παράλληλα μονοπάτια πρόσβασης στη μνήμη, με την ομαδοποίηση φυσικών πόρων μνήμης σε αδρομερούς εικονικούς πόρους, που ονομάζεται σωροί (heaps), όπως απεικονίζεται και στο Σχήμα 5.4. Κάθε σωρός διαχειρίζεται από έναν αποκλειστικό ελεγκτή μνήμης, ο οποίος αποτελείται από δύο βασικές υπομονάδες υλικού, i) τη δομή μνήμης free-list, η οποία αποθηκεύει πληροφορίες για τις δεσμευμένες και ελεύθερες θέσεις μνήμης στη σωρό και ii) τον αλγόριθμο δέσμευσης, ο οποίος αναζητά ελεύθερο χώρο στη δομή free-list σε κάθε νέα αίτηση δέσμευσης. Το πλαίσιο δεν επιβάλλει αυστηρή συσχέτιση σωρών-επιταχυντών. Αντίθετα, είναι δυνατόν περισσότεροι από έναν επιταχυντές να χρησιμοποιούν τον ίδιο σωρό καθώς και ένα επιταχυντής να δεσμεύει δεδομένα σε περισσότερους από έναν σωρό. Η αμφιμονοσήμαντη αυτή συσχέτιση επιτρέπει μεγάλη ευελιξία στη ΔΔΜ, η οποία πρέπει όμως να καθοριστεί επακριβώς προκειμένου να αποτραπούν συνθήκες ανταγωνισμού στους διαύλους δεδομένων και διευθύνσεων των σωρών. Ο αυξημένος εγγενής παραλληλισμός των επιταχυντών υλικού σε συνδυασμό με την επικαλυπτόμενη εκτέλεση που προσφέρεται από τη χρήση πολλαπλών σωρών ΔΔΜ, επιτρέπει την επίτευξη σημαντικών κερδών ρυθμαπόδοσης, όταν υφίσταται φόρτος εργασίας μεταβλητού μεγέθους κατά το χρόνο εκτέλεσης, σε FPGAs. Το Σχήμα 5.5 απεικονίζει ένα τέτοιο σενάριο λειτουργίας, στο οποίο προσδιορίζεται το κέρδος ρυθμαπόδοσης μέσω του παραλληλισμού των επιταχυντών και της επικάλυψης εκτέλεσης των σωρών. Συγκεκριμένα, το παράδειγμα περιγράφει την εκτέλεση τεσσάρων αιτήσεων πολλαπλασιασμού πινάκων ακεραίων (MMUL) με μεταβλητά διανύσματα εισόδου, σε ένα FPGA με μέγιστο μέγεθος μνήμης 90 BRAMs⁴, χρησιμοποιώντας περιβάλλον ΔΔΜ με δύο σωρούς. Τονίζεται ότι οι απαιτήσεις μνήμης και η καθυστέρηση εκτέλεσης των επιταχυντών αποτελούν πραγματικές μετρήσεις, που εξήχθησαν από το εργαλείο Vivado-HLS και Xilinx ISIM.

Όπως μπορεί εύκολα να διαπιστωθεί, η προτεινόμενη λύση παρέχει επιτάχυνση 42,8% σε σχέση με την λύση HLS στατικής δέσμευσης μνήμης, στην οποία ακολουθείται σειριακή εκτέλεση των διεργασιών. Όπως φαίνεται, η εκτέλεση με στατική δέσμευση απαιτεί περίπου 175ms για να ολοκληρωθεί, δίχως κάποια αλλαγή του πηγαίου κώδικα για το πλαίσιο DMM-HLS. Ωστόσο, καθώς η στατική δέσμευση μνήμης διαχειρίζεται από το Vivado-HLS ως δέσμευση μνήμης για τη χειρότερη περίπτωση, μόνο ένας επιταχυντής με μέγεθος δέσμευσης μνήμης 50 BRAMs μπορεί να συντεθεί επιτυχώς στη συσκευή FPGA. Συνεπώς, ο επεξεργαστικός φόρτος εργασίας σειριοποιείται σε έναν μοναδικό επιταχυντή. Ωστόσο, στη περίπτωση εφαρμογής του πλαισίου DMM-HLS, έως τέσσερις επιταχυντές υλικού μπορούν να συντεθούν και να ενεργοποιηθούν δυναμικά κατά το χρόνο εκτέλεσης, ολοκληρώνοντας την εκτέλεσή τους σε 105 ms, αποδίδοντας κέρδος εκτέλεσης κατά 40%. Εάν χρησιμοποιηθούν περαιτέρω δύο σωροί, ο παραλληλισμός των επιταχυντών αυξάνεται και μπορεί να σημειώσει κέρδος 42,8% σε σχέση με τη στατική δέσμευση μνήμης που χρησιμοποιεί το εργαλείο Vivado-HLS.

⁴Στο συγκεκριμένο παράδειγμα, κάθε BRAM εμπεριέχει φυσικούς πόρους μνήμης μεγέθους 18Kb.



Σχήμα 5.6 Αρχιτεκτονική διάταξη των ελεγκτών μνήμης του πλαισίου DMM-HLS για ΔΔΜ σε FPGAs.

Η εσωτερική δομή του ελεγκτή που διαθέτει κάθε σωρός του πλαισίου DMM-HLS, απεικονίζεται στο Σχήμα 5.3. Μια δομή μνήμης FreeBitMap συγκεντρώνει τη πληροφορία σχετικά με τις ελεύθερες θέσεις μνήμης του σωρού. Σε επίπεδο σχεδίασης υλοποιείται ως ένας πίνακας καταχωρητών στον οποίο κάθε δυαδικό ψηφίο αντιστοιχεί στη κατάσταση δέσμευσης ενός byte του σωρού. Η διαχείριση της δομής FreeBitMap πραγματοποιείται μέσω πράξεων σε επίπεδο ψηφίου, ώστε να καταστεί δυνατή ταχύτερη και αποτελεσματικότερη σύνθεση υλικού. Κάθε σωρός-*i* είναι παραμετροποιήσιμος σε ένα σύνολο σχεδιαστικών επιλογών, οι σημαντικότερες εκ των οποίων είναι (i) το βάθος του σωρού D_i^H , δηλαδή ο συνολικός αριθμός μοναδικών διευθύνσεων, (ii) το μήκος λέξης του σωρού L_i^H , δηλαδή ο αριθμός των bytes κάθε λέξης, (iii) το βάθος της δομής FreeBitMap D_i^F , δηλαδή ο συνολικός αριθμός των καταχωρητών της δομής free-list, (iv) το μήκος λέξης της δομής FreeBitMap L_i^F , δηλαδή ο αριθμός των bytes κάθε λέξης της δομής FreeBitMap, (v) η ευθυγράμμιση των δεσμεύσεων A_i , δηλαδή ο ελάχιστος αριθμός bytes ανά δέσμευση, έτσι ώστε κάθε νέα δέσμευση να ξεκινά από μια μοναδική διεύθυνση και (vi) το μέγεθος επικεφαλίδας μετα-πληροφοριών H_i , δηλαδή ο αριθμός bytes που καταλαμβάνει την πρώτη διεύθυνση(, ή/και διευθύνσεις) της κάθε δέσμευσης για την αποθήκευση μετα-δεδομένων που σχετίζονται με την δέσμευση. π.χ. μέγεθος δέσμευσης. Το σύνολο αυτών των παραμέτρων οδηγεί σε ένα ευέλικτο χώρο σχεδιασμού ανά σωρό, που μπορεί να προσαρμοστεί ανάλογα με τις αιτήσεις εκχώρησης μνήμης που κάθε σωρός πρόκειται να εξυπηρετήσει. Σε αυτή την εργασία χρησιμοποιήθηκε η ακόλουθη διαμόρφωση ανά σωρό: $[D_i^H, L_i^H, D_i^F, L_i^F, A_i, H_i] = [8192, 4, 1024, 4, 4, 2]$. Η προαναφερθείσα αρχιτεκτονική διάταξη των ελεγκτών μνήμης του προτεινόμενου πλαισίου DMM-HLS για ΔΔΜ σε FPGAs απεικονίζεται στο Σχήμα 5.6.

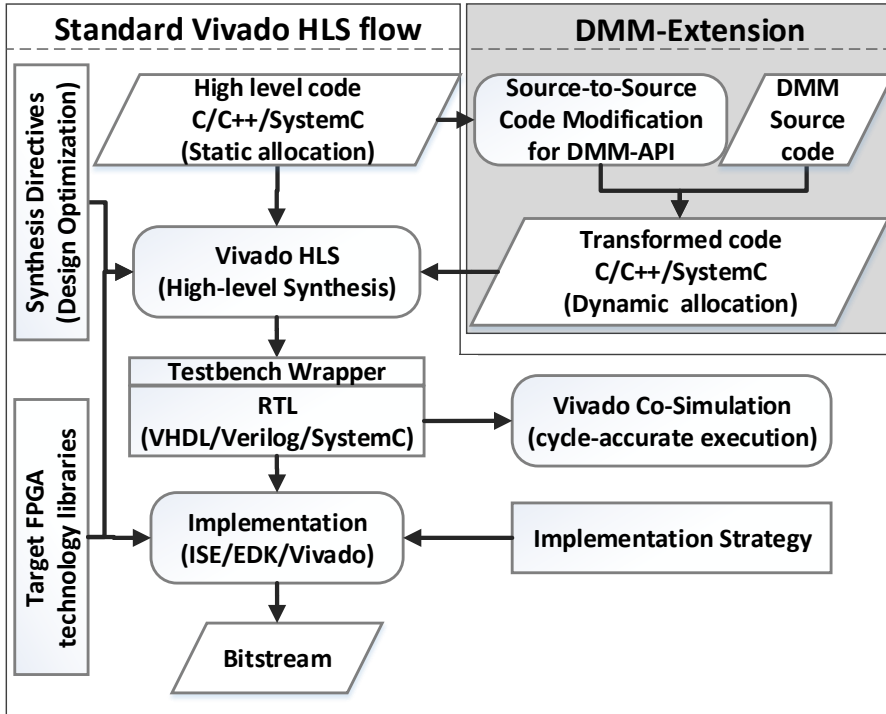
Αναγνωρίζονται τρία σημαντικά ζητήματα που σχετίζονται με τη ΔΔΜ για συ-

στήματα πολλαπλών επιταχυντών. Στα πλαίσια αυτής της έρευνας αναφέρονται ως ο κατακερματισμός μνήμης (*memory fragmentation*), η συνοχή μνήμης (*memory coherency*) και οι συγκρούσεις πρόσβασης στη μνήμη (*memory access conflicts*). Ο κατακερματισμός μνήμης ορίζεται ως η ποσότητα της πραγματικής μνήμης που χρησιμοποιείται από έναν επιταχυντή σε σχέση με την ποσότητα της μνήμης που ζητήθηκε από αυτόν. Στα συστήματα πολλαπλών επιταχυντών αναγνωρίζονται δύο τύποι κατακερματισμού, ο κατακερματισμός ευθυγράμμισης και ο κατακερματισμός αίτησης. Ο κατακερματισμός ευθυγράμμισης αντιπροσωπεύει τα επιπλέον bytes που δεσμεύονται για τη πλήρωση κάθε δέσμευσης σε ακέραιο πολλαπλάσιο του μήκους λέξης της σωρού L_i^H , συμπεριλαμβανομένου του απόλυτου μεγέθους δέσμευσης και των πληροφοριών κεφαλίδας μετα-δεδομένων. Εφόσον το μέγεθος των αιτήσεων DMM (*malloc/free*) είναι πολλαπλάσιο του μεγέθους L_i^H , ο κατακερματισμός ευθυγράμμισης είναι μηδενικός.

Ο κατακερματισμός αίτησης αναφέρεται στην κατάσταση κατά την οποία ένα αίτημα μνήμης παρακάμπει περιοχές μνήμης προκειμένου να εντοπίσει συνεχή ελεύθερο χώρο μνήμης που ισοδυναμεί με το ακριβές μέγεθος της αίτησης. Στη χειρότερη περίπτωση, η αίτηση δεν μπορεί να εξυπηρετηθεί ακόμη και αν το μέγεθος της ελεύθερης μνήμης στο σωρό είναι διαθέσιμο σε μη-συνεχείς περιοχές. Ο κατακερματισμός αίτησης εξαρτάται σε μεγάλο βαθμό από τα χαρακτηριστικά των δεσμεύσεων μνήμης του κάθε επιταχυντή. Σε περίπτωση ενός ομογενούς συστήματος πολλαπλών επιταχυντών, δηλαδή όπου κάθε επιταχυντής δεσμεύει το ίδιο ή/και ακέραιο πολλαπλάσιο μέγεθος μνήμης, ο κατακερματισμός αίτησης είναι μηδενικός. Λόγω της χρήσης αλγορίθμου αναζήτησης *first-fit*, ο χειρότερος κατακερματισμός αίτησης εμφανίζεται σε συστήματα πολλαπλών επιταχυντών με ετερογενή αιτήματα δέσμευσης μνήμης και μόνο στην περίπτωση διαδοχικών αιτήσεων όπου η μεταγενέστερη δέσμευση αιτείται μεγαλύτερου μεγέθους μνήμης από την προγενέστερη.

Τα προβλήματα συνοχής μνήμης αποκλείονται εγγενώς από το πλαίσιο DMM-HLS, δεδομένου ότι κάθε επιταχυντής έχει το δικό του χώρο στη μνήμη, συνεπώς, κανένας άλλος επιταχυντής δεν μπορεί να έχει πρόσβαση σε αυτόν. Ωστόσο, οι συγκρούσεις πρόσβαση στη μνήμη αποτελούν ιδιαίτερο πρόβλημα απόδοσης σε περίπτωση που ένα μεγάλο σύνολο επιταχυντών μοιράζονται την ίδια σωρό. Όπως αναφέρθηκε προηγουμένως, το πλαίσιο DMM-HLS υποστηρίζει διαμορφώσεις πολλαπλών σωρών προκειμένου να υποστηρίξει ευελιξία στο σύνολο μνήμης, κατά τη ΔΔΜ.

Το Σχήμα 5.7 απεικονίζει μία επισκόπηση του προτεινόμενου πλαισίου DMM-HLS για σχεδίαση και επαλήθευση ορθής λειτουργίας συστημάτων πολλαπλών επιταχυντών σε FPGAs. Η σχεδιαστική ροή βασίζεται στο εργαλείο Xilinx Vivado-HLS, ένα σύγχρονο και βιομηχανικό εργαλείο HLS. Η επέκταση του πλαισίου DMM-HLS επενεργεί στο επίπεδο πηγαίου κώδικα της εφαρμογής, με αποτέλεσμα την προγραμματιστικά εύκολη εφαρμογή του. Το στάδιο τροποποίησης του πηγαίου κώδικα περιλαμβάνει τη μετατροπή των στατικών δεσμεύσεων μνήμης σε δυναμικές χρησιμοποιώντας τις κλείσεις του DMM-HLS API. Ένα παραδειγματικό σενάριο αυτών των μετασχηματισμών κώδικα απεικονίζεται στον Πίνακα 5.1. Ο μετασχηματισμένος κώδικας προσαυξάνεται με τις κλήσεις των



Σχήμα 5.7 Προτεινόμενη επέκταση του εργαλείου Vivado HLS για την υποστήριξη ΔΔΜ σε συστήματα πολλαπλών επιταχυντών σε FPGAs.

συναρτήσεων DMM-HLS και συντίθεται σε κώδικα RTL μέσω του εργαλείου Vivado HLS. Το πλαίσιο προβλέπει την επαλήθευση ορθής λειτουργίας μέσω κατάλληλων testbenches και του εγγενή μηχανισμού συν-εξομοίωσης του εργαλείου Vivado-HLS (Xilinx ISIM). Μετά την επιτυχή ολοκλήρωση αυτού του σταδίου, ακολουθεί παραγωγή κατάλληλου bitstream προγραμματισμού για την υλοποίηση του συστήματος σε FPGA.

Όσον αφορά τη παύση εκτέλεσης ενός επιταχυντή, στη περίπτωση όπου δεν υπάρχει διαθέσιμη ελεύθερη μνήμη στο σωρό, χρησιμοποιείται κατάλληλος βρόχος αναμονής στη κλήση HlsMalloc. Όταν υπάρχει διαθέσιμη ελεύθερη μνήμη σε μία συγκεκριμένη περιοχή του σωρού, ο διαχειριστής μνήμης επιστρέφει την πρώτη διεύθυνση αυτής της περιοχής μνήμης. Όταν δεν υπάρχει ελεύθερη μνήμη στο σωρό, ο διαχειριστής μνήμης επιστρέφει την αναγνωριστική τιμή “-1”. Χρησιμοποιώντας το βρόχο καθυστέρησης, επιτυγχάνεται καθυστέρηση εκτέλεσης του επιταχυντή στο σημείο κλήσης της συνάρτησης HlsMalloc, στον πηγαίο κώδικα. Την ίδια στιγμή, οι υπόλοιποι επιταχυντές εκτελούνται παράλληλα και συνεπώς ελευθερώνουν μνήμη κατά τον τερματισμό τους. Έως αυτό το σημείο, οι επιταχυντές που έχουν αδρανοποιηθεί ελέγχουν περιοδικά για διαθέσιμο ελεύ-

Πίνακας 5.1 Παραδειγματικό σενάριο μετατροπής στατικών δεσμεύσεων μνήμης σε δυναμικές, χρησιμοποιώντας κλείσεις του DMM-HLS API.

<u>Original Code</u>	<u>Transformed Code for DMM-HLS</u>
	<code>const unsigned int T=500; /* check period */</code>
	<code>int *IN,*OUT;</code>
<code>int IN [10];</code>	<code>while((IN =HlsMalloc(10,0))!=-1){ HlsSleep(T)}</code>
<code>int OUT[10];</code>	<code>while((OUT=HlsMalloc(10,0))!=-1){ HlsSleep(T)}</code>
<code>foo(IN,&OUT);</code>	<code>foo(IN,&OUT);</code>
	<code>HlsFree(IN,0);</code>
	<code>HlsFree(OUT,0);</code>

θερο χώρο μνήμης (π.χ. κάθε T = 500 κύκλους).

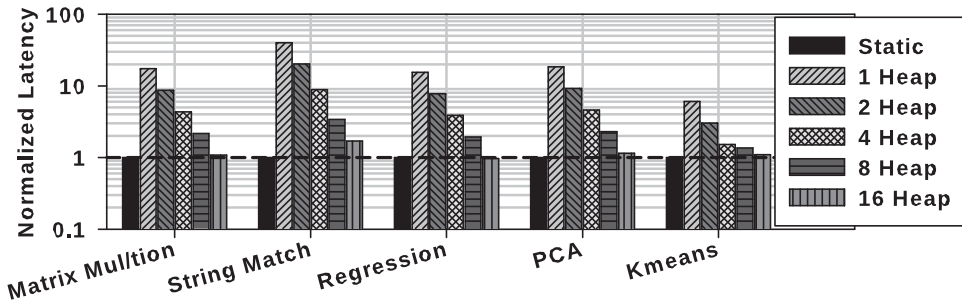
5.1.3. Αξιολόγηση του πλαισίου DMM-HLS

Στη παρούσα έρευνα χρησιμοποιήθηκε η συσκευή Virtex UltraScale XVCU190, η οποία διαθέτει το μεγαλύτερο μέγεθος μνήμης στην αγορά, σήμερα (132.9 Mb). Αξιολογήθηκε η αποτελεσματικότητα του προτεινόμενου πλαισίου DMM-HLS θεωρώντας αρχιτεκτονικές πολλαπλών επιταχυντών, σε σύγχρονες εφαρμογές των πεδίων της τεχνητής νοημοσύνης, της επιστημονικής πληροφορικής (scientific computing) και της επιχειρηματικής πληροφορικής (enterprise computing). Συγκεκριμένα, χρησιμοποιήθηκαν έξι εφαρμογές, οι οποίες αναφέρονται στον Πίνακα 5.2 και οι οποίες χρησιμοποιούνται στο πλαίσιο Phoenix MapReduce για συστήματα διαμοιραζόμενης μνήμης [25]. Θεωρήθηκε ένα πείραμα αξιολόγησης χιλίων (1.000) διεργασιών. Οι απαιτήσεις μεγέθους μνήμης των διεργασιών προέρχεται από μια κανονική κατανομή $N(\frac{MaxSize}{2}, MaxSize)$, όπου η μεταβλητή *MaxSize* ορίζεται στην τελευταία στήλη του Πίνακα 5.2. Οι διεργασίες εισάγονται στο σύστημα την ίδια χρονική στιγμή και εκτελούνται όποτε υπάρχει διαθέσιμη μνήμη. Σε περίπτωση ανταγωνισμού των διεργασιών επιλέγονται οι διεργασίες με μεγαλύτερο μέγεθος αιτήματος μνήμης.

Πίνακας 5.2 Χαρακτηρισμός εφαρμογών αξιολόγησης.

Εφαρμογή	Περιγραφή	Παράμετροι
Histogram	Determine frequency of RGB channels in image.	$M_{size} = 640 \times 480$ pixels
Matrix Mul/tion	Dense integer matrix multiplication.	$M_{size} = 100 \times 100$
String Match	Search file with keys for an encrypted word	$N_{file-keys} = 307,200$
Linear Regr.	Compute the best fit line for a set of points.	$N_{points} = 100,000$
PCA	Principal components analysis on a matrix.	$M_{size} = 250 \times 250$
<i>K</i> _{means}	Iterative clustering algorithm to classify <i>n</i> -D data points into groups.	$N_{points} = 20,000,$ $P_{clusters} = 10, n_{dimensions} = 3$

Προκειμένου να αξιολογηθεί η επιβάρυνση καθυστέρησης ανά επιταχυντή που οφείλεται στους μηχανισμούς ΔΔΜ, το Σχήμα 5.8 απεικονίζει την κανονικοποιημένη καθυστέρηση των επιταχυντών χρησιμοποιώντας 1,2,4,8 και 16 σωρούς. Όπως φαίνεται, η μέση επιβάρυνση για όλους τους επιταχυντές είναι 19,9× όταν

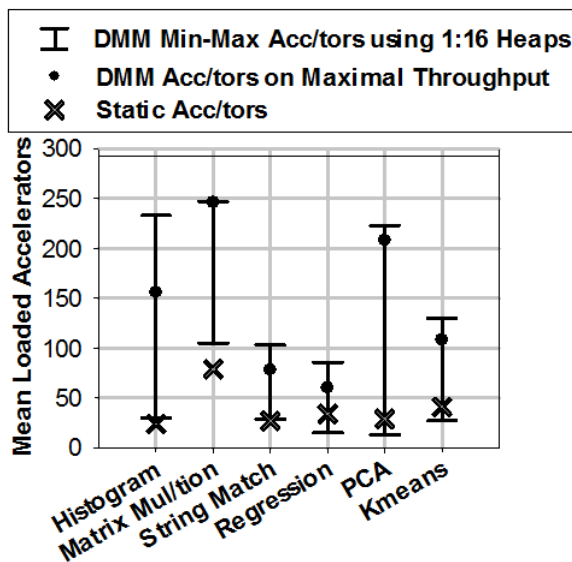


Σχήμα 5.8 Κανονικοποιημένη καθυστέρηση ανά επιταχυντή, λόγω μηχανισμών ΔΔΜ, αναφορικά προς τον αριθμό των χρησιμοποιούμενων σωρών.

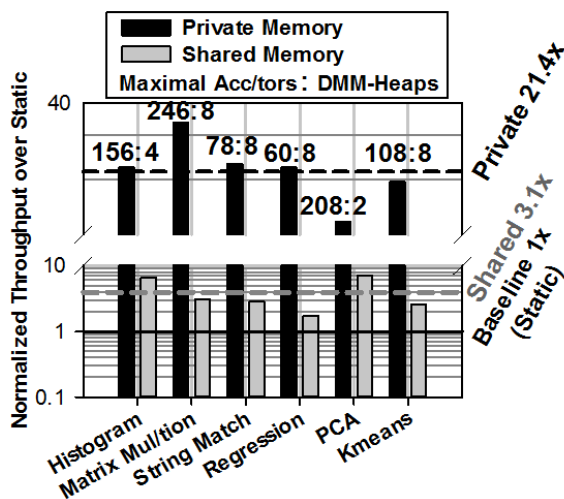
χρησιμοποιείται μόνο ένας σωρός, ενώ μειώνεται σε 10x, 4,7x, 2,3x και 1,2x κάθε φορά που οι σωροί διπλασιάζονται. Η αύξηση του σωρών επιτρέπει υψηλότερο επίπεδο παραλληλισμού, δεδομένου ότι λιγότεροι επιταχυντές μοιράζονται την ίδια σωρό. Ωστόσο, ακόμη και στην περίπτωση που κάθε επιταχυντής έχει αποκλειστικά το δικό του μοναδικό σωρό, π.χ. 16-επιταχυντές και 16-σωροί, υφίσταται ακόμη επιβάρυνση καθυστέρησης της τάξης του 1,2x, λόγω της εσωτερικής λειτουργίας της ΔΔΜ, δηλαδή του έλεγχος free-list και της εκτέλεσης του αλγορίθμου αναζήτησης first-fit. Ωστόσο, το πλαίσιο DMM-HLS επιτρέπει την αποτελεσματική ρύθμιση του συστήματος και της βιβλιοθήκης ΔΔΜ, η οποία συμβιβάζει την ικανότητα επιτάχυνσης και τη κοινή χρήση σωρών.

Στη συνέχεια αξιολογούνται οι πρακτικές βελτιώσεις επιδόσεων του προτεινόμενου πλαισίου DMM-HLS με δύο τρόπους. Αρχικά, αξιολογείται η επεξεργαστική ισχύς, δηλαδή ο αριθμός των επιταχυντών που μπορεί να προγραμματιστούν στο FPGA ταυτόχρονα. Το Σχήμα 5.9 απεικονίζει τα κέρδη και τις απώλειες του πλαισίου DMM-HLS σε σύγκριση με το σενάριο στατικής δέσμεισης μνήμης. Οι μπάρες εκφράζουν το ελάχιστο/μέγιστο αριθμό επιταχυντών που μπορούν να προγραμματιστούν στο FPGA με τη χρήση 1:16 σωρών. Όπως φαίνεται, το προτεινόμενο πλαίσιο DMM-HLS επιτυγχάνει τη σύσταση αρχιτεκτονικών λύσεων πολλαπλών επιταχυντών με 3,8x περισσότερους επιταχυντές, κατά μέσο όρο (έως 9,7x για τον αλγόριθμο Histogram), σε σύγκριση με τη στατική ΔΜ. Τα υψηλά κέρδη επεξεργαστικής ισχύος προέρχονται από τη χρήση μιας ενός σωρού, εφόσον αυτή η σύσταση παρέχει την ελάχιστη δυνατή επιβάρυνση, αναφορικά προς τους πόρους που καταναλώνονται από τον διαχειριστή ΔΔΜ. Όσο αυξάνεται ο αριθμός των σωρών, απαιτούνται περισσότεροι πόροι για την υλοποίηση των αντίστοιχων διαχειριστών ΔΔΜ και επομένως μειώνεται η δυναμική των επιταχυντών (μέγιστο αριθμός επιταχυντών), π.χ. η χρήση 16 σωρών επιτυγχάνει μέση αύξηση επιταχυντών κατά μόλις 1,7x.

Σημειώνεται ότι αποκλειστικά η αύξηση του αριθμού των επιταχυντών δεν



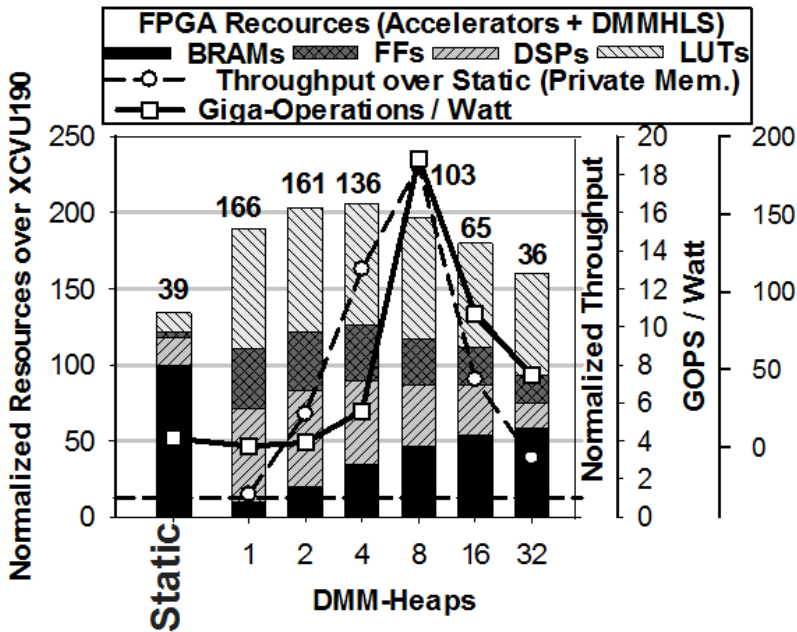
Σχήμα 5.9 Σύγκριση επεξεργαστικής ισχύος μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔΜΜ DMM-HLS.



Σχήμα 5.10 Σύγκριση ρυθμαπόδοσης μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔΜΜ DMM-HLS.

συνεπάγεται αντίστοιχα κέρδη απόδοσης. Το Σχήμα 5.10 απεικονίζει τον αριθμό των επιταχυντών, στους οποίους το σύστημα εμφανίζει συνολικά μέγιστη από-

δοση, ως προς την ρυθμαπόδοση⁵. Το Σχήμα 5.10 απεικονίζει την κανονικοποιημένη απόδοση σε σχέση με τη στατική ΔΜ για όλες τις εφαρμογές επαλήθευσης. Οι παρατηρούμενες διαφορές στην απόδοση και την επεξεργαστική ισχύ προέρχεται από τις διαφορετικές απαιτήσεις πόρων και χαρακτηριστικά φόρτου εργασίας των εφαρμογών. Θεωρούμε δύο σενάρια λειτουργίας αναφορικά προς το τρόπο εισαγωγής των δεδομένων: i) ιδιωτική μνήμη, δηλαδή οι επιταχυντές αρχικοποιούνται με στοιχεία από μεμονωμένο και μοναδικό χώρο μνήμης, και ii) κοινόχρηστη μνήμη, δηλαδή οι επιταχυντές επεξεργάζονται τα ίδια δεδομένα. Ένα τέτοιο σενάριο κοινόχρηστης μνήμης είναι η εύρεση διαφορετικών αλφαριθμητικών στο ίδιο έγγραφο με τον αλγόριθμο String Match. Τα αποτελέσματα αναφέρουν την διαμόρφωση που παρέχει τη μέγιστη απόδοση για κάθε εφαρμογή. Η διαμόρφωση συμβολίζεται ως επιταχυντές: αριθμός σωρών στην κορυφή της κάθε στήλης. Μετρήθηκε μέση αύξηση της απόδοσης κατά 21,4x με χρήση ιδιωτικής μνήμης και κατά 3,1x με χρήση κοινόχρηστης μνήμης, σε σχέση με τη στατική ΔΜ που προσφέρει το εργαλείο Vivado-HLS.



Σχήμα 5.11 Σύγκριση επιβάρυνσης πόρων σε σχέση με τη ρυθμαπόδοση και την ενέργεια, μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔΜM DMM-HLS.

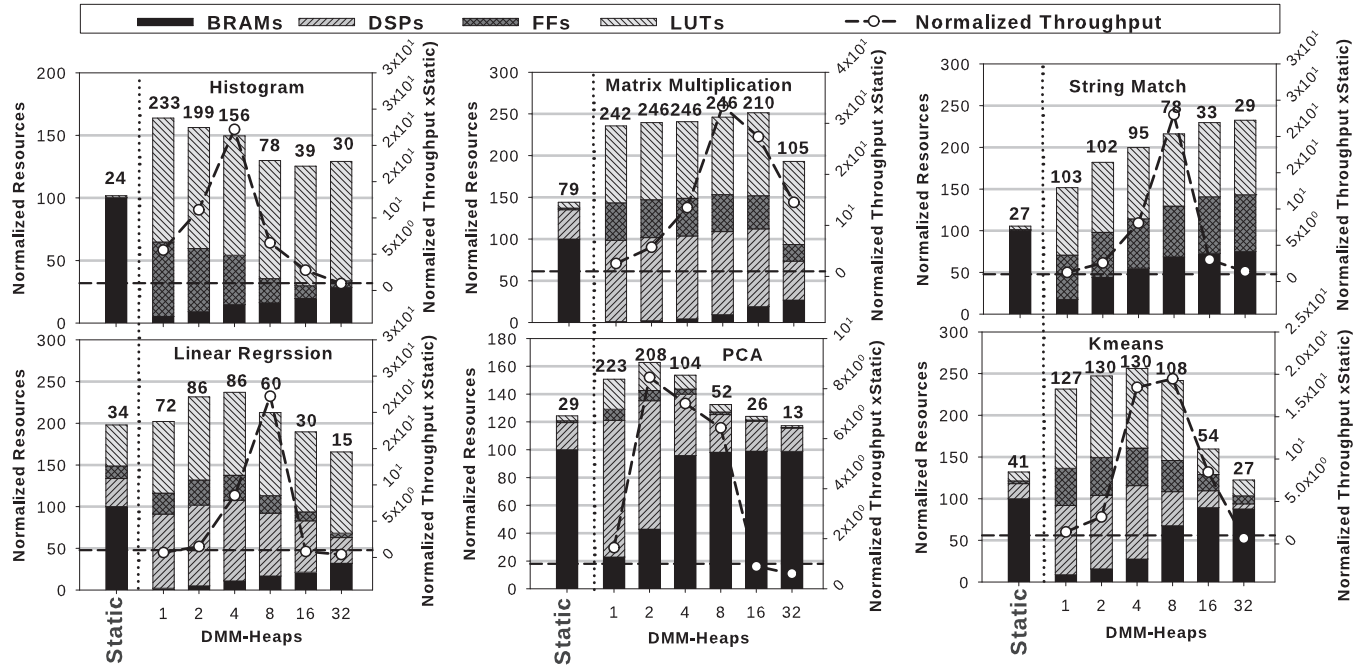
Το Σχήμα 5.11 απεικονίζει συγκεντρωτικά το μέσο όρο των αποτελεσμάτων, όσον αφορά τον αριθμό των επιταχυντών, τη χρήση των πόρων υλικού και τη ρυθμαπόδοση. Εξετάζεται τόσο η στατική δέσμευση ΔΜ όσο και η δυναμική, με αριθμό σωρών που κυμαίνονται από 1 έως 32, αντίστοιχα. Ο αριστερός κατακό-

⁵Η ρυθμαπόδοση υπολογίζεται ως το μέγεθος του φόρτου εργασίας (σε μέγεθος δεδομένων Mbytes) ως προς τη καθυστέρηση επεξεργασίας αυτού (σε μονάδα χρόνου us).

ρυθος άξονας αναφέρεται στους κανονικοποιημένους πόρους υλικού του FPGA⁶. Ο μέσος όρος της επεξεργαστικής ισχύος (δηλαδή ο μέγιστος αριθμός προγραμματιζόμενων επιταχυντών) για κάθε διαμόρφωση αναφέρεται στην κορυφή κάθε κατακόρυφης μέτρησης. Ο δεξιός κατακόρυφος άξονας αναφέρεται στην κανονικοποιημένη ρυθμαπόδοση σε σχέση με τη στατική ΔΜ, η οποία απεικονίζεται ως μία διακεκομμένη οριζόντια γραμμή. Ακολουθώντας την ίδια τάση με την ρυθμαπόδοση, η ενεργειακή απόδοση αποτιμάται επίσης σε όρους Γίγα-πράξεων ανά δευτερόλεπτο / Watt (GOPs/W). Ενώ η διαμόρφωση ενός σωρού προσφέρει το μεγαλύτερο αριθμό επιταχυντών (3,8x), εντούτοις, παρουσιάζει χαμηλή ρυθμαπόδοση δεδομένου ότι οι προσπελάσεις μνήμης εκτελούνται σειριακά. Με την προσθήκη περισσότερων σωρών, το συνολικό σύστημα εμφανίζει υψηλότερη απόδοση μέχρι το σημείο που οι επιπλέον πόροι που δεσμεύονται από τους πολλαπλούς σωρούς προκαλούν τη μείωση του αριθμού των επιταχυντών. Η επιβάρυνση σε πόρους υλικού οφείλεται στους ελεγκτές ΔΔΜ των σωρών και στην εκτεταμένη προγραμματιστική διασύνδεση των επιταχυντών. Η επιλογή μίας σωρού, επιβαρύνει κατά μέσο όρο το συνολικό σύστημα με 0,3% FFs και 1,2% LUTs σε σχέση με τις στατική ΔΜ. Η επιβάρυνση κλιμακώνεται έως 10,7% FFs και 55,2% LUTs για την περίπτωση των 32 σωρών. Ωστόσο, λαμβάνοντας υπόψη συνολικά τη χρήση του ΔΔΜ καθώς και των επιταχυντών υλικού, η επιβάρυνση στη μνήμη από τη ΔΔΜ είναι περίπου 60%. Δηλαδή φαίνεται ότι οι υπόλοιποι πόροι υπερβαίνουν το ανώτερο όριο χρήσης του FPGA, καταλαμβάνοντας κατά μέσο όρο 6,3x, 17,6x και 29,7x DSPs, FFs και LUTs αντίστοιχα. Το Σχήμα 5.12 απεικονίζει αναλυτικά τα συγκεντρωτικά αποτελέσματα για όλες τις εφαρμογές αξιολόγησης.

Σε αυτή την ενότητα παρουσιάστηκε μία αρχιτεκτονική κλιμάκωσης για τα σύγχρονα υπολογιστικά συστήματα, η οποία προσφέρει ευελιξία και ευκολία προγραμματισμού, μέσω χρήσης τεχνικών σχεδιασμού High-level Synthesis. Δείχθηκε ότι η on-chip μνήμη αποτελεί αποτρεπτικό παράγοντα στη σχεδίαση συστημάτων πολλαπλών επιταχυντών, οδηγώντας σε προβλήματα υπό-αξιοποίησης των επεξεργαστικών πόρων των σύγχρονων συσκευών FPGAs. Ως λύση, προτάθηκε η ενσωμάτωση τεχνικών δυναμικής διαχείρισης μνήμης σε σύγχρονα βιομηχανικά εργαλεία HLS, προκειμένου να αποφευχθεί η στατική δέσμευση μνήμης. Η προτεινόμενη προσέγγιση αξιολογήθηκε εκτεταμένα σε σύγχρονες εφαρμογές υψηλών επεξεργαστικών απαιτήσεων και αποδείχθηκε ότι η υιοθέτησή της επιτυγχάνει αξιολόγηση αύξηση υπολογιστικής ισχύος και ρυθμαπόδοσης.

⁶Κάθε κατακόρυφη γραμμή περιέχει το συσσωρευμένο συνολικό ποσοστό των τεσσάρων πόρους υλικού σε συσκευές FPGAs (Brams, DSPs, FFs και LUTs). Συνεπώς η μέγιστη θεωρητική τιμή του αριστερού άξονα, στο διάγραμμα (γ) του Σχήματος 5.11, είναι 400 %.



Σχήμα 5.12 Σύγκριση επιβάρυνσης πόρων σε σχέση με τη ρυθμιστική και την ενέργεια, μεταξύ στατικής ΔΜ και σεναρίων χρήσης του πλαισίου ΔΜ DMM-HLS, για όλες τις εφαρμογές αξιολόγησης.

5.2. Κλιμάκωση συστημάτων πολλαπλών επιταχυντών σε σταθμούς εργασίας

5.2.1. Εισαγωγή

Στην προηγούμενη ενότητα παρουσιάστηκε μία αρχιτεκτονική κλιμάκωσης της επεξεργαστικής ισχύος, με τη χρήση πολλαπλών επιταχυντών υλικού σε FPGAs. Το προτεινόμενο πλαίσιο εφαρμόστηκε σε μεμονωμένες εφαρμογές μεγάλης υπολογιστικής πολυπλοκότητας και απέδειξε την υπεροχή του σε σχέση με τις υφιστάμενες μεθοδολογίες και εργαλεία σχεδιασμού. Σε αυτή την εργασία αναλύεται μία παρόμοια προσέγγιση για σταθμούς εργασίας (Workstations). Σκοπός της παρούσας έρευνας είναι η διερεύνηση καταλληλότητας των συστημάτων πολλαπλών επιταχυντών σε πραγματικά σενάρια λειτουργίας συστημάτων μεγάλης κλίμακας.

Οι πρόσφατες απαιτήσεις των εφαρμογών έχουν καθιερώσει τη σύσταση νέων τεχνολογικών όρων, προκειμένου να περιγράψουν τις ιδιαιτερότητες τους. Μεταξύ αυτών, οι όροι “Big Data” και “Internet-of-Things” έχουν επικρατήσει και περιγράφουν τις συνεχώς αυξανόμενες απαιτήσεις των εφαρμογών σε μέγεθος δεδομένων επεξεργασίας και επικοινωνίας. Αυτές οι απαιτήσεις αυξάνουν την πολυπλοκότητα επεξεργασίας στους σταθμούς εργασίας (Workstations) και στα κέντρα δεδομένων (Data Centers). Η αύξηση του όγκου πληροφορίας και επεξεργασίας σε αυτά τα τερματικά κέντρα έχει επίσης ως αποτέλεσμα την αύξηση της κατανάλωσης ηλεκτρικής ενέργειας. Οι επεξεργαστές των διακομιστών πρέπει συνεχώς να κλιμακώνουν την επεξεργαστική τους απόδοση, διατηρώντας χαμηλά επίπεδα κατανάλωσης ενέργειας. Σήμερα, μια από τις κύριες προκλήσεις για τους σταθμούς εργασίας και τα κέντρα δεδομένων είναι η κατανάλωση ενέργειας, καθώς αυτή αποτιμάται σε ποσοστό 45% της συνολικής κατανάλωσης ηλεκτρικής ενέργειας παγκοσμίως [26].

Συνεπώς, η ερευνητική κοινότητα καλείται να σχεδιάσει νέες αρχιτεκτονικές, οι οποίες αυξάνουν την απόδοση των διακομιστών, αλλά ταυτόχρονα είναι ενεργειακά αποδοτικές. Οι συσκευές FPGAs αποτελούν ιδανικές πλατφόρμες υλοποίησης, καθώς εξυπηρετούν ταυτόχρονα και τις δύο αυτές απαιτήσεις. Ωστόσο, το κύριο μειονέκτημα της χρήσης FPGAs στα κέντρα δεδομένων είναι η υψηλή πολυπλοκότητα προγραμματισμού. Σε αυτή την εργασία, παρουσιάζεται ένα νέο πλαίσιο που επιτρέπει την απρόσκοπτη ανάπτυξη επιταχυντών υλικού για κέντρα δεδομένων και σταθμούς εργασίας. Το πλαίσιο στηρίζεται στη χρήση της τεχνικής προγραμματισμού και εκτέλεσης εργασιών MapReduce. Το πλαίσιο αυτό είναι ένα ευρέως χρησιμοποιούμενο API για εφαρμογές σε συστήματα διακομιστών, όπως τα κέντρα δεδομένων και οι σταθμοί εργασίας, καθώς υποστηρίζει τον διαμοιρασμό των επεξεργαστικών διεργασιών σε κατανεμημένα συστήματα για την επεξεργασία μεγάλου όγκου δεδομένων. Στη παρούσα εργασία αναπτύσσεται μία ροή εργασίας για συστήματα πολλαπλών επιταχυντών που αξιοποιούν το πλαίσιο MapReduce. Στόχος της εργασίας είναι η σύγκριση της απόδοσης και της ενεργειακής αποδοτικότητας του προτεινόμενου πλαισίου με σύγχρονους πολυπύρηνους επεξεργαστές διακομιστών. Οι κύριες συνεισφο-

ρές της παρούσας εργασίας είναι οι εξής:

- αναπτύσσεται μία καινοτόμα συνθέσιμη αρχιτεκτονική MapReduce,
- σχεδιάζονται επιταχυντές υλικού για τυπικές εφαρμογές MapReduce,
- το προτεινόμενο πλαίσιο αξιολογείται σε πραγματικό συστημικό σενάριο, με μία σύγχρονη συσκευή Virtex7 FPGA,
- επιδεικνύονται κέρδη ρυθμαπόδοσης κατά 4.3X,
- επιδεικνύονται κέρδη κατανάλωσης ενέργειας κατά δύο τάξεις μεγέθους.

Η ερευνητική κοινότητα έχει ήδη ερευνήσει συστήματα επιταχυντών υλικού για εφαρμογές διακομιστών μέσω του πλαισίου MapReduce. Στην εργασία [27] παρουσιάζεται ένα πλαίσιο MapReduce για επαναδιαμορφούμενες αρχιτεκτονικές, το οποίο ωστόσο, υλοποιείται ως ένας εξειδικευμένος RTL-σχεδιασμός για την υλοποίηση αποκλειστικά της εφαρμογής RankBoost. Αν και η βασική αρχιτεκτονική αυτής της εργασίας είναι αρκετά σχετική με την προτεινόμενη, σημειώνεται ότι η προτεινόμενη προσέγγιση υλοποιείται σε επίπεδο C/C++ και συντίθεται αυτόματα σε επίπεδο RTL χρησιμοποιώντας εργαλεία HLS. Αντίθετα, στην εργασία [27] οι εσωτερικές διαδικασίες του πλαισίου MapReduce υλοποιούνται αποκλειστικά σε επίπεδο γλώσσας περιγραφής υλικού και επομένως κάθε νέα εφαρμογή που υλοποιείται με το συγκεκριμένο πλαίσιο απαιτεί εκ νέου προσεκτικό και χρονοβόρο σχεδιασμό σε γλώσσα HDL. Στην εργασία [28] προτείνεται ένα πλαίσιο MapReduce για συσκευές FPGA. Αποτελείται από ένα σύνολο κόμβων επεξεργασίας, κάθε ένας από τους οποίους αποτελείται από επεξεργαστικά στοιχεία γενικού σκοπού καθώς και εξειδικευμένα. Παρά το γεγονός ότι η προσέγγιση αυτή προσφέρει μεγάλη ευελιξία και βελτιστοποίηση κατά το χρόνο εκτέλεσης, εξακολουθεί να αυξάνει τη δυσκολία προγραμματισμού της ετερογενούς αρχιτεκτονικής του κάθε κόμβου, ενώ οι νέες εφαρμογές πρέπει να σχεδιαστούν και να απεικονιστούν αποκλειστικά για αυτή την υβριδική αρχιτεκτονική των πολλαπλών ετερογενών κόμβων. Επίσης το πλαίσιο MapReduce έχει διερευνηθεί ως προς την υλοποίησή του σε πλατφόρμες GPU [29]. Ωστόσο, οι αρχιτεκτονικές GPU προσφέρουν συσχετισμένη πρόσβαση στη μνήμη, γεγονός που τις καθιστά μη-ιδανικές πλατφόρμες για τη διαχείριση περίπλοκων δομών δεδομένων. Επίσης η αρχιτεκτονική SIMT που διαθέτουν, περιορίζει τη κλιμάκωση απόδοσης υπολογισμού σε ακανόνιστες (irregular) εφαρμογές [27]. Παράλληλα, η εργασία [30] υιοθετεί μια υβριδική αρχιτεκτονική που συνδυάζει πλατφόρμες GPU και FPGA για την υλοποίηση ενός πλαισίου MapReduce. Η συγκεκριμένη αρχιτεκτονική προτείνει το προγραμματισμό της ροής ελέγχου στη CPU και αξιοποιεί τις πλατφόρμες GPU και FPGA ως επιταχυντές επεξεργασίας. Η εργασία [31] παρουσιάζει ένα πλαίσιο MapReduce βασισμένο σε μία ενσωματωμένη πλατφόρμα πολυπύρηνου δικτύου-σε-ψηφίδα (NoC). Ωστόσο, ανεξάρτητα από το μέσο υλοποίησης, μία πρόσφατη έρευνα για τις εφαρμογές που αξιοποιούν το πλαίσιο MapReduce αναφέρει ότι η ιδανική πλατφόρμα για το πλαίσιο MapReduce προσομοιάζει ένα σύστημα διαχείρισης βάσεων δεδομένων, το

οποίο διαθέτει κλιμακούμενα και ευέλικτα παράλληλα επεξεργαστικά στοιχεία για την ανάλυση πολύπλοκων τύπων δεδομένων [32]. Στην εργασία [33] η εταιρεία Microsoft χρησιμοποιεί FPGAs προκειμένου να αυξήσει την απόδοση της εφαρμογής κατάταξης σελίδας (Page Ranking). Συγκεκριμένα, χρησιμοποιήθηκε μία διάταξη από 1.632 διακομιστές, στους οποίους αναπτύχθηκε κατάλληλη διεπαφή επιτάχυνσης μέσω FPGAs. Το πλαίσιο αυτό αξιολογήθηκε στη μηχανή αναζήτησης Bing. Αναφέρθηκαν βελτιώσεις απόδοσης του κάθε διακομιστή κατά ένα συντελεστή 95%. Ωστόσο, πρακτικά σήμερα, η χρήση πλατφορμών FPGAs σε κέντρα δεδομένων και σταθμούς εργασίας περιορίζεται κυρίως λόγω της υψηλής πολυπλοκότητας προγραμματισμού των FPGAs.

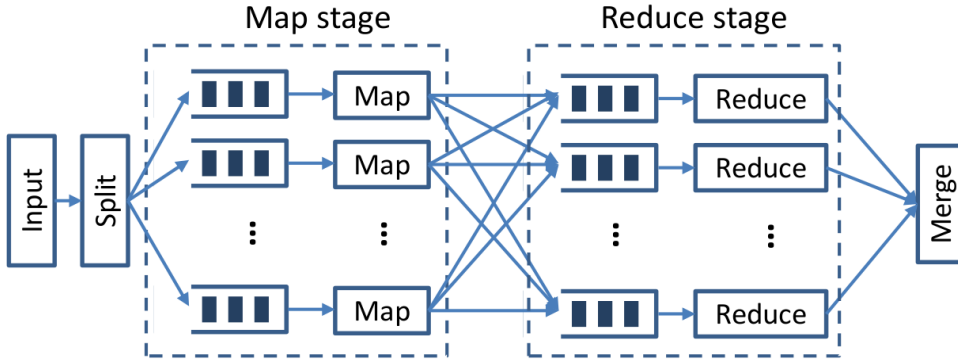
Στην εργασία [34] περιγράφεται ένας επιταχυντής υλικού για το πλαίσιο MapReduce, ο οποίος συνδυάζεται αποδοτικά με επεξεργαστές ARM προκειμένου να σχηματίσει πλήρη προγραμματιζόμενες πλατφόρμες. Στη συγκεκριμένη εργασία επιλέγεται ως πλατφόρμα υλοποίησης η συσκευή Zynq-7000 All Programmable SoC, η οποία περιέχει ένα διπύρηνο Cortex-A9 επεξεργαστή και επαναπρογραμματιζόμενη λογική στην ίδια ψηφίδα. Ο προτεινόμενος επιταχυντής χρησιμοποιείται για την επιτάχυνση της λειτουργίας Reduce. Η λειτουργία Map, καθώς και η δημιουργία των ενδιάμεσων ζευγαριών κλειδιών/τιμών, εκτελούνται στον ενσωματωμένο επεξεργαστή. Η αξιολόγηση της συγκεκριμένης αρχιτεκτονικής έδειξε ότι είναι εφικτή επιτάχυνση κατά 1.8x σε τυπικές εφαρμογές MapReduce. Χρησιμοποιώντας αυτή την αρχιτεκτονική ως βάση, η παρούσα εργασία προτείνει την πλήρη αποδέσμευση των μονοπατιών δεδομένων του πλαισίου MapReduce σε μεμονωμένους διαύλους, οι οποίοι επεξεργάζονται από ιδιωτικά επεξεργαστικά στοιχεία. Με αυτόν το τρόπο αποφεύγεται η χρήση ενός κεντρικού επεξεργαστή-επόπτη (supervisor on-board CPU). Ωστόσο, σε αυτή την αρχιτεκτονική υπάρχει ανάγκη ολικής μετάβασης του κώδικα C/C++ του πλαισίου MapReduce σε επίπεδο RTL. Στη παρούσα εργασία χρησιμοποιούνται εργαλεία HLS για αυτό το στάδιο και κατάλληλοι μετασχηματισμοί κώδικα, ώστε να είναι δυνατή η βελτιστοποίηση της αρχιτεκτονικής καθώς και η άμεση υλοποίηση νέων εφαρμογών στο πλαίσιο.

5.2.2. Η αρχιτεκτονική HLSMapReduceFlow

5.2.2.1. Το πλαίσιο Phoenix MapReduce

Ως αρχική βάση αρχιτεκτονικής επιλέγεται το πλαίσιο ανοικτού κώδικα Phoenix MapReduce [25]. Στο πλαίσιο αυτό, οι χρήστες καθορίζουν μία λειτουργία Map, η οποία επεξεργάζεται ένα ζεύγος κλειδιού/τιμής εισόδου, προκειμένου να δημιουργήσει ένα σύνολο ενδιάμεσων ζευγών κλειδιού/τιμής, καθώς και μία λειτουργία Reduce, η οποία συγχωνεύει όλες τις ενδιάμεσες τιμές που σχετίζονται με το ίδιο ενδιάμεσο κλειδί. Τέλος, κατά το τελευταίο στάδιο επεξεργασίας, συγχωνεύονται όλα τα ζεύγη κλειδιού/τιμής. Τα προγράμματα που είναι γραμμένα σε αυτό το μοτίβο λειτουργικού προγραμματισμού (functional programming) παραλληλοποιούντε αυτόματα μέσω του πλαισίου Phoenix και εκτελούνται σε ένα σύστημα πολλαπλών κόμβων υπολογισμού. Το σύστημα επεξεργασίας κατά το χρόνο εκτέλεσης φροντίζει για την εισαγωγή των δεδομένων, τον προγραμματισμό εκτέλεσης των διεργασιών σε ένα σύνολο κόμβων υπολογισμού, παρέχει

ανοχή σε βλάβες, και διαχειρίζεται την επικοινωνία μεταξύ των επεξεργαστικών κόμβων. Αυτό το επίπεδο αφαίρεσης επιτρέπει στους προγραμματιστές δίχως εμπειρία παράλληλων και κατανεμημένων συστημάτων, να χρησιμοποιούν εύκολα τους πόρους ενός μεγάλου κατανεμημένου συστήματος.



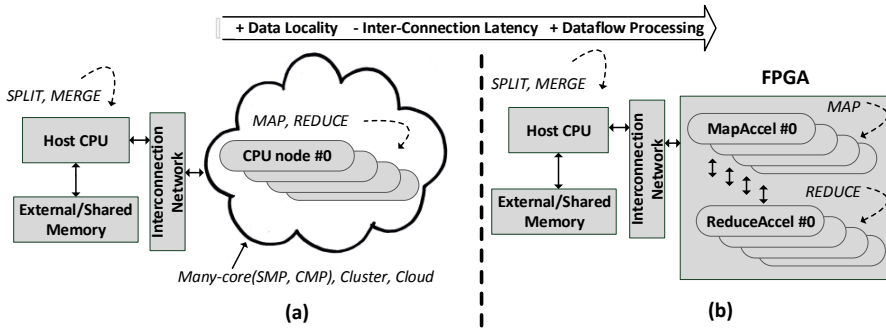
Σχήμα 5.13 Το πλαίσιο MapReduce.

Το Σχήμα 5.13 απεικονίζει τη βασικά ροή δεδομένων του πλαισίου MapReduce, κατά το χρόνο εκτέλεσης. Βασική υπομονάδα του πλαισίου αποτελεί ο χρονοπρογραμματιστής, ο οποίος ελέγχεται από τον προγραμματιστή, μέσω κατάλληλου κώδικα διεπαφής. Ο χρονοπρογραμματιστής δημιουργεί και διαχειρίζεται τα νήματα εκτέλεσης για τις λειτουργίες Map και Reduce. Επίσης, διαχειρίζεται τα ρυθμιστικά διαλύματα που χρησιμοποιούνται για την επικοινωνία εργασίας. Μετά το στάδιο της αρχικοποίησης, ο χρονοπρογραμματιστής καθορίζει τον αριθμό των πυρήνων επεξεργασίας. Για κάθε πυρήνα, δημιουργείται ένα νήμα επεξεργασίας, στο οποίο εκχωρείται δυναμικά μία διεργασία Map ή Reduce.

5.2.2.2. Επιτάχυνση σε επίπεδο ροής δεδομένων

Όπως προαναφέρθηκε, η βασική αρχιτεκτονική της προτεινόμενης εργασίας εμπνεύστηκε από την εργασία [34]. Οι συγγραφείς σε αυτή την εργασία ανέπτυξαν έναν επαναδιαμορφούμενο επιταχυντή MapReduce, ο οποίος επιταχύνει τη λειτουργία Reduce, ενώ υλοποιεί τη λειτουργία Map καθώς και τη δημιουργία των ενδιάμεσων ζευγών κλειδιών/τιμών στον ενσωματωμένο επεξεργαστή. Χρησιμοποιώντας ως βάση αυτή την αρχιτεκτονική, δημιουργείται η κατάλληλη αρχιτεκτονική για την επιτάχυνση της λειτουργίας Map. Το Σχήμα 5.1 απεικονίζει τη βασική διαφοροποίηση της προτεινόμενης αρχιτεκτονικής σε σύγκριση με την υφιστάμενη προσέγγιση.

Στη τυπική εκτέλεση ενός πλαισίου MapReduce, οι λειτουργίες Map και Reduce εκτελούνται μέσω λογισμικού, ως νήματα εκτέλεσης στους πυρήνες CPU της υποστηριζόμενης αρχιτεκτονικής MapReduce (Σχήμα 5.14(a)). Η επιτυχία μιας τέτοιας αρχιτεκτονικής βασίζεται στη διαθεσιμότητα μιας μεγάλης κοινόχρηστης μνήμης, η οποία επιτρέπει την άμεση επικοινωνία με την ελάχιστη αντιγραφή δεδομένων, προκειμένου τα δεδομένα να φτάσουν από την πηγή εισόδου (μέσο αποθήκευσης, δίκτυο, κτλ.) στους τελικούς επεξεργαστικούς πό-

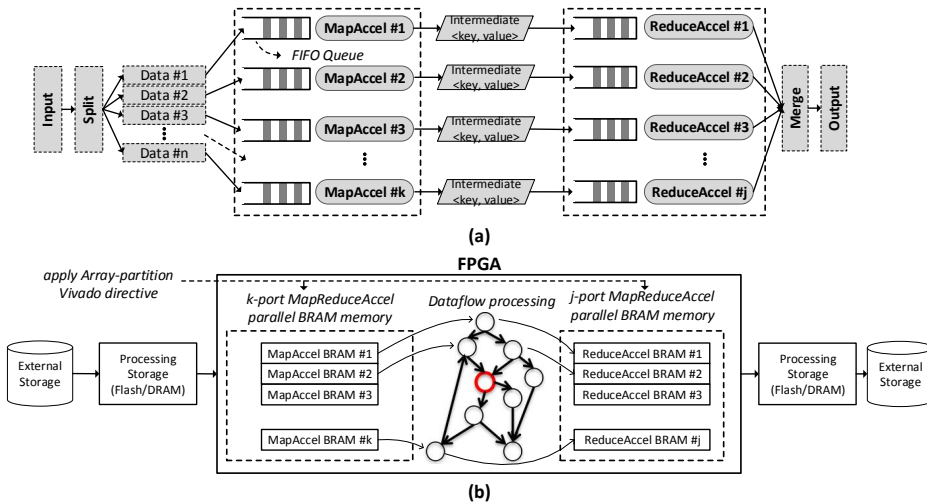


Σχήμα 5.14 Τοπολογία αρχιτεκτονικής για a) υφιστάμενο τυπικό πλαίσιο MapReduce και b) προτεινόμενο πλαίσιο HLSMapReduceFlow.

ρους. Επιπλέον, κατά το χρόνο εκτέλεσης, σημαντικός είναι ο ρόλος του χρονο-προγραμματιστή, ο οποίος είναι υπεύθυνος για την κατανομή των διεργασιών σε όλες τους διαθέσιμους επεξεργαστικούς πόρους, προκειμένου να επιτευχθεί ισορροπία φορτίου και μεγιστοποίηση απόδοσης.

Σε αυτή την εργασία, προτείνεται η δημιουργία προσαρμοσμένων επιταχυντές της λειτουργίας Map, οι οποίοι εκμεταλλεύονται την τοπικότητα (locality) των αναφορών και επομένως, εξαλείφεται η ανάγκη για αρχιτεκτονικές μεγάλης κοινόχρηστης μνήμης και για κατανεμημένα συστήματα. Σε αντίθεση με τον αυθαίρετο διαχωρισμό των δεδομένων εισόδου σε πολλαπλά υποσύνολα επεξεργασίας, το προτεινόμενο πλαίσιο προτείνει το διαχωρισμό της εισόδου ανάλογα με τη ροή επεξεργασίας των δεδομένων της εφαρμογής, κατά τρόπο που τα διαιρεμένα υποσύνολα δεδομένων προς επεξεργασία, επεξεργάζονται από ανεξάρτητους διακριτούς επιταχυντές υλικού. Χρησιμοποιώντας αυτή την προσέγγιση, είναι δυνατή η αύξηση της απόδοσης του συστήματος μέσω α) της αυξανόμενης τοπικότητας των αναφορών στα δεδομένα, β) της μείωσης των ένδο-καθυστερήσεων στις διασυνδέσεις μεταξύ των επεξεργαστικών κόμβων και γ) της αύξησης του παραλληλισμού με την αξιοποίηση επεξεργασίας ροής δεδομένων (dataflow processing).

Η εργασία αυτή προτείνει την εύρεση του βέλτιστου σημείου επεξεργασίας ροής δεδομένων για κάθε εφαρμογή, μέσω της ανάλυσης του γράφου ροής ελέγχου (CFG), του γράφου ροής δεδομένων (DFG) και του χρονικού πλαισίου εγκυρότητας δεδομένων (variable liveness analysis - LA). Με βάση αυτές τις πληροφορίες, σχεδιάζονται οι επιταχυντές της λειτουργίας Map. Το Σχήμα 5.15(α) απεικονίζει τη βασική αρχιτεκτονική HLSMapReduceFlow, ενώ το Σχήμα 5.15(β) απεικονίζει τις καινοτόμες αρχιτεκτονικές εισηγήσεις της προτεινόμενης προσέγγισης. Πρώτον, η ενσωματωμένη μνήμη Bram του FPGA οργανώνεται σε ιδιωτικές υπομονάδες μνήμης. Κάθε υπομονάδα έχει τους δικούς της μοναδικούς διαύλους διευθύνσεων και δεδομένων, ενώ προσπελάζεται από ένα μόνο επεξεργαστικό κόμβο. Με αυτό το τρόπο, το σύστημα επιτρέπει την πλήρη παράλληλη και ταυτόχρονη λειτουργία των επεξεργαστικών κόμβων στο FPGA.

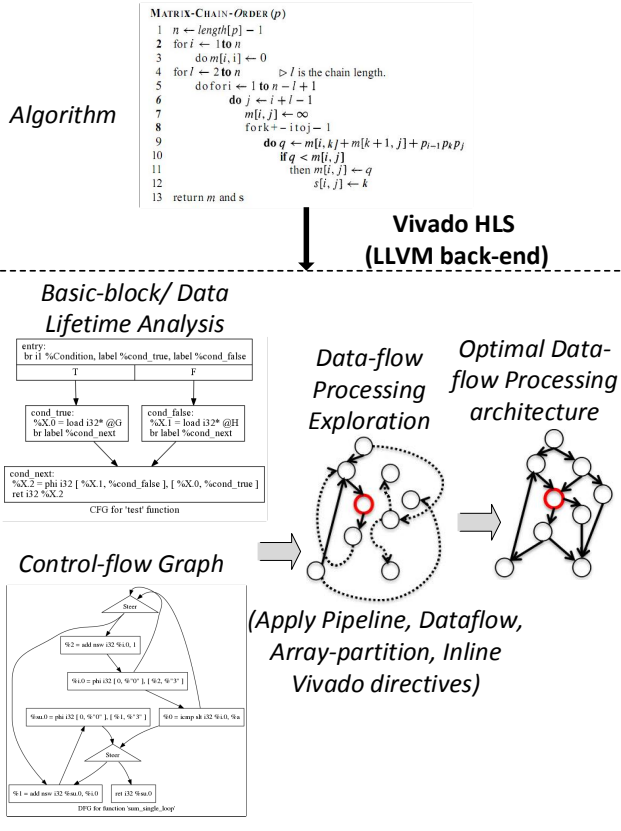


Σχήμα 5.15 Η αρχιτεκτονική HLS MapReduce Flow: Κάθε επεξεργαστικός dataflow-κόμβος εργάζεται επί αποκλειστικής μνήμης. Η συνολική μνήμη του συστήματος διαιρείται σε μνήμες k -port και j -port για έναν αριθμό νημάτων εκτέλεσης k -map και j -reduce.

Προκειμένου να υλοποιηθούν ανεξάρτητα μονοπάτια υπολογισμού, χρησιμοποιείται το εργαλείο Xilinx Vivado HLS. Εκτός από τη λειτουργία της σύνθεσης υψηλού επιπέδου, το εργαλείο αυτό παρέχει ένα πλήθος επιλογών αρχιτεκτονικής διερεύνησης, μέσω της χρήσης εξειδικευμένων ντιρεκτίβων προς τον προεπεξεργαστή του πηγαίου κώδικα. Σε αυτή την εργασία χρησιμοποιούνται οι ντιρεκτίβες *DATAFLOW*, *INLINE* και *ARRAY PARTITION*.

Αρχικά χρησιμοποιούνται οι ντιρεκτίβες *PARTITION*, *MAP*, *RESHAPE* προκειμένου να ανασχηματιστούν οι πίνακες μνήμης σύμφωνα με τον τρόπο που προσπελάζονται. Οι πίνακες μνήμης διαιρούνται σε μικρότερες υπομονάδες μνήμης, κάθε μία εκ των οποίων διαθέτει μοναδική και αποκλειστική διεπαφή. Αυτή η διαδικασία περιλαμβάνει τη δυνατότητα λεπτομερούς ανάλυσης των πινάκων μνήμης. Κάθε διεπαφή υπομονάδας μνήμης διαθέτει μοναδική θύρα πρόσβασης. Συνεπώς εξασφαλίζεται μέγιστη παραλληλία πρόσβασης. Ωστόσο η αυθαίρετη υποδιαίρεση της μνήμης μπορεί να δημιουργήσει προβλήματα στη δρομολόγηση και συνεπώς, στην ικανότητα φυσικής σχεδίασης στο FPGA. Όσο περισσότερο διαιρείται η μνήμη τόσο πιθανότερο είναι να μην υπάρχουν εξαρτήσεις δεδομένων ανάμεσα στα δεδομένα διαφορετικών υπομονάδων μνήμης. Ανάλογα με την τοπικότητα των αναφορών της εκάστοτε εφαρμογής, είναι δυνατή η εξαγωγή αντίστοιχου παραλληλισμού, σύμφωνα με τις επιλογές κατάτμησης της μνήμης εισόδου.

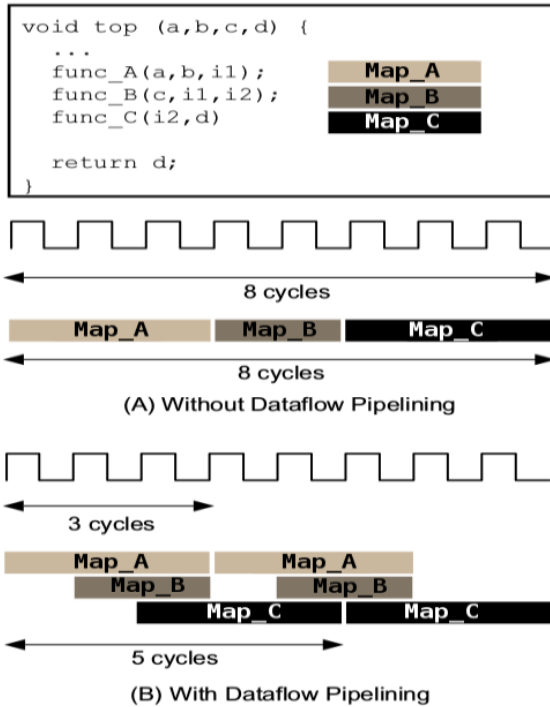
Εφόσον πραγματοποιηθεί η κατάτμηση της μνήμης εισόδου, ακολουθεί η εξερεύνηση της μικρο-αρχιτεκτονικής, μέσω του εργαλείου Vivado HLS, ακολουθώντας ένα μοντέλο υπολογισμού ροής δεδομένων (*dataflow*). Προτείνεται μία πλή-



Σχήμα 5.16 Εξερεύνηση ροής δεδομένων από τη ροή έλεγχου της εφαρμογής εισόδου, με το εργαλείο Vivado HLS.

ρος χωρική αρχιτεκτονική (spatial) για κάθε πυρήνα επεξεργασίας της λειτουργίας Map, μέσω της αναδρομικής χρήσης της ντιρεκτίβας inline του εργαλείου Vivado HLS, δηλαδή `#pragma AP inline recursive`. Παρόλο που η συγκεκριμένη ντιρεκτίβα οδηγεί σε αυξημένη χρησιμοποίηση των πόρων, εντούτοις, επιτρέπει την επίτευξη μέγιστης δυνατής παραλληλίας. Πρακτικά, η συγκεκριμένη ντιρεκτίβα ωθεί το εργαλείο σύνθεσης στη δημιουργία πολλαπλών αντιγράφων υλικού, κάθε φορά που υπάρχει ανάγκη επεξεργασίας του ίδιου κώδικα. Με αυτό το τρόπο, δημιουργούνται μοναδικές - και όχι διαμοιραζόμενες - ροές επεξεργασίας, ακόμα και για τις πιο απλές στοιχειώδεις πράξεις (λόγω αναδρομής).

Υστερα από την εφαρμογή των προηγούμενων βελτιστοποιήσεων, έχει υλοποιηθεί ένα σύνολο από παράλληλους επιταχυντές Map, οι οποίοι δεν έχουν κινές εξαρτήσεις σε δεδομένα και επεξεργαστικούς πόρους. Το τελευταίο βήμα περιλαμβάνει τον έλεγχο του τρόπου τροφοδότησης των δεδομένων σε αυτούς τους επιταχυντές. Το Σχήμα 5.16 απεικονίζει τη βασική ιδέα της προτεινόμε-



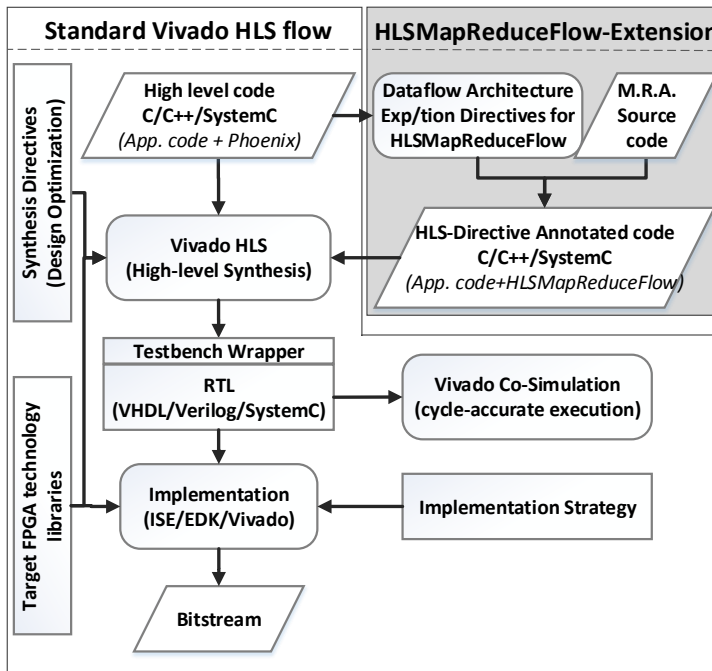
Σχήμα 5.17 Παράδειγμα χρήσης ντιρεκτίβας *DATAFLOW* (a) Ακολουθιακή εκτέλεση (b) Παράλληλη επικαλυπτόμενη επεξεργαστική ροή.

νης προσέγγισης. Ο κώδικας εισόδου συντίθεται από το εργαλείο Vivado HLS (back-end LLVM compiler) σε βασικές μονάδες υπολογισμού (basic blocks). Στη συνέχεια, μέσω της ντιρεκτίβας *#pragma AP dataflow* ωθείται μεταβολή της ροής επεξεργασίας σε ροή δεδομένων (dataflow). Πρακτικά, η σειριακή εκτέλεση διεργασιών του κώδικα C/C++ μετατρέπεται σε παράλληλη ροή εκτέλεσης πολλαπλών στοιχειωδών πράξεων, που συντίθενται σε πραγματικό υλικό στο FPGA. Η βελτιστοποίηση μέσω της ντιρεκτίβας *Dataflow*, που προσφέρεται από το εργαλείο Vivado HLS είναι μία πολύ αποτελεσματική διαδικασία, η οποία αυξάνει τη ρυθμαπόδοση του συστήματος. Όπως φαίνεται και στο Σχήμα 5.17, η βελτιστοποίηση *DATAFLOW* επιτρέπει τη χρονική επικάλυψη των διεργασιών, όταν αυτές δεν μοιράζονται δεδομένα ή/και επεξεργαστικούς πόρους, μειώνοντας τη καθυστέρηση επεξεργασίας και αυξάνοντας τη συστημική ρυθμαπόδοση.

5.2.2.3. Η μεθοδολογία του πλαισίου *HLSMapReduceFlow* με το εργαλείο *Vivado-HLS*

Το Σχήμα 5.18 απεικονίζει τη μεθοδολογία του προτεινόμενου πλαισίου *HLSMapReduceFlow* και την επαλήθευση ροής σχεδιασμού. Η σχεδιαστική ροή βασίζεται στο σύγχρονο βιομηχανικό εργαλείο Xilinx Vivado-HLS. Η επέκταση του πλαισίου *HLSMapReduceFlow* εφαρμόζεται σε υψηλό επίπεδο του πηγαιού κώδικα της εφαρμογής. Βασική προϋπόθεση είναι ο μετασχηματισμός του αρχικού πηγαιού κώδικα σε συνθέ-

σιμη μορφή. Αυτοί οι μετασχηματισμοί είναι αναγκαίοι λόγω των περιορισμών διαχείρισης του εργαλείου Vivado-HLS σε συγκεκριμένες τεχνικές κώδικα και τύπους δεδομένων, όπως η έλλειψη υποστήριξης δυναμικής διαχείρισης μνήμης, αριθμητικής δεικτών, πλήρους υποστήριξης του ANCI C κλπ. Επιπλέον, σε αυτό το στάδιο γίνεται η αρχιτεκτονική βελτιστοποίηση μέσω της εισαγωγής των κατάλληλων ντιρεκτίβων. Επί του παρόντος, αυτό το βήμα εκτελείται κατά μη-αυτόματο τρόπο, παρόλο που είναι δυνατή μία αυτοματοποιημένη ροή μετασχηματισμού του κώδικα. Ο μετασχηματισμένος κώδικας προσαυξάνεται από τις κλήσεις συναρτήσεων του πλαισίου HLSMapReduceFlow, δηλαδή *Emit_Intermediate_accelerator* (κλειδί, τιμή) και συντίθεται σε επίπεδο RTL, μέσω του εργαλείου Vivado HLS.



Σχήμα 5.18 Προτεινόμενη επέκταση του εργαλείου Vivado HLS για υποστήριξη του πλαισίου MapReduce σε πλατφόρμες FPGA.

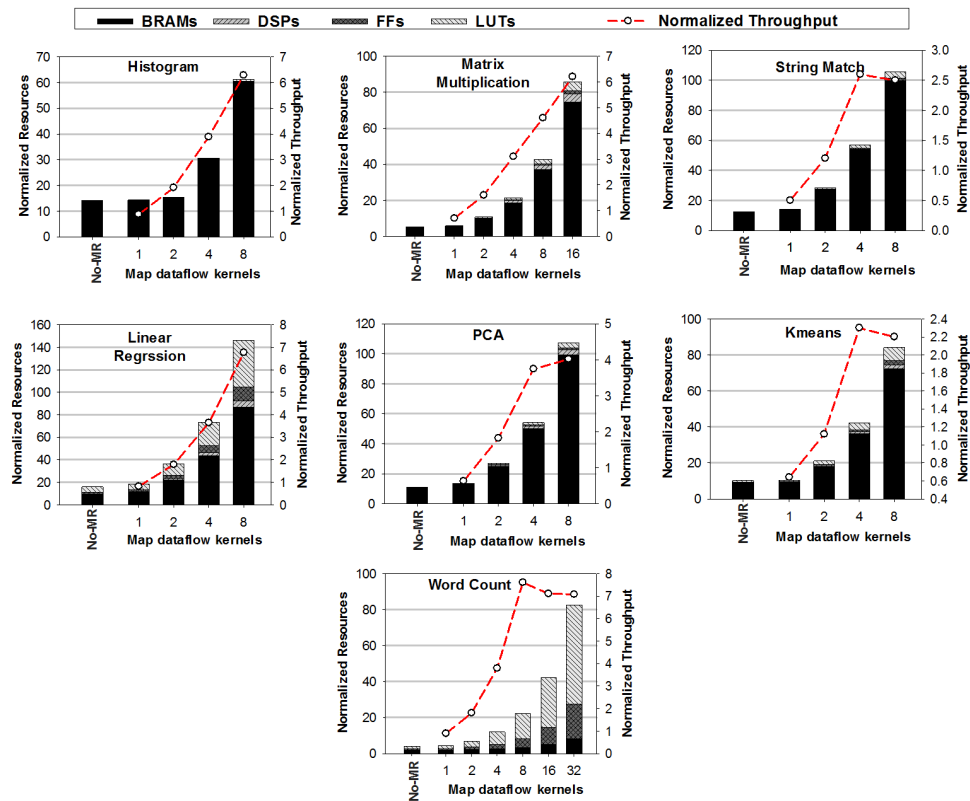
5.2.3. Πειραματικά αποτελέσματα

Αυτή η υποενότητα περιγράφει την πειραματική διάταξη που χρησιμοποιήθηκε για την αξιολόγηση του πλαισίου HLSMapReduceFlow, καθώς και τα αντίστοιχα αποτελέσματα. Χρησιμοποιήθηκαν οι εφαρμογές αξιολόγησης του πλαισίου Phoenix MapReduce [25]. Τα χαρακτηριστικά των εφαρμογών είναι ακριβώς ίδια με αυτά της ανάλυσης της προηγούμενης ενότητας 5.1 και απει-

κονίζονται λεπτομερώς στον Πίνακα 5.2.

Ο κύριος σκοπός της παρούσας μελέτης είναι η επιτάχυνση του πλαισίου MapReduce με τη χρήση συσκευών FPGAs για εφαρμογές σταθμών εργασίας. Η εργασία [25] αναφέρει ότι για τις χρησιμοποιούμενες εφαρμογές, η λειτουργία Map αποτελεί το 95% του συνολικού χρόνου εκτέλεσης του πλαισίου Phoenix. Συνεπώς, στη παρούσα έρευνα δίδεται μεγαλύτερη βαρύτητα στην επιτάχυνση της λειτουργίας Map μέσω FPGA, ενώ για τη λειτουργία Reduce χρησιμοποιείται μόνο μία διεργασία σε επίπεδο RTL, όπως αυτή που αναφέρεται στην εργασία [34]. Τα δεδομένα εισόδου αποθηκεύονται αποκλειστικά στις μνήμες BRAMs του FPGA. Αυτή η προσέγγιση θεωρεί ότι η μεταφορά δεδομένων σε αυτές τις μνήμη γίνεται από κάποιο σύστημα-επόπτη, ο οποίος τροφοδοτεί τα δεδομένα από τις συσκευές παραγωγής (π.χ. τερματικά δικτύου) στις μνήμες BRAMs.

Το Σχήμα 5.19 απεικονίζει τα συνολικά αποτελέσματα επιδόσεων κλιμάκωσης για κάθε εφαρμογή, με τη χρήση, ή μη, του πλαισίου HLSMapReduceFlow. Οι οριζόντιοι άξονες αναφέρουν το πλήθος των παράλληλων επιταχυντών της λειτουργίας Map που χρησιμοποιούνται. Ο αριθμός αυτός περιορίζεται από τους διαθέσιμους πόρους του FPGA, καθώς μεγαλύτερος αριθμός από επιταχυντές επιφέρει αντίστοιχη δέσμευση πόρων, ανάλογα με την υλοποιημένη εφαρμογή. Όπως προκύπτει από τη σύγκριση της αρχιτεκτονικής δίχως τη χρήση του πλαισίου MapReduce (No-MR), και τη χρήση ενός επιταχυντή Map, το πλαίσιο HLSMapReduceFlow εισάγει επιβάρυνση στους πόρους και στο χρόνο εκτέλεσης, κατά 18% και 38% αντίστοιχα. Ωστόσο, καθώς προστίθενται περισσότεροι επιταχυντές Map, η απόδοση αναφορικά προς τη ρυθμικό απόδοση αυξάνεται σχεδόν γραμμικά. Η αύξηση αυτή φτάνει σε κάποιο σημείο κορεσμού, για κάποιες εφαρμογές, καθώς η χρήση περισσότερων επιταχυντών δεν οδηγεί σε αντίστοιχη επιτάχυνση εκτέλεσης. Αυτή είναι η περίπτωση για τις εφαρμογές String Match, PCA, Kmeans και Word Count. Ύστερα από σχετική ανάλυση, διαπιστώθηκε ότι η δυναμική συμπεριφορά αυτών των εφαρμογών, καθώς και τα δεδομένα εξάρτησης μεταξύ των υπολογισμών αυτών των εφαρμογών αποτρέπει το εργαλείο Vivado HLS σε περαιτέρω αποτελεσματικές βελτιστοποιήσεις επεξεργασίας δεδομένων. Για παράδειγμα, η εφαρμογή PCA περιέχει έναν αλγόριθμο δίχως δυναμική συμπεριφορά. Ωστόσο, η εκτέλεσή της εμφανίζει υψηλή εξάρτηση δεδομένων, χωρίς μεγάλη τοπικότητα των αναφορών. Συνεπώς δεν είναι αποδοτική η λεπτομερής διαίρεση των δεδομένων εισόδου σε μικρά υποσύνολα δεδομένων, ώστε αυτά να προσπελαύνονται μόνο από έναν επιταχυντή Map.



Σχήμα 5.19 Ανάλυση επιδόσεων κλιμάκωσης με τη χρήση του πλαισίου HLSMapReduceFlow.

Πίνακας 5.3 Σύγκριση του πλαισίου HLSMapReduceFlow-FPGA και επεξεργαστή γενικού σκοπού, σε αντιπροσωπευτικές εφαρμογές σταθμού εργασίας.

Πλαίσιο	GNU/Linux 3.18.6 x86-64 / GCC-glibc			HLSMapReduceFlow				Αναλογία		
	AMD 8-core FX-8350 4GHz			Virtex7-XC7VX485T 150MHz						
Μετρική	Χρόνος(ms)	Ισχύς(W)	Ενέργεια(J)	T_p (ms)	T_c (ms)	Ισχύς(W)	Ενέργεια(J)	T	P	E
Histogram	344	41.1	14.1	72.2	4.8	1.84	0.13	0.21	0.04	0.009
Matrix Mul/tion	177	41.3	7.3	208	0.6	1.02	0.21	1.17	0.03	0.029
String Match	206	41.6	8.5	95	4.9	2.33	0.22	0.46	0.06	0.026
Word Count	172	40.8	7.0	84	1.4	1.87	0.16	0.48	0.05	0.023
Linear Reg/sion.	158	41.6	6.6	73	6.4	2.08	0.15	0.46	0.05	0.023
PCA	392	41.9	16.4	964	4.1	1.17	1.13	2.45	0.03	0.070
K_{means}	435	40.3	17.5	503	3.8	1.03	0.52	1.16	0.03	0.029
Μέσος όρος	269	41.2	11	285	3.7	1.62	0.36	1.06×	0.04×	0.03×

Εν συνεχεία, αναφέρονται τα αποτελέσματα από μία αντιπροσωπευτική μελέτη σχετικά με την καταλληλότητα χρήσης συσκευών FPGAs σε σταθμούς εργασίας. Σε αυτή την εργασία αξιολογείται η προαναφερθείσα πλατφόρμα του πλαισίου HLSMapReduceFlow σε σύγκριση με ένα σύγχρονο σταθμό εργασίας. Ο σταθμός εργασίας αποτελείται από τον οκταπύρρηνο επεξεργαστή AMD FX-8350 χρονισμένο στη συχνότητα 4GHz. Αυτός ο επεξεργαστής έχει μέγιστη τιμή ανοχής ενέργειας (TDP) 125 Watt. Οι χρησιμοποιούμενες εφαρμογές συντέθηκαν με το μεταγλωττιστή GCC (v4.9.2) και συνδέθηκαν κατά το χρόνο εκτέλεσης με το πλαίσιο glibc, σε λειτουργικό σύστημα GNU/Linux (Kernel 3.18.6) 64-bit, επιτρέποντας πολλές βελτιστοποιήσεις του μεταγλωττιστή (-O2), συμπεριλαμβανομένων των βελτιστοποιήσεων επεξεργασίας διανυσμάτων (SSE, AVX κλπ). Ο Πίνακας 5.3 αναφέρει τα συνολικά αποτελέσματα. Οι τρεις πρώτες στήλες του πίνακα αυτού αναφέρονται στο χρόνο εκτέλεσης, στην ισχύ και στην ενέργεια κατανάλωσης. Οι επόμενες τέσσερις στήλες εμπεριέχουν τις αντίστοιχες μετρήσεις για ένα σύστημα αποτελούμενο από μία πλατφόρμα Virtex-7 FPGA (XC7VX485T), χρονισμένη στη συχνότητα 150MHz, η οποία χρησιμοποιεί το πλαίσιο HLSMapReduceFlow. Η επικοινωνία με έναν υπολογιστή υποδοχής, ο οποίος μεταφέρει δεδομένα στο FPGA, πραγματοποιείται μέσω ενός διαύλου PCI Express 3.0, συνολικού ρυθμού μεταφοράς δεδομένων 8Gbps. Ο συνολικός χρόνος εκτέλεσης των εφαρμογών στο FPGA συνυπολογίζει το χρόνο επεξεργασίας T_p καθώς και το χρόνο μεταφοράς δεδομένων T_c . Η τελευταία μετρική συμπεριλαμβάνει το χρόνο μεταφορά δεδομένων εισόδου και το χρόνο αποστολής δεδομένων εξόδου στον υπολογιστή υποδοχής.

Στη παρούσα εργασία δεν έχει χρησιμοποιηθεί κάποια τεχνική επικάλυψης του χρόνου εκτέλεσης με το χρόνο επικοινωνίας, παρόλο που μία τέτοια προσέγγιση αναμένεται να αυξήσει τα κέρδη επιτάχυνσης (π.χ. με χρήση τεχνικών σωλήνωσης). Οι μετρικές ισχύος και κατανάλωσης ενέργειας υπολογίσθηκαν με τη χρήση των εργαλείων PowerTop⁷ και Xilinx Xpower⁸ για τον επεξεργαστή του σταθμού εργασίας και τη συσκευή FPGA, αντίστοιχα. Όπως φαίνεται, το προτεινόμενο πλαίσιο παρέχει αποδοτικές λύσεις, μεγάλης απόδοσης-ανά-watt, σημειώνοντας έως δύο τάξεις μεγέθους μικρότερη κατανάλωση ενέργειας για τον ίδιο χρόνο εκτέλεσης. Κατά συνέπεια, αποδεικνύεται ότι η προτεινόμενη προσέγγιση αποτελεί μία ευέλικτη λύση για τη κλιμάκωση των σύγχρονων κέντρων δεδομένων και σταθμών εργασίας.

5.3. Συμπεράσματα

Οι δύο εργασίες του παρόντος κεφαλαίου αποτελούν προτεινόμενες λύσεις αρχιτεκτονικής υπολογιστικής κλιμάκωσης. Και οι δύο εργασίες προτείνουν ένα νέο υπολογιστικό πρότυπο και το αντίστοιχο περιβάλλον προγραμματισμού για την υλοποίηση συστημάτων αυξημένης υπολογιστικής ισχύος, χρησιμοποιώντας, για αυτό το σκοπό, επαναδιαμορφούμενες πλατφόρμες FPGA. Κάθε εργασία έχει διαφορετικούς στόχους υλοποίησης και κριτήρια βελτιστοποίησης, ενώ προσπαθεί να λύσει διαφορετικά σχεδιαστικά προβλήματα. Η πρώτη, η οποία παρουσιάζεται στην Ενότητα 5.1, στοχεύει στην ανάπτυξη συστημάτων πολλαπλών επιταχυντών υλικού και επιτυγχάνει αύξηση ρυθμαπόδοσης έως 21.4x. Η δεύτερη, η οποία παρουσιάζεται στην Ενότητα 5.2 επικεντρώνεται στην ανάπτυξη ενός αρχιτεκτονικού προτύπου επιτάχυνσης του πλαισίου MapReduce, στοχεύοντας την αγορά των σταθμών εργασίας και επιτυγχάνοντας ελάττωση της κατανάλωσης ενέργειας κατά δύο τάξεις μεγέθους. Οι δύο εργασίες χρησιμοποιούν τεχνικές προγραμματισμού και βελτιστοποίησης με εργαλεία High-Level-Synthesis προκειμένου να επιτρέψουν ταχεία, συστημική και αφαιρετική σχεδίαση. Επίσης οι εργασίες είναι ορθογώνια εφαρμόσιμες για συστήματα κλιμακούμενης υπολογιστικής ισχύος και ο συνδυασμός τους αποτελεί μέλλουσα ερευνητική εργασία αυτής της διατριβής.

⁷<https://01.org/powertop>

⁸http://www.xilinx.com/products/design_tools/logic_design/verification/xpower.htm

Βιβλιογραφία

- [1] G. Moore, *Cramming more components onto integrated circuits*, *Proceedings of the IEEE* **86**, 82 (1998).
- [2] CNET.com, *End of moore's law: It's not just about physics*, .
- [3] Recode.net, *Moore's law hits 50, but it may not see 60*, .
- [4] R. Dennard, F. Gaensslen, H.-N. YU, V. Rideout, E. BASSOUS, and A. R. LEBLANC, *Design of ion-implanted mosfet's with very small physical dimensions*, *Proceedings of the IEEE* **87**, 668 (1999).
- [5] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, *Power challenges may end the multicore era*, *Commun. ACM* **56**, 93 (2013).
- [6] M. Shafique, S. Garg, J. Henkel, and D. Marculescu, *The eda challenges in the dark silicon era: Temperature, reliability, and variability perspectives*, in *Proceedings of the 51st Annual Design Automation Conference*, DAC '14 (ACM, New York, NY, USA, 2014) pp. 185:1–185:6.
- [7] C. Mack, *Fifty years of moore's law*, *Semiconductor Manufacturing, IEEE Transactions on* **24**, 202 (2011).
- [8] M. Taylor, *Is dark silicon useful? harnessing the four horsemen of the coming dark silicon apocalypse*, in *Design Automation Conference (DAC), 2012 49th ACM/EDAC/IEEE* (2012) pp. 1131–1136.
- [9] J. Shalf, D. Quinlan, and C. Janssen, *Rethinking hardware-software codesign for exascale systems*, *Computer* **44**, 22 (2011).
- [10] Q. Zhu, B. Akin, H. Sumbul, F. Sadi, J. Hoe, L. Pileggi, and F. Franchetti, *A 3d-stacked logic-in-memory accelerator for application-specific data intensive computing*, in *3D Systems Integration Conference (3DIC), 2013 IEEE International* (2013) pp. 1–7.
- [11] G. Venkatesh, J. Sampson, N. Goulding, S. Garcia, V. Bryksin, J. Lugo-Martinez, S. Swanson, and M. B. Taylor, *Conservation cores: Reducing the energy of mature computations*, *SIGARCH Comput. Archit. News* **38**, 205 (2010).
- [12] Y.-T. Chen, J. Cong, M. Ghodrati, M. Huang, C. Liu, B. Xiao, and Y. Zou, *Accelerator-rich cmps: From concept to real hardware*, in *Computer Design (ICCD), 2013 IEEE 31st International Conference on* (2013) pp. 169–176.
- [13] J. Cong, M. A. Ghodrati, M. Gill, B. Grigorian, and G. Reinman, *Architecture support for domain-specific accelerator-rich cmps*, *ACM Trans. Embed. Comput. Syst.* **13**, 131:1 (2014).

- [14] A. Putnam, A. Caulfield, E. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. P. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J.-Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Y. Xiao, and D. Burger, *A reconfigurable fabric for accelerating large-scale datacenter services*, in *41st Annual International Symposium on Computer Architecture (ISCA)* (2014).
- [15] J. Cong, B. Liu, S. Neuendorffer, J. Noguera, K. Vissers, and Z. Zhang, *High-level synthesis for fpgas: From prototyping to deployment*, *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* **30**, 473 (2011).
- [16] M. Araya-Polo, J. Cabezas, M. Hanzich, M. Pericas, F. Rubio, I. Gelado, M. Shafiq, E. Morancho, N. Navarro, E. Ayguade, J. Cela, and M. Valero, *Assessing accelerator-based hpc reverse time migration*, *Parallel and Distributed Systems, IEEE Transactions on* **22**, 147 (2011).
- [17] M. J. Lyons, M. Hempstead, G.-Y. Wei, and D. Brooks, *The accelerator store: A shared memory framework for accelerator-based systems*, *ACM Trans. Archit. Code Optim.* **8**, 48:1 (2012).
- [18] E. Cota, P. Mantovani, M. Petracca, M. Casu, and L. Carloni, *Accelerator memory reuse in the dark silicon era*, *IEEE Computer Architecture Letters* **99**, 1 (2012).
- [19] L. Semeria and G. De Micheli, *Spc: synthesis of pointers in c application of pointer analysis to the behavioral synthesis from c*, in *Computer-Aided Design, 1998. ICCAD 98. Digest of Technical Papers. 1998 IEEE/ACM International Conference on* (1998) pp. 340–346.
- [20] L. Séméria, K. Sato, and G. De Micheli, *Resolution of dynamic memory allocation and pointers for the behavioral synthesis form c*, in *Proceedings of the Conference on Design, Automation and Test in Europe*, DATE '00 (ACM, New York, NY, USA, 2000) pp. 312–319.
- [21] M. Shalan and V. J. Mooney, *A dynamic memory management unit for embedded real-time system-on-a-chip*, in *Proceedings of the 2000 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*, CASES '00 (ACM, New York, NY, USA, 2000) pp. 180–186.
- [22] M. J. Lyons, M. Hempstead, G.-Y. Wei, and D. Brooks, *The accelerator store: A shared memory framework for accelerator-based systems*, *ACM Trans. Archit. Code Optim.* **8**, 48:1 (2012).
- [23] S. Xydis, A. Bartzas, I. Anagnostopoulos, D. Soudris, and K. Z. Pekmetzi, *Custom multi-threaded dynamic memory management for multiprocessor system-on-chip platforms*, in *ICSAMOS* (2010) pp. 102–109.

- [24] Y. Sade, M. Sagiv, and R. Shaham, *Optimizing c multithreaded memory management using thread-local storage*, in *Proceedings of the 14th International Conference on Compiler Construction*, CC'05 (2005) pp. 137–155.
- [25] C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, and C. Kozyrakis, *Evaluating mapreduce for multi-core and multiprocessor systems*, in *High Performance Computer Architecture, 2007. HPCA 2007. IEEE 13th International Symposium on* (2007) pp. 13–24.
- [26] ITRS, *International technology roadmap for semiconductos*, (2012).
- [27] Y. Shan, B. Wang, J. Yan, Y. Wang, N. Xu, and H. Yang, *Fpmr: Mapreduce framework on fpga*, in *Proceedings of the 18th Annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, FPGA '10 (ACM, New York, NY, USA, 2010).
- [28] D. Yin, G. Li, and K.-d. Huang, *Scalable mapreduce framework on fpga accelerated commodity hardware*, in *Internet of Things, Smart Spaces, and Next Generation Networking*, Lecture Notes in Computer Science, Vol. 7469, edited by S. Andreev, S. Balandin, and Y. Koucheryavy (Springer Berlin Heidelberg, 2012).
- [29] B. He, W. Fang, Q. Luo, N. K. Govindaraju, and T. Wang, *Mars: A mapreduce framework on graphics processors*, in *Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques*, PACT '08 (ACM, New York, NY, USA, 2008) pp. 260–269.
- [30] J. H. C. Yeung, C. C. Tsang, K. H. Tsoi, B. S. H. Kwan, C. C. C. Cheung, A. P. C. Chan, and P. H. W. Leong, *Map-reduce as a programming model for custom computing machines*, in *Proceedings of the 2008 16th International Symposium on Field-Programmable Custom Computing Machines*, FCCM '08 (IEEE Computer Society, Washington, DC, USA, 2008) pp. 149–159.
- [31] K. Gyftakis, I. Anagnostopoulos, D. Soudris, and D. Reisis, *A mapreduce framework implementation for network-on-chip platforms*, in *Electronics, Circuits and Systems (ICECS), 2014 21st IEEE International Conference on* (2014) pp. 120–123.
- [32] K.-H. Lee, Y.-J. Lee, H. Choi, Y. D. Chung, and B. Moon, *Parallel data processing with mapreduce: A survey*, *SIGMOD Rec.* **40**, *11* (2012).
- [33] A. Putnam, A. Caulfield, E. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmailzadeh, J. Fowers, G. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J.-Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Xiao, and D. Burger, *A reconfigurable fabric for accelerating large-scale datacenter services*, in *Computer Architecture (ISCA), 2014 ACM/IEEE 41st International Symposium on* (2014) pp. 13–24.

- [34] C. Kachris, G. Sirakoulis, and D. Soudris, *A reconfigurable mapreduce accelerator for multi-core all-programmable socs*, in *System-on-Chip (SoC), 2014 International Symposium on* (2014) pp. 1–6.

6

Συμπεράσματα και Μελλοντικές Προεκτάσεις

Στο παρόν κεφάλαιο αποτυπώνονται τα βασικά συμπεράσματα που προέκυψαν από τη διδακτορική διατριβή. Συγκεκριμένα, σημειώνονται οι κύριοι άξονες καινοτομίας που εισήγαγαν τα προτεινόμενα μεθοδολογικά πλαίσια και εργαλεία σχεδίασης ενώ παράλληλα αναφέρονται οι μελλοντικές ερευνητικές προεκτάσεις της παρούσας έρευνας.

6.1. Επισκόπηση διδακτορικής διατριβής

Η διδακτορική διατριβή “Πολυεπίπεδη Ταχεία Προτυποποίηση και Σύνθεση Εξειδικευμένων και Επαναδιαμορφούμενων Συστημάτων Πολλαπλών Επιταχυντών Υλικού” εισηγήθηκε νέα αρχιτεκτονικά πρότυπα και καινοτόμες σχεδιαστικές προσεγγίσεις προκειμένου να δώσει λύσεις στα σύγχρονα σχεδιαστικά προβλήματα των ολοκληρωμένων συστημάτων-σε-ψηφίδα. Η στόχευση της διατριβής είναι διττή, τεχνολογική και εμπορική. Η διατριβή προσπαθεί να συνεισφέρει ανταγωνιστικές λύσεις, οι οποίες αφενός υπόσχονται να υπακούσουν στις τάσεις της αγοράς και αφετέρου να είναι βιώσιμες, αναφορικά προς τη κλιμάκωση της τεχνολογίας σε επίπεδο υλικού ημιαγωγών, εργαλείων σχεδιασμού και χρόνου ανάπτυξης. Αυτή η προσέγγιση επιτυγχάνεται με την ανάπτυξη ενός ολιστικού μεθοδολογικού πλαισίου, το οποίο κατηγοριοποιεί τα επίπεδα σχεδίασης και προτείνει διακριτές και μη επικαλυπτόμενες διαδικασίες βελτιστοποίησης. Παράλληλα προτείνονται αρχιτεκτονικά πρότυπα τα οποία συνάδουν με τις εμπορικές κατευθύνσεις, προσφέροντας αυξημένη συστημική πολυπλοκότητα, μικρό χρόνο σχεδίασης και επαλήθευσης ορθής λειτουργίας, καθώς και ένα σύνολο μετα-χαρακτηριστικών λειτουργίας, όπως η αξιοπιστία, ο επαναπρογραμματισμός και η προσαρμοστικότητα, τα οποία αποτελούν ορόσημο για τη μετάβαση στην νέα γενιά ολοκληρωμένων συστημάτων ημιαγωγών. Βασική

κατεύθυνση αυτής της γενιάς είναι η γεφύρωση του χάσματος της σχεδιαστικής πολυπλοκότητας και της τεχνολογικής παραγωγικότητας, η οποία και επιτυγχάνεται με την διαγώνια πορεία σχεδιασμού, ανάμεσα στους δύο κάθετους άξονες ανάπτυξης, του τεχνολογικού και του εμπορικού. Αναφορικά προς τη σειρά έκθεσης της παρούσας έρευνας στα κεφάλαια αυτής της διατριβής, οι επί μέρους συνεισφορές απαριθμίζονται ως εξής:

1. Αναπτύχθηκε ένα νέο μεθοδολογικό πλαίσιο ταχείας εικονικής προτυποποίησης για ενσωματωμένα συστήματα καθώς και συστήματα που συνδυάζουν επεξεργαστές γενικού σκοπού με επιταχυντές υλικού (Κεφάλαιο 2). Το πλαίσιο αυτό προτείνει την υιοθέτηση εικονικών μοντέλων των επεξεργαστικών στοιχείων υλικού, κατά τη φάση σχεδίασης, προκειμένου να διασφαλίσει ένα λειτουργικό επίπεδο της υπό-ανάπτυξης πλατφόρμας για την ανάπτυξη του λογισμικού. Αυτή η νέα προσέγγιση συν-σχεδιασμού καθιστά δυνατή την εκκίνηση της ανάπτυξης λογισμικού, καθώς και τη δοκιμή και την αποσφαλμάτωση αυτού, πολύ νωρίτερα από ό,τι είναι δυνατό με συμβατικές ακολουθιακές μεθοδολογίες συν-σχεδιασμού. Η προτεινόμενη μέθοδος ανάπτυξης επαληθεύεται με σενάρια ελέγχου που εκτείνονται από το σχεδιασμό εξειδικευμένων πυρήνων υλικού, έως το συν-σχεδιασμό υλικού/λογισμικού ετερογενών ενσωματωμένων συστημάτων [1, 2].
2. Αναπτύχθηκε ένα μεθοδολογικό πλαίσιο για τον αποτελεσματικό συν-σχεδιασμό υλικού/λογισμικού και την κατάτμηση των επεξεργαστικών διεργασιών στους διαθέσιμους υπολογιστικούς πόρους ενός ετερογενούς συστήματος, αποτελούμενο από επεξεργαστές γενικού σκοπού και συσκευές επαναδιαμορφώμενης αρχιτεκτονικής, βάση μιας διαδικασίας χαρακτηρισμού των εφαρμογών (Ενότητα 3.4). Η καινοτομία που εισάγει το συγκεκριμένο πλαίσιο αφορά την πολυ-κριτηριακή και συστηματική ανάλυση των εφαρμογών προκειμένου να επιλεγεί η σωστή απόφαση για τοποθέτηση της κάθε επεξεργαστικής διεργασίας στον βέλτιστο πόρο επεξεργασίας, π.χ. λογισμικό-CPU / υλικό-FPGA. Η προτεινόμενη συστηματική μεθοδολογία εξετάζει την υπολογιστική πολυπλοκότητα, τις απαιτήσεις μνήμης (μέγεθος αποθήκευσης και διάρκεια ζωής δεδομένων) καθώς και το κόστος επικοινωνίας, ώστε να βρεθεί η βέλτιστη αρχιτεκτονική αδρομερών επιταχυντών υλικού ως μέσο επιτάχυνσης της συνολικής απόδοσης του συστήματος. Η προσέγγιση αυτή χρησιμοποιήθηκε στη σχεδίαση ενός ετερογενούς συστήματος, το οποίο υποστηρίζει αυτόνομη πλοήγηση σε διαστημικά ρομποτικά οχήματα μέσω χρήσης αλγορίθμων μηχανικής όρασης. Το υλοποιημένο σύστημα κατάφερε να επιτύχει τις προδιαγραφές που τέθηκαν από τον οργανισμό ESA, μέσω της σχεδίασης εξειδικευμένων επιταχυντών υλικού, σύμφωνα από τις υποδείξεις συν-σχεδίασης και κατάτμησης του αναφερόμενου πλαισίου [3-5].
3. Αναπτύχθηκε ένα μεθοδολογικό πλαίσιο καθώς και τα αντίστοιχα εργαλεία σχεδίασης για την εκτίμηση του κέρδους σχεδιασμού από την υιοθέτηση της τρισδιάστατης τεχνολογίας ολοκλήρωσης ημιαγωγών (3-D) με

τη χρήση υφιστάμενων εμπορικών συμβατικών εργαλείων σχεδίασης CAD (Ενότητα 3.5). Η προτεινόμενη προσέγγιση εισάγει νέα εργαλεία σχεδίασης τα οποία καταφέρνουν να προσθέσουν τους περιορισμούς που θέτει η τρίτη διάσταση ολοκλήρωσης στα συμβατικά εργαλεία 2-D. Το σύνολο των αναπτυχθέντων εργαλείων χρησιμοποιεί ανοιχτούς τύπους αρχείων για τους σχεδιασμούς προκειμένου να αποδώσει ανταγωνιστική υπεροχή έναντι της περιορισμένης και βιομηχανικά ιδιόκτητης διαθεσιμότητας των σύγχρονων εργαλείων σχεδίασης 3-D. Το πλαίσιο χρησιμοποιήθηκε κατά το σχεδιασμό δύο συστημάτων 3-D SoC για επεξεργασία σήματος βασισμένα στους ενσωματωμένους επεξεργαστές *Leon 3* και *OpenRISC*, σε τεχνολογίες ολοκλήρωσης TSMC 45nm και 130nm, πετυχαίνοντας εξοικονόμηση ενέργειας κατά ποσοστό 20% συγκρινόμενη με την συμβατική δυσδιάστατη υλοποίηση καθώς και βελτιστοποίηση της μέγιστης συχνότητας λειτουργίας έως και 1,26x, αντίστοιχα [1, 6–8].

4. Αναπτύχθηκε μία αποτελεσματική και ταχεία λύση στο πρόβλημα τοποθέτησης για επαναδιαμορφούμενες αρχιτεκτονικές FPGA (Κεφάλαιο 4). Ο προτεινόμενος αλγόριθμος τοποθέτησης χρησιμοποιεί τους γενετικούς αλγόριθμους, προκειμένου να εξερευνήσει αποτελεσματικότερα το χώρο λύσεων. Παράλληλα το προτεινόμενο εργαλείο τοποθέτησης, εκμεταλλεύομενο τον εγγενή παραλληλισμό των γενετικών αλγορίθμων, επιτυγχάνει βέλτιστη κλιμάκωση στους σύγχρονους επεξεργαστές πολλαπλών πυρήνων, μέσω εξειδικευμένου αδρομερούς προγραμματιστικού παραλληλισμού. Για τη βελτίωση της ποιότητας των λύσεων τοποθέτησης, το προτεινόμενο μεθοδολογικό πλαίσιο χρησιμοποιεί ένα στάδιο διερεύνησης των παραμέτρων του γενετικού αλγορίθμου, προσαρμοσμένο στην εκάστοτε εφαρμογή. Τα πειραματικά αποτελέσματα επαλήθευσαν την υπεροχή του προτεινόμενου εργαλείου σε σχέση με τα αποδοτικότερα υφιστάμενα εργαλεία, προσφέροντας λύσεις οι οποίες, είτε διερευνώνται ταχύτερα κατά ένα παράγοντα έως 67x, είτε είναι ποιοτικότερες αναφορικά με την επιτευχθείσα συχνότητα λειτουργίας κατά 1,16x, ενώ ταυτόχρονα διατηρούν σταθερή κλιμάκωση όσον αφορά τη δέσμευση μνήμη και τις υπολογιστικές απαιτήσεις πόρων.
5. Αναπτύχθηκε ένα μεθοδολογικό πλαίσιο το οποίο προτείνει την αυτοματοποιημένη εξερεύνηση της μικρο-αρχιτεκτονικής ενός συστήματος-σε-ψηφίδα, με σκοπό την ελαχιστοποίηση της έκκληση θερμότητας και κατ'επέκταση την αύξηση της αξιοπιστίας του ολοκληρωμένου κυκλώματος (Ενότητα 3.2). Το πλαίσιο, μέσα από σχετικές μετρήσεις, αναγνωρίζει ότι η πυκνότητα θερμότητας σε ένα ολοκληρωμένο σύστημα πυριτίου είναι αφενός ανομοιογενής και αφετέρου σημειώνει μεγάλες τιμές σε εκείνα τα τμήματα υλικού τα οποία έχουν αυξημένη λειτουργία και μικρή επιφάνεια πυριτίου. Αυτό το πρόβλημα οδηγεί σε περιορισμένο βαθμό έκκλησης θερμότητας και συνεπώς σε συνεχή καταπόνηση, λόγω θερμοκρασίας, της συγκεκριμένης επιφάνειας πυριτίου. Το πλαίσιο προτείνει την επιλεκτική αντιγραφή των μονάδων υλικού που αντιμετωπίζουν το προαναφερθέν πρόβλημα, ενώ μέσω

της σχεδίασης ενός εξειδικευμένου ελεγκτή πραγματικού χρόνου, το υλοποιημένο σύστημα είναι ικανό να διατηρεί σταθερή θερμική συμπεριφορά ανά εφαρμογή, χρονοπρογραμματίζοντας τις διεργασίες των εφαρμογών στα κατανομημένα αντίγραφα υλικού, βάση της τρέχουσας θερμοκρασίας τους. Τα πειραματικά αποτελέσματα από την εφαρμογή της προτεινόμενης προσέγγισης σε ένα σύστημα ασύρματης ραδιοζεύξης (SDR), έδειξαν ότι το πλαίσιο μπορεί να επιτύχει τη μείωση της μέσης θερμοκρασίας κατά 17 βαθμούς Kelvin καθώς και την μείωση της γήρανσης του ολοκληρωμένου συστήματος κατά 14%, με μία αύξηση της επιφάνειας πυριτίου, μόλις κατά 15% [9, 10].

6. Αναπτύχθηκε μία σχεδιαστική ροή αρχιτεκτονικών τύπου Δικτύου-σε-Ψηφίδα (NoC) με χρήση της τρισδιάστατης τεχνολογίας ολοκλήρωσης (Ενότητα 3.3). Καθώς η χρήση των TSVs επηρεάζει τόσο το κόστος κατασκευής ενός ολοκληρωμένου κυκλώματος όσο και τα φυσικά χαρακτηριστικά σχεδίασης (συχνότητα λειτουργίας, καθυστέρηση, επιφάνεια πυριτίου, εκλυόμενη θερμοκρασία κτλ.), σκοπός της εργασίας είναι ο σχεδιασμός κατάλληλων παραμετρικών δρομολογητών 2-D και 3-D οι οποίοι επιτρέπουν τη βελτιστοποίηση της φυσικής σχεδίασης παρέχοντας ένα στάδιο μελέτης συμβιβασμών μεταξύ των μετρικών ποιότητας σχεδίασης. Τα αποτελέσματα από την υλοποίηση αυτής της αρχιτεκτονικής σε ευρέως διαδεδομένους αλγόριθμους επεξεργασίας εικόνας και βίντεο, έδειξαν ότι είναι εφικτή η επίτευξη μεγαλύτερης συχνότητας λειτουργίας κατά 25% με ταυτόχρονη μείωση της καταναλισκόμενης ενέργειας κατά 39%, σε σύγκριση με συμβατικές μεθόδους σχεδιασμού [11].
7. Αναπτύχθηκε ένα νέο αρχιτεκτονικό πρότυπο για την υποστήριξη συστημάτων πολλαπλών επιταχυντών υλικού, αδρομερούς υψής, σε επαναδιαμορφούμενες πλατφόρμες (Ενότητα 5.4). Διαπιστώνεται, μέσω μετρήσεων, ότι το κύριο πρόβλημα για την ανάπτυξη τέτοιων επιταχυντών είναι αφενός ο αποδοτικός προγραμματισμός τους, και η χρονο-διαχείριση των πολλαπλών διεργασιών και αφετέρου η περιορισμένη διαθέσιμη μνήμη που βρίσκεται στις επαναδιαμορφούμενες πλατφόρμες. Ειδικά μέσα από την εργασία αυτή, τονίζεται ότι παρόλο που η ερευνητική κοινότητα αναφέρει την παροχή επαρκούς ενέργειας ως το μεγαλύτερο πρόβλημα στη σχεδίαση συστημάτων μαζικών επιταχυντών υλικού (Dark Silicon) [12, 13], εντούτοις, το πρόβλημα της περιορισμένης on-chip μνήμης, ακόμα και στα σύγχρονα FPGAs, μπορεί να οδηγήσει σε συστήματα με 2.5 φορές λιγότερους επιταχυντές, σε σχέση με το μέγιστο ανεκτό θερμικό και ενεργειακό κατώφλι των σύγχρονων ολοκληρωμένων συστημάτων. Σημαντικό ρόλο σε αυτό το πρόβλημα αποτελεί η στατική δέσμευση μνήμης, που είναι μονόδρομος για τα εργαλεία σχεδίασης FPGAs. Προς αυτή τη κατεύθυνση, το προτεινόμενο πλαίσιο προτείνει την υιοθέτηση εργαλείων υψηλού επιπέδου σύνθεσης, για τον αφαιρετικό προγραμματισμό και την ευέλικτη χρονο-διαχείριση των πολλαπλών διεργασιών για συστήματα πολλαπλών επιταχυντών, ενώ παράλληλα προσφέρει βιβλιοθήκες δυναμικής διαχείρισης μνήμης, προκει-

μένου να ξεπεραστεί το εμπόδιο της στατικής δέσμευσης μνήμης καθ' όλη τη διάρκεια εκτέλεσης μίας διεργασίας. Το προτεινόμενο πλαίσιο εφαρμοσθήκε σε σύγχρονες εφαρμογές του χώρου των Υπολογιστών Υψηλής Επίδοσης (High Performance Computing, HPC) στο ωριμότερο, βιομηχανικά, εργαλείο HLS για FPGAs, το Xilinx Vivado HLS και κατάφερε να αυξήσει τους επιταχυντές κατά 3.8× και κατ' επέκταση τη ρυθμαπόδοση του συστήματος κατά 3.1× και 21.4× σε αρχιτεκτονικές με κοινή και ιδιωτική μνήμη ανάμεσα στους επιταχυντές, αντίστοιχα [14, 15].

8. Αναπτύχθηκε ένα αρχιτεκτονικό πρότυπο και το αντίστοιχο προγραμματιστικό περιβάλλον για κλιμακούμενες σχεδιαστικές λύσεις σε κέντρα δεδομένων και σταθμούς εργασίας (Ενότητα 5.2). Το πλαίσιο στηρίζεται στη χρήση της τεχνικής προγραμματισμού και εκτέλεσης εργασιών MapReduce και στοχεύει στη σχεδίαση επεξεργαστικών κόμβων σε καταναμημένα συστήματα για την επεξεργασία μεγάλου όγκου δεδομένων. Συγκεκριμένα, αναπτύσσεται μία ροή εργασίας για συστήματα πολλαπλών επιταχυντών που αξιοποιούν το πλαίσιο MapReduce σε FPGA. Η εφαρμογή της εργασίας σε ένα τυπικό σταθμό εργασίας έδειξε ότι η προτεινόμενη αρχιτεκτονική είναι ενεργειακά αποδοτικότερη κατά δύο τάξεις μεγέθους, σε σχέση με τους σύγχρονους πολυπύρηνους επεξεργαστές διακομιστών γενικού σκοπού [16].

6.2. Μελλοντικές Προεκτάσεις

Τα κυριότερα θέματα για μελλοντική έρευνα που απορρέουν από τη παρούσα διατριβή, σε συνάφεια με τις ερευνητικές κατευθύνσεις που μελετά η επιστημονική ομάδα του εργαστηρίου Μικροϋπολογιστών και Ψηφιακών Συστημάτων του Ε.Μ.Π., συνοψίζονται ως εξής:

- Επέκταση πλαισίου εικονικής προτυποποίησης σε επίπεδο λειτουργικού συστήματος προκειμένου να καταστεί δυνατή η πρόωγη επαλήθευση λογισμικού εφαρμογών που εξαρτώνται από βιβλιοθήκες λειτουργικού συστήματος.
- Συνδυασμένη βελτιστοποίηση μετρικών θερμοκρασίας, κατανάλωσης ενέργειας και αξιοπιστίας στον γενετικό αλγόριθμο τοποθέτησης για συσκευές FPGA, προκειμένου να διερευνάνται ο χώρος λύσεων σε κάθε βήμα εκτέλεσης.
- Υποστήριξη μερικού-επαναπρογραμματισμού στο προτεινόμενο εργαλείο τοποθέτησης προκειμένου να αυξηθεί η ευελιξία αυτό-προσαρμοστικότητας των συσκευών FPGA στις δυναμικές συνθήκες λειτουργίας.
- Μελέτη πολλαπλών καναλιών τάσης τροφοδοσίας στη τρισδιάστατη τεχνολογία ολοκλήρωσης προκειμένου να υποστηριχθούν αρχιτεκτονικές χαμηλής ενεργειακής και θερμικής απόδοσης ώστε να εξισορροπείται η χαμηλή απαγωγή θερμότητας που χαρακτηρίζει τα κυκλώματα 3-D.

- Συνδυασμένη εξερεύνηση των παραμέτρων της δυναμικής διαχείρισης μνήμης σε επίπεδο υλικού, προκειμένου να υποστηριχθεί βέλτιστη προσαρμογή του υποσυστήματος μνήμης στις απαιτήσεις δεδομένων των δυναμικών εφαρμογών για συστήματα πολλαπλών επιταχυντών υλικού.
- Ανάπτυξη μεθοδολογίας αυτοματοποιημένης διερεύνησης παραμέτρων υψηλού επιπέδου σύνθεσης για αρχιτεκτονικές πολλαπλών επιταχυντών υλικού, ώστε να συσχετισθεί ο χώρος λύσεων με τις παραμέτρους βελτιστοποίησης HLS.
- Ενσωμάτωση δυναμικού διαχειριστή μνήμης σε περιβάλλον MapReduce για εφαρμογές σταθμών εργασίας και κέντρων δεδομένων, ώστε να υποστηριχθούν ευέλικτες τεχνικές μεταφοράς δεδομένων με προσαρμοσμένο αποτύπωμα μνήμης στις πραγματικές ανάγκες της δυναμικής εφαρμογής και συνεπώς, αύξηση της επεξεργαστικής ισχύος ανά μονάδα κατανάλωσης ενέργειας.

Βιβλιογραφία

- [1] D. Diamantopoulos, E. Sotiriou-Xanthopoulos, K. Siozios, G. Economakos, and D. Soudris, *Plug&chip: A framework for supporting rapid prototyping of 3d hybrid virtual socs*, *ACM Trans. Embed. Comput. Syst.* **13**, 168:1 (2014).
- [2] D. Diamantopoulos, K. Siozios, E. Sotiriou-Xanthopoulos, G. Economakos, and D. Soudris, *Hvsocs: A framework for rapid prototyping of 3-d hybrid virtual system-on-chips*, in *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2013 IEEE 27th International* (2013) pp. 2194–2199.
- [3] I. Kostavelis, L. Nalpantidis, E. Boukas, M. A. Rodrigalvarez, I. Stamoulias, G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, and A. Gasteratos, *Spartan: Developing a vision system for future autonomous space exploration robots*, *Journal of Field Robotics* **31**, 107 (2014).
- [4] D. Diamantopoulos, K. Siozios, G. Lentaris, D. Soudris, and M. Rodrigalvarez, *Spartan project: On profiling computer vision algorithms for rover navigation*, in *Adaptive Hardware and Systems (AHS), 2012 NASA/ESA Conference on* (2012) pp. 174–181.
- [5] G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, and M. Rodrigalvarez, *Hardware implementation of stereo correspondence algorithm for the exomars mission*, in *Field Programmable Logic and Applications (FPL), 2012 22nd International Conference on* (2012) pp. 667–670.
- [6] D. Diamantopoulos, I. Galanis, K. Siozios, G. Economakos, and D. Soudris, *A framework for rapid system-level synthesis targeting to reconfigurable platforms*, in *Workshop on Reconfigurable Computing (WRC), 2015, Netherlands*.
- [7] D. Diamantopoulos, K. Siozios, and D. Soudris, *Framework for performing rapid evaluation of 3d socs*, *Electronics Letters* **48**, 679 (2012).
- [8] D. Diamantopoulos, K. Siozios, D. Bekiaris, and D. Soudris, *A novel methodology for architecture-level exploration of 3d socs*, in *Design Technology of Integrated Systems in Nanoscale Era (DTIS), 2011 6th International Conference on* (2011) pp. 1–6.
- [9] D. Diamantopoulos, K. Siozios, S. Xydis, and D. Soudris, *A systematic methodology for reliability improvements on soc-based software defined radio systems*, *VLSI Des.* **2012**, 13:13 (2012).
- [10] D. Diamantopoulos, K. Siozios, S. Xydis, and D. Soudris, *Thermal optimization for micro-architectures through selective block replication*, in *Embedded Computer Systems (SAMOS), 2011 International Conference on* (2011) pp. 59–66.
- [11] E. Sotiriou-Xanthopoulos, D. Diamantopoulos, K. Siozios, G. Economakos, and D. Soudris, *A framework for rapid evaluation of heterogeneous 3-d noc architectures*, *Microprocessors and Microsystems* **38**, 292 (2014).

- [12] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, *Dark silicon and the end of multicore scaling*, in *Proceedings of the 38th Annual International Symposium on Computer Architecture*, ISCA '11 (ACM, New York, NY, USA, 2011) pp. 365–376.
- [13] N. Hardavellas, M. Ferdman, B. Falsafi, and A. Ailamaki, *Toward dark silicon in servers*, *Micro, IEEE* **31**, 6 (2011).
- [14] D. Diamantopoulos, S. Xydis, K. Siozios, and D. Soudris, *Mitigating memory-induced dark silicon in many-accelerator architectures*, *Computer Architecture Letters PP*, 1 (2015).
- [15] D. Diamantopoulos, S. Xydis, K. Siozios, and D. Soudris, *Dynamic memory management in vivado-hls for scalable many-accelerator architectures*, in *Applied Reconfigurable Computing*, Lecture Notes in Computer Science, Vol. 9040, edited by K. Sano, D. Soudris, M. Hübner, and P. C. Diniz (Springer International Publishing, 2015) pp. 117–128.
- [16] D. Diamantopoulos and C. Kachris, *High-level synthesizable dataflow mapreduce accelerator for fpga-coupled data centers*, in *Embedded Computer Systems (SAMOS), 2015 International Conference on* (2015).

Κεφάλαια σε Βιβλία

1. Konstantinos Tatas, Kostas Siozios, Dimitrios Soudris, Axel Jantsch, “Designing 2D and 3D Network-on-Chip Architectures”, Authors refer on “Chapter 9, On Designing 3-D Platforms” that: “This chapter was contributed by D. Diamantopoulos, Kostas Siozios, George Economakos, and Dimitrios Soudris of the School of ECE, National Technical University of Athens.”, pp.209-236, doi:10.1007/978-1-4614-4274-5_9, Springer, 2014.

Διεθνή Περιοδικά

7. D. Diamantopoulos, S. Xydis, K. Siozios, D. Soudris, “Mitigating Memory-induced Dark Silicon in Many-Accelerator Architectures,” IEEE Computer Architecture Letters , vol.PP, no.99, pp.1,1 doi: 10.1109/LCA.2015.2410791, March 2015.
6. D. Diamantopoulos, K. Siozios, S. Xydis, D. Soudris. “GENESIS: Parallel Application Placement onto Reconfigurable Architectures (Invited for the Special Issue on Runtime Management).” ACM Transactions on Embedded Computing Systems (TECS) vol. 14, no. 1: 18, doi: 10.1145/2629651, January 2015.
5. I. Kostavelis, L. Nalpantidis, E. Boukas, M. Aviles Rodrigalvarez, I. Stamoulias, G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, A. Gasteratos. “SPARTAN: Developing a vision system for future autonomous space exploration robots.” Journal of Field Robotics vol. 31, no. 1, pp.107-140. doi:10.1002/rob.21484, 2014.
4. E. Sotiriou-Xanthopoulos, D. Diamantopoulos, K. Siozios, G. Economakos, D. Soudris. “A framework for rapid evaluation of heterogeneous 3-D NoC architectures.” Elsevier Microprocessors and Microsystems vol. 38, no. 4, pp. 292-303, doi:10.1016/j.micpro.2013.09.003, June 2014.
3. D. Diamantopoulos, E. Sotiriou-Xanthopoulos, K. Siozios, G. Economakos, D. Soudris. “Plug&Chip: A Framework for Supporting Rapid Prototyping of 3D Hybrid Virtual SoCs”. ACM Transactions on Embedded Computing Systems (TECS), vol. 13, no. 5s, Article 168, pp. 1-25, 25 pages, doi:10.1145/2661634, December 2014.

2. D. Diamantopoulos, K. Siozios, D. Soudris, “Framework for performing rapid evaluation of 3D SoCs,” *IET Electronics Letters*, vol.48, no.12, pp.679,681, doi: 10.1049/el.2012.1321, June 2012.
1. D. Diamantopoulos, K. Siozios, S. Xydis, and D. Soudris. “A systematic methodology for reliability improvements on SoC-based software defined radio systems”, *Hindawi VLSI Design*, Article ID 784945, 15 pages, doi: 10.1155/2012/784945, January 2012.

Διεθνή Συνέδρια με Κριτές

16. D. Diamantopoulos, C. Kachris. ”High-level Synthesizable Dataflow MapReduce Accelerator for FPGA-coupled Data Centers”, *Embedded Computer Systems (SAMOS)*, 2015 International Conference on, Accepted for publication. To be presented in July 2015.
15. D. Diamantopoulos, S. Xydis, K. Siozios, D. Soudris. ”Dynamic memory management in Vivado-hls for scalable many-accelerator architectures.” In *Applied Reconfigurable Computing (ARC)*, pp. 117-128. Springer International Publishing, March 2015.
14. E. Sotiriou-Xanthopoulos, D. Diamantopoulos, G. Economakos. ”Evaluation of High-Level Synthesis Techniques for Memory and Datapath Tradeoffs in FPGA Based SoC Architectures.” In *Applied Reconfigurable Computing (ARC)*, pp. 321-330. Springer International Publishing, March 2015.
13. K. Siozios, P. Figuli, H. Sidiropoulos, C. Tradowsky, D. Diamantopoulos, K. Maragos, S. Percy Delicia, D. Soudris, J. Becker. ”TEACHER: TEAch AdvanCED Reconfigurable Architectures and Tools.” In *Applied Reconfigurable Computing (ARC)*, pp. 103-114. Springer International Publishing, March 2015.
12. G. Lentaris, I. Stamoulias, D. Diamantopoulos, K. Maragos, K. Siozios, D. Soudris, M. Aviles Rodrigalvarez, M. Lourakis, X. Zabulis, I. Kostavelis, L. Nalpantidis, E. Boukas, A. Gasteratos, ”SPARTAN/SEXTANT/COMPASS: Advancing Space Rover Vision via Reconfigurable Platforms.” In *Applied Reconfigurable Computing (ARC)*, pp. 475-486. Springer International Publishing, March 2015.
11. D. Diamantopoulos, G. Economakos, D. Reisis, ”Using high-level synthesis to build memory and datapath optimized DSP accelerators,” *Electronics, Circuits and Systems (ICECS)*, 2014 21st IEEE International Conference on, pp.714,717, doi: 10.1109/ICECS.2014.7050085, December 2014.
10. D. Diamantopoulos, C. Economakos, D. Soudris, G. Economakos, ”A new design paradigm for floating point DSP applications based on ESL/HLS

- and FPGAs,” Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on, pp.000404,000409, 12-15, doi: 10.1109/ISSPIT.2013.6781915, December 2013.
9. D. Diamantopoulos, K. Siozios, E. Sotiriou-Xanthopoulos, G. Economakos, D. Soudris, ”HVSoCs: A Framework for Rapid Prototyping of 3-D Hybrid Virtual System-on-Chips,” Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2013 IEEE 27th International, pp.2194,2199, 20-24, doi: 10.1109/IPDPSW.2013.202, May 2013.
 8. G. Lentaris, D. Diamantopoulos, G. Stamoulias, K. Siozios, D. Soudris, M.A. Rodrigalvarez, ”FPGA-based path-planning of high mobility rover for future planetary missions,” Electronics, Circuits and Systems (ICECS), 2012 19th IEEE International Conference on, pp.85,88, doi: 10.1109/ICECS.2012.6463793, December 2012.
 7. D. Diamantopoulos, K. Siozios, G. Lentaris, D. Soudris, M.A. Rodrigalvarez, ”SPARTAN project: On profiling computer vision algorithms for rover navigation,” Adaptive Hardware and Systems (AHS), 2012 NASA/ESA Conference on, pp.174,181, doi: 10.1109/AHS.2012.6268647, June 2012.
 6. G. Lentaris, D. Diamantopoulos, K. Siozios, D. Soudris, M.A. Rodrigalvarez, ”Hardware implementation of stereo correspondence algorithm for the Exo-Mars mission,” Field Programmable Logic and Applications (FPL), 2012 22nd International Conference on , pp.667,670, doi: 10.1109/FPL.2012.6339173, August 2012.
 5. D. Diamantopoulos, P. Galiatsatos, A. Karachalios, G. Lentaris, D. Reisis, D. Soudris, ”Configurable baseband digital transceiver for Gbps wireless 60 GHz communications,” Electronics, Circuits and Systems (ICECS), 2011 18th IEEE International Conference on, pp.192,195, doi: 10.1109/ICECS.2011.6122246, December 2011.
 4. E. Sotiriou-Xanthopoulos, D. Diamantopoulos, G. Economakos, D. Soudris, ”Design and experimentation with low-power morphable multipliers,” Electronics, Circuits and Systems (ICECS), 2011 18th IEEE International Conference on , vol., no., pp.752,755, doi: 10.1109/ICECS.2011.6122383, December 2011.
 3. K. Siozios, D. Diamantopoulos, I. Kostavelis, E. Boukas, L. Nalpantidis, D. Soudris, A. Gasteratos, M. Aviles, I. Anagnostopoulos, ”SPARTAN project: Efficient implementation of computer vision algorithms onto reconfigurable platform targeting to space applications,” in Proceedings of the 6th International Workshop on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC), doi: 10.1109/ReCoSoC.2011.5981524, pp.1,9, June 2011.

2. D. Diamantopoulos, K. Siozios, S. Xydis, D. Soudris, "Thermal optimization for micro-architectures through selective block replication," *Embedded Computer Systems (SAMOS)*, 2011 International Conference on, pp.59,66, doi: 10.1109/SAMOS.2011.6045445, July 2011.
1. D. Diamantopoulos, K. Siozios, D. Bekiaris, D. Soudris, "A novel methodology for architecture-level exploration of 3D SoCs," *Design & Technology of Integrated Systems in Nanoscale Era (DTIS)*, 2011 6th International Conference on, pp.1,6, doi: 10.1109/DTIS.2011.5941425, April 2011

Διεθνή Workshops

14. D. Diamantopoulos, S. Xydis, K. Siozios and D. Soudris, "High-Level-Synthesis Extensions for Scalable Single-Chip Many-Accelerators on FPGAs", *International Workshop on Reconfigurable Computing for High-Performance Computing (HPC) and High-Performance Data Analytics (HPDA) (ReC4P)*, August 2015, London, UK.
13. D. Diamantopoulos, I. Galanis, K. Siozios, G. Economakos, and D. Soudris, "A Framework for Rapid System-Level Synthesis Targeting to Reconfigurable Platforms", *Workshop on Reconfigurable Computing (WRC)*, 2015.
12. D. Diamantopoulos, K. Siozios, E. Sotiriou-Xanthopoulos, G. Economakos and D. Soudris, "HVSoCs: A Framework for Rapid Prototyping of 3-D Hybrid Virtual System-on-Chips", *Workshop on Virtual Prototyping of Parallel and Embedded Systems (VIPES)*, May 20-24th, 2013.
11. D. Diamantopoulos, P. Galiatsatos, A. Karachalios, G. Lentaris, D. Reisis and D. Soudris, *A Reconfigurable Baseband Architecture for Gbps Wireless 60 GHz Communications*, *Fifth Friday Workshop on Designing for Embedded Parallel Computing Platforms (DEPCP)*, March 22, 2013, Grenoble, France.
10. G. Lentaris, D. Diamantopoulos, K. Siozios, I. Stamoulias, I. Kostavelis, E. Boukas, L. Nalpantidis, D. Soudris, A. Gasteratos, and M. Aviles, "SPARTAN: Efficient Implementation of Computer Vision Algorithms for Autonomous Rover Navigation", *Workshop on Reconfigurable Computing (WRC)*, Jan. 2013, Germany.
9. G. Lentaris, I. Stamoulias, D. Diamantopoulos, K. Siozios, and D. Soudris, "An FPGA implementation of the SURF algorithm for the ExoMars programme", *Workshop on Reconfigurable Computing (WRC)*, Jan. 2013, Germany.
8. D. Diamantopoulos, K. Siozios, and D. Soudris, "A Framework for Performing Fault-Tolerant Placement Based on Genetic Algorithm", *Workshop on Reconfigurable Computing (WRC)*, Jan. 2013, Germany.

7. D. Diamantopoulos, K. Siozios, S. Xydis and D. Soudris, A genetic algorithm-based FPGA placer for multi-core processors, Fifth Friday Workshop on Designing for Embedded Parallel Computing Platforms (DEPCP), March 22, 2013, Grenoble, France
6. D. Diamantopoulos, K. Siozios, I. Stamoulias, G. Lentaris, D. Soudris and M. Aviles, Towards Computer Vision FPGA Acceleration, DATE Friday Workshop on Reconfigurable Computing (Configcomp), March 22, 2013, Grenoble, France.
5. D. Diamantopoulos, K. Siozios, S. Xydis and D. Soudris, “A Framework for Supporting Parallel Application Placement onto Reconfigurable Platforms”, Workshop on Parallel Programming and Run-time Management Techniques for Many-core Architectures (PARMA), Jan. 2013.
4. K. Siozios, H. Sidiropoulos, D. Diamantopoulos, P. Figuli, D. Soudris, M. Hubner and J. Becker, “On Designing Self-Aware Reconfigurable Platforms”, Workshop on Self-Awareness in Reconfigurable Computing Systems (SRCS), pp. 14-17, Norway, 2012.
3. D. Diamantopoulos, G. Lentaris, K. Siozios, D. Soudris and M. Aviles, “Towards Accelerating Computer Vision Algorithms Targeting to Space Applications with a Heterogeneous Platform”, Friday Workshop on Designing for Embedded Parallel Computing Platforms: Architectures, Design Tools, and Applications at DATE 2012, Germany, 2012.
2. K. Siozios, D. Diamantopoulos, H. Sidiropoulos, A. Papanikolaou, and D. Soudris, Rapid Evaluation of 3-D Interconnection Schemes , DATE 2011 3D Integration Workshop, Grenoble, 2011, France.
1. M. Aviles, K. Siozios, D. Diamantopoulos, L. Nalpantidis, I. Kostavelis, E. Boukas, D. Soudris and A. Gasteratos, A Co-design Methodology for Implementing Computer Vision Algorithms for Rover Navigation onto Reconfigurable Hardware, Workshop on Computer Vision on Low-Power Reconfigurable Architectures, International Conference on Field Programmable Logic and Applications, Sept. 2011, Chania, Greece.

Βραβεία - Διακρίσεις

6. Cadence Design Systems: 2^η θέση στο διαγωνισμό “Cadence Thesis Contest for Automotive Embedded Systems”, 2013.
5. Βράβευση καλύτερης εργασίας (Best Paper Award) στο 4th Workshop on Parallel Programming and Run-Time Management Techniques for Many-core Architectures (PARMA) January 23, 2013, Berlin, Germany.

4. University Booth στο συνέδριο IEEE International Conference on Design Automation and Test in Europe (DATE 2013).
3. Πρόσκληση δημοσίευσης στο περιοδικό “ACM Transactions on Embedded Computing Systems (TECS)” λόγω υψηλής βαθμολογίας κρίσης εργασίας στο “Workshop on Virtual Prototyping of Parallel and Embedded Systems (VIPES), 2013”
2. University Booth στο συνέδριο IEEE International Conference on Design Automation and Test in Europe (DATE 2012).
1. HiPEAC: Επιχορήγηση συμμετοχής για το “International Summer School (7 days) on Advanced Computer Architecture and Compilation for High-Performance and Embedded Systems”, Fiuggi, Italy, July 2011.

Βιογραφικό



Ο Διονύσιος Διαμαντόπουλος γεννήθηκε την 12η Φεβρουαρίου του έτους 1985, στην Αθήνα. Μετά από εισαγωγικές πανελλήνιες εξετάσεις στο Ενιαίο Λύκειο Πύλου Μεσσηνίας, εισήχθη στο Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, της Πολυτεχνικής Σχολής του Πανεπιστημίου Πατρών. Το έτος 2009 έλαβε από το τμήμα αυτό το δίπλωμα του Μηχανικού Ηλεκτρονικών Υπολογιστών και Πληροφορικής.

Το έτος 2009 έγινε δεκτός στο μεταπτυχιακό πρόγραμμα της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Την ίδια περίοδο έγινε μέλος του εργαστηρίου Μικροϋπολογιστών και Ψηφιακών Συστημάτων VLSI και ξεκίνησε τη διδακτορική του διατριβή στο ερευνητικό πεδίο των επαναδιαμορφούμενων & εξειδικευμένων ενσωματωμένων συστημάτων, υπό την επίβλεψη του Καθηγητή κ. Δημήτριου Σούντρη.

Από το έτος 2009 εργάζεται ως ερευνητής στο Ερευνητικό Πανεπιστημιακό Ινστιτούτο Συστημάτων Επικοινωνιών & Υπολογιστών (ΕΠΙΣΕΥ) σε ευρωπαϊκά ερευνητικά έργα. Επίσης έως σήμερα εργάζεται ως βασικός ερευνητής σε ερευνητικά έργα της Ευρωπαϊκής Υπηρεσίας Διαστήματος (ESA) καθώς και σε ερευνητικά έργα του Εθνικού Στρατηγικού Πλαισίου Αναφοράς (ΕΣΠΑ). Οι αναφερθείσες δραστηριότητες εξασφάλισαν χρηματοδότηση της διδακτορικής του διατριβής από ευρωπαϊκά και εθνικά κεφάλαια.

Ο Διονύσιος, έως σήμερα, έχει δημοσιεύσει επτά (7) άρθρα σε διεθνή επιστημονικά περιοδικά των εκδοτικών οίκων IEEE, ACM, Elsevier και Hindawi, εικοσιεννιά (29) άρθρα σε διεθνή συνέδρια και workshops. Από τις δημοσιεύσεις του έχει λάβει, έως σήμερα, δεκατέσσερις (14) ετεροαναφορές. Επίσης έχει συνεπιβλέψει επτά (7) διπλωματικές εργασίες προπτυχιακών φοιτητών. Παράλληλα, προσφέρει υπηρεσίες στις ερευνητικές κοινότητες IEEE και ACM, ως κριτής σε διεθνή συνέδρια και περιοδικά.

Τα ερευνητικά του ενδιαφέροντα περιλαμβάνουν τις επαναδιαμορφούμενες αρχιτεκτονικές, τους αλγόριθμους πολυ-κριτηριακής βελτιστοποίησης για εργαλεία EDA, τις αρχιτεκτονικές πολλαπλών επιταχυντών υλισμικού και τις τεχνικές υψηλού επιπέδου σύνθεσης για επαναδιατάξιμες ετερογενείς πλατφόρμες.

Ο Διονύσιος είναι μέλος του Τεχνικού Επιμελητηρίου Ελλάδας (ΤΕΕ), του ευρωπαϊκού οργανισμού HiPEAC και της επιστημονικής κοινότητας IEEE.

