

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Εργαστήριο Υπολογιστικών Συστημάτων
www.cslab.ece.ntua.gr

Διπλωματική εργασία

*Μελέτη και αξιολόγηση εναλλακτικών τεχνικών αποθήκευσης αραιών
πινάκων βασισμένων σε κανονικότητες*

Καθηγητής: Νεκτάριος Κοζύρης (nkoziris@cslab.ece.ntua.gr)

Επικοινωνία: Βασίλειος Καρακάσης (bkk@cslab.ece.ntua.gr)

Γιώργος Γκούμας (goumas@cslab.ece.ntua.gr)

Άτομα: 1

Εισαγωγή

Οι αραιοί πίνακες αποτελούν αναπόσπαστο κομμάτι μίας πληθώρας επιστημονικών εφαρμογών, οι οποίες πολύ συχνά καταλήγουν στον επαναληπτικό υπολογισμό ενός πολλαπλασιασμού αραιού πίνακα με διάνυσμα. Ο Πολλαπλασιασμός Αραιού Πίνακα με Διάνυσμα (ΠΑΠΔ) (Sparse Matrix-Vector Multiplication – SpMV) αποτελεί εδώ και χρόνια ισχυρό πεδίο ενδιαφέροντος στην κοινότητα των Υπολογιστικών Συστημάτων Υψηλών Επιδόσεων. Ο λόγος πίσω από αυτό το ιδιαίτερο ενδιαφέρον είναι ότι τα υπολογιστικά χαρακτηριστικά αυτού του πυρήνα είναι τέτοια, ώστε η επιτυγχάνουσα επίδοση να είναι μόνο ένα πολύ μικρό ποσοστό της μέγιστης επίδοσης των σύγχρονων υπολογιστικών συστημάτων.

Ο κύριος παράγοντας που εμποδίζει την επίτευξη υψηλών επιδόσεων για τον πυρήνα ΠΑΠΔ είναι περισσότερο οι μεγάλες του απαιτήσεις σε εύρος ζώνης μνήμης (memory bandwidth) παρά σε επεξεργαστική ισχύ. Οι απαιτήσεις αυτές γίνονται ιδιαίτερα εμφανείς σε αρχιτεκτονικές μοιραζόμενης μνήμης, όπου τέσσερις ή οκτώ πυρήνες μοιράζονται τον κοινό δίαυλο προς την κύρια μνήμη. Αυτό έχει ως αποτέλεσμα ο ΠΑΠΔ στην κλασική του μορφή να μην παρουσιάζει σχεδόν καθόλου επιτάχυνση πάνω από τέσσερα νήματα λόγω συμφόρησης στον κοινό δίαυλο δεδομένων.

Για την αντιμετώπιση των αυξημένων απαιτήσεων του ΠΑΠΔ σε εύρος ζώνης μνήμης έχουν προταθεί διάφορες βελτιστοποιήσεις του τρόπου αποθήκευσης (αποθήκευση σε μπλοκ, συμπίεση δεδομένων κ.ά.), ώστε να μειωθούν κατά το δυνατόν περισσότερο τα δεδομένα που πρέπει να μεταφερθούν από την κύρια μνήμη κατά την διάρκεια εκτέλεσης του πυρήνα. Η ιδέα πίσω από αυτές τις βελτιστοποιήσεις είναι

ότι οι αραιοί πίνακες –παρόλο που η ακριβής δομή τους δεν είναι γνωστή πριν τον χρόνο εκτέλεσης– παρουσιάζουν κάποιες κανονικότητες (patterns), π.χ., πυκνά μπλοκ στοιχείων, συνεχόμενα στοιχεία κατά μία διεύθυνση κ.ά. Γνωρίζοντας, επομένως, αυτές τις κανονικότητες, μπορεί κανείς χρησιμοποιώντας κάποια μετα-πληροφορία ανά κανονικότητα, και όχι ανά στοιχείο, να μειώσει τον όγκο των δεδομένων που πρόκειται να μεταφερθούν από την κύρια μνήμη, πράγμα που, τελικά, βοηθάει στην επίδοση.

Σκοπός

Ο σκοπός της συγκεκριμένης διπλωματικής εργασίας είναι να μελετηθεί και να αξιολογηθεί ένας νέος τρόπος αποθήκευσης αραιών πινάκων βασισμένος σε κανονικότητες (Pattern-Based Row format – PBR) σε σχέση με πιο κλασσικές τεχνικές βελτιστοποίησης αραιών πινάκων, όπως είναι οι μπλοκ μέθοδοι (βιβλιοθήκη OSKI). Ενδεικτικά, τα στάδια υλοποίησης της διπλωματικής εργασίας μπορούν να συνοψιστούν στα εξής:

1. Μελέτη σχετικής βιβλιογραφίας.
2. Εγκατάσταση βιβλιοθηκών PBR και OSKI και πιθανή τροποποίησή τους, ώστε να υποστηρίζονται διάφορα σχήματα παραλληλίας.
3. Οργάνωση και διάρθρωση πειραμάτων ώστε να ληφθούν αποτελέσματα επίδοσης και χρόνου προεπεξεργασίας των δεδομένων σε διάφορες παράλληλες αρχιτεκτονικές υπολογιστών.
4. Μελέτη και ανάλυση αποτελεσμάτων.
5. Συγγραφή διπλωματικής εργασίας.

Προαπαιτούμενη γνώση και γνώση που θα αποκτηθεί

Για την εκπόνηση της διπλωματικής εργασίας απαιτείται εξοικείωση με τις γλώσσες προγραμματισμού C/C++. Επιθυμητή είναι η γνώση μεταφραζόμενων γλωσσών, π.χ., Bash, Python, ενώ επίσης επιθυμητή είναι κάποια εξοικείωση με τις αρχιτεκτονικές μοιραζόμενης μνήμης.

Κατά την διάρκεια της διπλωματικής εργασίας θα αποκτηθεί βαθύτερη γνώση σε θέματα βελτιστοποίησης κώδικα για αρχιτεκτονικές μοιραζόμενης μνήμης, σε τεχνικές παραγωγής κώδικα στον χρόνο εκτέλεσης και σε προχωρημένες τεχνικές αποθήκευσης αραιών πινάκων.

Χρήσιμοι σύνδεσμοι

- M. Belgin, G. Back, and C. J. Ribbens. Pattern-based sparse matrix representation for memory-efficient SMVM kernels. In *Proceedings of the 23rd International Conference on Supercomputing (ICS'09)*. Yorktown Heights, NY, USA, 2009.
- M. Belgin, G. Back, and C. J. Ribbens. A Library for Pattern-based Sparse Matrix Vector Multiply. Technical Report, <http://eprints.cs.vt.edu/archive/00001103/>
- OSKI: Optimized Sparse Kernel Interface. <http://bebop.cs.berkeley.edu/oski/>.