

Software & Programming of HPC Systems

ΕΡΓΑΣΙΑ 3: OPENMP and GPUs

Φλουρής Παντελής, up1093507

Κολτσάκης Χρυσάφης, up1084671

1. OpenMP and Complex Matrix Multiplication

Η υλοποίηση με χρήση OpenMP GPU είναι πολύ απλή, καθώς απαιτούνται μονάχα μικρές αλλαγές στον σειριακό CPU κώδικα που είχαμε ήδη απο το 2ο set.

Αντιγράψαμε τις συναρτήσεις της CPU, τις μετονομάσαμε σε GPUvecAdd, GPUvecSub, GPUmatMult και προσθέσαμε πριν τους βρόγχους for το OpenMP directive

#pragma omp target teams distribute parallel for simd

βεβαιώνοντας ότι οι βρόγχοι αποστέλλονται σε ομάδες νημάτων της κάρτας γραφικών, και εκμεταλλεύονται τις simd δυνατότητες της.

Στην main, οι εντολές για τους ζητούμενους υπολογισμούς στην GPU βρίσκονται μέσα στο OpenMP directive

***#pragma omp target data map(to: A, B, C, D) map(alloc: temp)
map(from: E_gpu, F_gpu)***

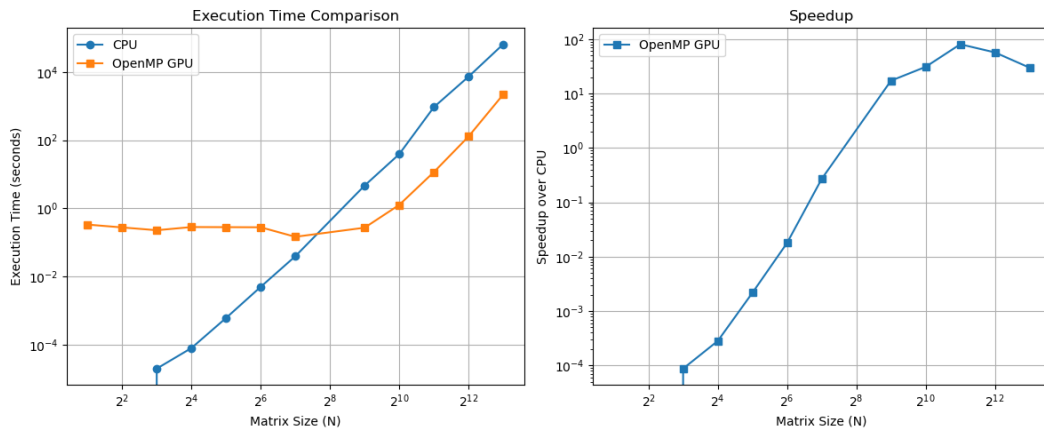
Έτσι αντιγράφουμε τα μητρώα A, B, C, D που χρειάζονται για τους υπολογισμούς, δεσμεύουμε μνήμη για την ενδιάμεση μεταβλητή που απαιτείται, και τα αποτελέσματα αντιγράφονται στην CPU.

Με αυτό τον τρόπο αποφεύγουμε όσο μπορούμε χρονοβόρες αντιγραφές απο και πρὸς την GPU, βελτιώνοντας την απόδοση.

Αποτελέσματα

Η βελτίωση συγκριτικά με την CPU είναι προφανής και μεγάλη, αλλά η υλοποίηση με OpenMP είναι **σημαντικά** πιο αργή από την αντίστοιχη με CUDA της προηγούμενης άσκησης.

Π.χ. Για μητρώα 4096x4096, η υλοποίηση με cuBLAS υπολογίζει τα αποτελέσματα σε **0.33s**, η OpenMP χρειάζεται **128s (!)**.



Η χρονοβελτίωση φαίνεται να σταματά γύρω στο 100, ενώ στην CUDA υλοποίηση του προηγούμενου υποερωτήματος, η πιο αργή υλοποίηση ξεπερνά το 1000.

