

**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΑΡΑΛΛΗΛΑ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ**

**Heat 2D (MPI-OpenMP)**

**Χρήστος Πατσούρας Μ1612**

**ΑΘΗΝΑ**

**ΜΑΡΤΙΟΣ 2019**

**ΠΕΡΙΕΧΟΜΕΝΑ**

[1. ΕΙΣΑΓΩΓΗ 6](#_Toc6927575)

[2. ΣΧΕΔΙΑΣΜΟΣ ΔΙΑΜΟΙΡΑΣΜΟΥ ΔΕΔΟΜΕΝΩΝ ΣΤΙΣ ΔΙΕΡΓΑΣΙΕΣ 7](#_Toc6927576)

[2.1 Partitioning 7](#_Toc6927577)

[2.2 Communication 7](#_Toc6927578)

[2.3 Agglomeration 7](#_Toc6927579)

[2.3.1 Διαχωρισμός σε λωρίδες 8](#_Toc6927580)

[2.3.2 Διαχωρισμός σε blocks 9](#_Toc6927581)

[2.3.3 Σύγκριση διαχωρισμών 9](#_Toc6927582)

[2.4 Mapping 12](#_Toc6927583)

[3. ΣΧΕΔΙΑΣΜΟΣ ΚΩΔΙΚΑ 14](#_Toc6927584)

[3.1 Σχεδιασμός MPI προγράμματος 14](#_Toc6927585)

[3.2 Σχεδιασμός υβριδικού προγράμματος 17](#_Toc6927586)

[4. ΜΕΤΡΗΣΕΙΣ 18](#_Toc6927587)

[4.1 MPI πρόγραμμα χωρίς έλεγχο σύγκλισης 18](#_Toc6927588)

[4.2 MPI πρόγραμμα με έλεγχο σύγκλισης κάθε ν επαναλήψεις 19](#_Toc6927589)

[4.3 Υβριδικό MPI με OpenMP 21](#_Toc6927590)

[5. ΣΥΓΚΡΙΣΗ ΜΕ ΑΡΧΙΚΟ ΠΡΟΓΡΑΜΜΑ 23](#_Toc6927591)

[5.1 Σύγκριση πραγματικών υπολογισμών 23](#_Toc6927592)

[5.2 Σύγκριση αναλυτικών μετρήσεων 24](#_Toc6927593)

[6. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΠΑΡΑΤΗΡΗΣΕΙΣ – ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ 28](#_Toc6927594)

[7. ΕΞΟΔΟΣ mpiP 29](#_Toc6927595)

**ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ**

Σχήμα 1: Σύγκριση χρόνου εκτέλεσης με βάση το μέγεθος δεδομένων μεταξύ μπλοκ και λωρίδων 10

Σχήμα 2: Σύγκριση επιτάχυνσης με βάση το μέγεθος δεδομένων μεταξύ μπλοκ και λωρίδων 10

Σχήμα 3: Σύγκριση αποδοτικότητας με βάση το μέγεθος δεδομένων μεταξύ μπλοκ και λωρίδων 11

Σχήμα 4: Σύγκριση χρόνου εκτέλεσης με βάση το πλήθος των επεξεργαστών μεταξύ μπλοκ και λωρίδων 11

Σχήμα 5: Σύγκριση επιτάχυνσης με βάση το πλήθος των επεξεργαστών μεταξύ μπλοκ και λωρίδων 12

Σχήμα 6: Σύγκριση αποδοτικότητας με βάση το πλήθος των επεξεργαστών μεταξύ μπλοκ και λωρίδων 12

**ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ**

[Εικόνα 1: Γείτονες σημείου 6](#_Toc6927602)

[Εικόνα 2: Απλό μοντέλο επικοινωνίας 8](#_Toc6927603)

[Εικόνα 3: Δισδιάστατο καρτεσιανό πλέγμα διεργασιών 14](#_Toc6927604)

[Εικόνα 4: Πίνακας u και διαχωρισμός του σε blocks 15](#_Toc6927605)

[Εικόνα 5: Εσωτερικά, εξωτερικά και αλώ σημεία υποπίνακα 16](#_Toc6927606)

**ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ**

[Πίνακας 1: Μετρήσεις χρόνου MPI προγράμματος χωρίς έλεγχο σύγκλισης 18](#_Toc6927607)

[Πίνακας 2: Επιτάχυνση MPI προγράμματος χωρίς έλεγχο σύγκλισης 19](#_Toc6927608)

[Πίνακας 3: Αποδοτικότητα MPI προγράμματος χωρίς έλεγχο σύγκλισης 19](#_Toc6927609)

[Πίνακας 4: Μετρήσεις χρόνου MPI προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις 20](#_Toc6927610)

[Πίνακας 5: Επιτάχυνση MPI προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις 20](#_Toc6927611)

[Πίνακας 6: Αποδοτικότητα MPI προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις 21](#_Toc6927612)

[Πίνακας 7: Μετρήσεις χρόνου υβριδικού προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις 21](#_Toc6927613)

[Πίνακας 8: Επιτάχυνση υβριδικού προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις 22](#_Toc6927614)

[Πίνακας 9: Αποδοτικότητα υβριδικού προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις 22](#_Toc6927615)

[Πίνακας 10: Μετρήσεις χρόνου αρχικού προγράμματος προγράμματος χωρίς έλεγχο σύγκλισης 23](#_Toc6927616)

[Πίνακας 11: Επιτάχυνση υλοποίησή μας σε σχέση με το αρχικό πρόγραμμα 23](#_Toc6927617)

[Πίνακας 12: Θεωρητική μέτρηση χρόνου για διαχωρισμό σε λωρίδες 24](#_Toc6927618)

[Πίνακας 13: Θεωρητική μέτρηση επιτάχυνσης για διαχωρισμό σε λωρίδες 24](#_Toc6927619)

[Πίνακας 14: Θεωρητική μέτρηση αποδοτικότητας για διαχωρισμό σε λωρίδες 25](#_Toc6927620)

[Πίνακας 15: Θεωρητική μέτρηση χρόνου για διαχωρισμό σε μπλοκ 25](#_Toc6927621)

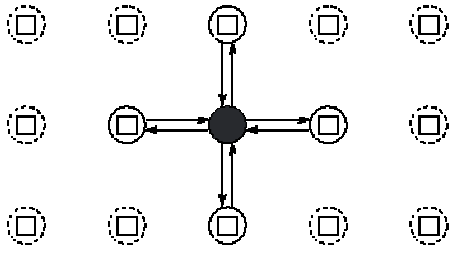
[Πίνακας 16: Θεωρητική μέτρηση επιτάχυνσης για διαχωρισμό σε μπλοκ 26](#_Toc6927622)

[Πίνακας 17: Θεωρητική μέτρηση αποδοτικότητας για διαχωρισμό σε μπλοκ 26](#_Toc6927623)

# ΕΙΣΑΓΩΓΗ

Στόχος της εφαρμογής είναι η προσομοίωση μεταφοράς θερμότητας σε επιφάνεια. Στην ουσία έχουμε ένα πίνακα διάστασης , όπου κάθε στοιχείο του πίνακα είναι ένα σημείο του χώρου και έχει μια συγκεκριμένη θερμοκρασία. Η αρχική θερμοκρασία είναι υψηλή στο κέντρο και μηδέν στα ακριανά στοιχεία. Το σύστημα αλλάζει κατάσταση με την πάροδο του χρόνου. Αυτό συμβαίνει διότι κάθε σημείο επηρεάζει και επηρεάζεται από τα γειτονικά του. Γειτονικά στοιχεία θεωρούνται αυτά που βρίσκονται πάνω, κάτω, δεξιά και αριστερά σε ένα στοιχείο, όπως στην Εικόνα 1. Τα στοιχεία που βρίσκονται στα άκρα δεν αλλάζουν τιμές καθώς θεωρούνται στοιχεία που απορροφούν ή εκπέμπουν θερμότητα στο σύστημα.

Το πρόγραμμα που δίνεται ως πρότυπο για τις απαιτήσεις της άσκησης είναι διαθέσιμο [εδώ](https://computing.llnl.gov/tutorials/mpi/samples/C/mpi_heat2D.c) και ζητούμενο είναι η αξιολόγηση, ο επανασχεδιασμός και η βελτίωση του με στόχο την καλύτερη κλιμάκωση και η σύγκριση χρονικά, σε speedup και σε efficiency του προγράμματος αυτού με αυτά που θα παραχθούν σε MPI και σε MPI+OpenMP (υβριδικό).



Εικόνα 1: Γείτονες σημείου

# ΣΧΕΔΙΑΣΜΟΣ ΔΙΑΜΟΙΡΑΣΜΟΥ ΔΕΔΟΜΕΝΩΝ ΣΤΙΣ ΔΙΕΡΓΑΣΙΕΣ

Partitioning

Εν γένει το partitioning μπορεί να γίνει ως προς τα δεδομένα και ως προς τους υπολογισμούς ή τη λειτουργία που πρέπει να επιτελεστεί. Σε αυτή την περίπτωση, το partitioning μπορεί να γίνει μόνο ως προς τα δεδομένα υπάρχει ένας μοναδικός υπολογισμός για όλα τα δεδομένα. Συγκεκριμένα με διάφορους τρόπους μπορούμε να αποφασίσουμε πως θα διαχωρίσουμε τα δεδομένα του πίνακα. Εφόσον ο φόρτος επεξεργασίας είναι ίδιος για όλο τον πίνακα δεν υπάρχει λόγος να γίνει κάποιος κυκλικός διαμοιρασμός. Οπότε τον πίνακα θα τον χωρίσουμε είτε σε γραμμές είτε σε στήλες είτε σε blocks. Στο δοσμένο πρόγραμμα ο διαχωρισμός γίνεται σε γραμμές. Εμείς θα αποδείξουμε ότι ο διαχωρισμός σε blocks είναι ορθότερος ως προς την κλιμάκωση και θα τον ακολουθήσουμε και στην υλοποίηση του νέου προγράμματος.

Communication

Όπως έχουμε αναφέρει, η τιμή ενός στοιχείου εξαρτάται από τα τέσσερα γειτονικά του σε διάταξη βοράς-νότος-ανατολή-δύση. Άρα σε περίπτωση διαχωρισμού του πίνακα, οι διεργασίες πρέπει να ανταλλάσσουν πληροφορίες των στοιχείων που χρειάζονται οι «γειτονικές» διεργασίες για τους υπολογισμούς που πραγματοποιούν. Αυτές θα είναι είτε οι ακραίες γραμμές είτε οι ακραίες στήλες στις περιπτώσεις διαχωρισμού σε οριζόντιες λωρίδες και σε κάθετες λωρίδες αντίστοιχα και συγκεκριμένα η πρώτη και η τελευταία κάθε υποπίνακα. Στην περίπτωση διαχωρισμού σε blocks θα είναι όλα τα ακραία στοιχεία, συγκεκριμένα η πρώτη και η τελευταία γραμμή και στήλη κάθε υποπίνακα. Θεωρούμε ότι η επικοινωνία μεταξύ των υπολογιστών είναι ίδιων τεχνικών χαρακτηριστικών και ότι δεν υπάρχει επικοινωνιακός φόρτος από άλλες εφαρμογές.

Η επικοινωνία στο συγκεκριμένο πρόβλημα είναι:

* Local: επικοινωνούμε με ένα μικρό τμήμα των διεργασιών, όχι με όλες
* Structured: Ακολουθούμε καρτεσιανή τοπολογία για να δομίσουμε την επικοινωνία και μιλάμε μόνο με τέσσερις γείτονες στην περίπτωση του διαχωρισμού σε blocks. Στην περίπτωση του διαχωρισμού σε γραμμές ή στήλες έχουμε μόνο δύο γείτονες, όπως στο παράδειγμα που δίνεται.
* Static: οι γείτονες είναι πάντα οι ίδιοι, δεν αλλάζουν με την πάροδο του χρόνου
* Synchronous: οι παραγωγοί και οι καταναλωτές των μηνυμάτων εκτελούνται συντονισμένα και συνεργάζονται για τη μεταφορά δεδομένων

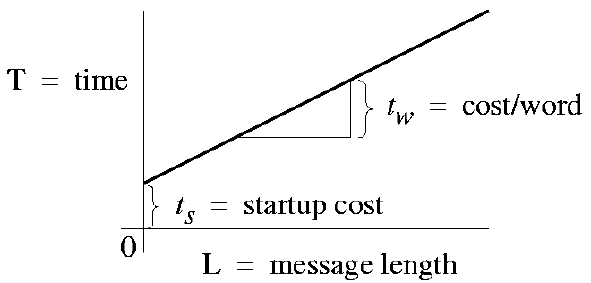
Στη δική μας υλοποίηση θα ακολουθήσουμε το διαχωρισμό σε blocks και τη διάταξη με τους τέσσερις γείτονες για κάθε διεργασία.

## Agglomeration

Θα εξετάσουμε και τις δύο προαναφερθείσες περιπτώσεις διαχωρισμού, δηλαδή σε λωρίδες (γραμμές ή στήλες) και blocks, και θα αποδείξουμε ότι ο διαχωρισμός σε blocks κλιμακώνει καλύτερα. Έστω ότι έχουμε ένα πίνακα . Χωρίς βλάβη της γενικότητας θα θεωρήσουμε ότι . Αφού έχουμε για να έχουμε λιγότερα στοιχεία για ανταλλαγή, ο διαχωρισμός σε λωρίδες θα είναι σε στήλες. Δε θα είχε διαφορά ο διαχωρισμός σε γραμμές. Ο συνολικός χρόνος εκτέλεσης ενός παράλληλου προγράμματος υπολογίζεται ως εξής:

όπου : χρόνος υπολογισμών, : χρόνος επικοινωνίας και : χρόνος αδράνειας/αναμονής.

Η διάρκεια του χρόνου αναμονής δε μπορεί να προβλεφθεί για αυτό και δε θα υπολογιστεί στο αποτέλεσμα. Αφού έχουμε ομοιογενή συστήματα, το για ένα στοιχείο θα είναι παντού ίδιο και δεν εξαρτάται από τον διαμοιρασμό, αλλά από τη φύση του προβλήματος. Επίσης για κάθε επικοινωνία που θα κάνει κάθε διεργασία το κόστος υπολογίζεται μέσω του τύπου



Εικόνα 2: Απλό μοντέλο επικοινωνίας

### Διαχωρισμός σε λωρίδες

Στην περίπτωση που γίνει ο διαχωρισμός σε λωρίδες, αφού , ο διαχωρισμός θα γίνει σε στήλες για να έχουμε λιγότερες ανταλλαγές στοιχείων. Άρα κάθε διεργασία θα έχει στοιχεία, αλλά οι πρώτες θα έχουν μια επιπλέον στήλη. Επίσης η πρώτη και η τελευταία διεργασία θα έχει μια λιγότερη στήλη. Άρα σύμφωνα με το παραπάνω ο χειρότερος χρόνος που μπορούμε να πετύχουμε είναι:

Για την επικοινωνία θα πάρουμε τη χειρότερη περίπτωση όπου μια διεργασία έχει κομμάτι του πίνακα που έχει δυο γείτονες. Από την ανταλλαγή των δυο στηλών που θα γίνει μεταξύ των δυο διεργασιών προκύπτει το εξής:

Συνεπώς:

* Execution time:
* Speedup:
* Efficiency:

### Διαχωρισμός σε blocks

Θεωρούμε ότι ο διαχωρισμός γίνεται σε P blocks και χωρίς βλάβη της γενικότητας θα θεωρήσουμε ότι ο P είναι τέλειο τετράγωνο και κάθε διάσταση μοιράζεται σε blocks. Σε αυτήν την περίπτωση σε κάθε μπλοκ θα έχουμε στοιχεία σε κάθε block. Οι πρώτες διεργασίες θα έχουν μια επιπλέον στήλη, ενώ οι πρώτες μια επιπλέον γραμμή. Επίσης όλες οι διεργασίες που θα περιλαμβάνουν ακραία στοιχεία θα έχουν μια λιγότερη γραμμή ή/και στήλη. Άρα σύμφωνα με το παραπάνω ο χειρότερος χρόνος που μπορούμε να πετύχουμε είναι:

Για την επικοινωνία θα πάρουμε τη χειρότερη περίπτωση όπου μια διεργασία έχει κομμάτι του πίνακα που έχει δυο γείτονες. Από την ανταλλαγή των δυο στηλών που θα γίνει μεταξύ των δυο διεργασιών προκύπτει το εξής:

Συνεπώς:

* Execution time:
* Speedup:
* Efficiency:

### Σύγκριση διαχωρισμών

Από μετρήσεις με το mpptest και από εκτέλεση του ανάλογου κώδικα σε ένα δοκιμαστικό πίνακα βρέθηκαν ότι , και .

Παρακάτω παρουσιάζονται τα διαγράμματα χρόνου, επιτάχυνσης και αποδοτικότητας των δυο επιλογών υλοποίησης μεταβάλλοντας το μέγεθος του πίνακα. Για την εξαγωγή τους χρησιμοποιήθηκε ένας πίνακας τετράγωνος (M=N) και P=100.

Σχήμα 1: Σύγκριση χρόνου εκτέλεσης με βάση το μέγεθος δεδομένων μεταξύ μπλοκ και λωρίδων

Σχήμα 2: Σύγκριση επιτάχυνσης με βάση το μέγεθος δεδομένων μεταξύ μπλοκ και λωρίδων

Σχήμα 3: Σύγκριση αποδοτικότητας με βάση το μέγεθος δεδομένων μεταξύ μπλοκ και λωρίδων

Ακολουθούν αντίστοιχα διαγράμματα στα οποία διατηρείται σταθερό το μέγεθος του πίνακα και μεταβάλλεται ο αριθμός των επεξεργαστών.

Σχήμα 4: Σύγκριση χρόνου εκτέλεσης με βάση το πλήθος των επεξεργαστών μεταξύ μπλοκ και λωρίδων

Σχήμα 5: Σύγκριση επιτάχυνσης με βάση το πλήθος των επεξεργαστών μεταξύ μπλοκ και λωρίδων

Σχήμα 6: Σύγκριση αποδοτικότητας με βάση το πλήθος των επεξεργαστών μεταξύ μπλοκ και λωρίδων

Όλα τα διαγράμματα συγκλίνουν στο συμπέρασμα ότι καλύτερη επιλογή είναι ο διαχωρισμός του προβλήματος σε blocks.

Mapping

Υπάρχουν δύο γενικές στρατηγικές στο mapping:

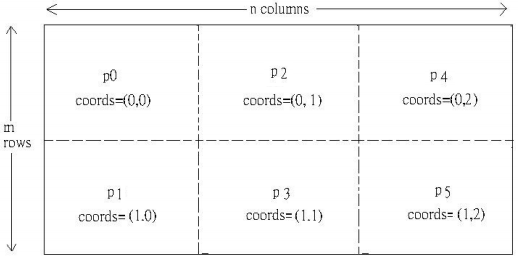
* Οι λειτουργίες που μπορούν να εκτελεστούν ταυτόχρονα τοποθετούνται σε διαφορετικούς επεξεργαστές, ώστε να μπορούμε να εκμεταλλευτούμε αυτή τη δυνατότητα
* Οι διεργασίες που επικοινωνούν συχνά μεταξύ τους, τοποθετούνται στον ίδιο κόμβο για να αποφεύγεται όσο είναι δυνατό η interprocessor επικοινωνία

Στην υλοποίησή μας κάθε διεργασία παίρνει το δικό της επεξεργαστή γιατί λαμβάνει ένα τμήμα του πίνακα και μπορεί να εκτελεστεί ταυτόχρονα από τις άλλες. Παράλληλα, στόχος μας είναι να φέρουμε όσο το δυνατόν πιο κοντά (ιδανικά στον ίδιο κόμβο) τις διεργασίες που επικοινωνούν, γνωρίζοντας ότι κάθε μία έχει τέσσερις σταθερούς γείτονες. Για να το πετύχουμε αυτό, θα εφαρμόσουμε καρτεσιανή τοπολογία.

# ΣΧΕΔΙΑΣΜΟΣ ΚΩΔΙΚΑ

Σχεδιασμός MPI προγράμματος

Σύμφωνα και με όσα αποδείχτηκαν στην ενότητα 2.3.3, ο ορθότερος ως προς την κλιμάκωση διαχωρισμός είναι σε blocks δύο διαστάσεων. Κάθε διεργασία θα αναλάβει έναν υποπίνακα. Για την επικοινωνία θα δημιουργήσουμε ένα δισδιάστατο καρτεσιανό πλέγμα διεργασιών στη μορφή που φαίνεται στην Εικόνα 3. Συγκεκριμένα, η πρώτη διεργασία θα πάρει το πάνω αριστερά κομμάτι και η ανάθεση θα συνεχιστεί πρώτα προς τα κάτω και μετα προς τα δεξιά. Η πρώτη βασική διαφοροποίηση με το δοσμένο πρόγραμμα είναι ότι θα χρησιμοποιήσουμε και τη master διεργασία για υπολογισμούς. Για αυτό η αρίθμιση στην Εικόνα 3 ξεκινάει απο το μηδέν. Κάθε διεργασία θα επικοινωνεί με τους τέσσερις γείτονες της (βοράς-νότος-ανατολή-δύση), όπως ορίζονται από το πλέγμα. Αν κάποιος γείτονας δεν υπάρχει (π.χ. ο p0 δεν έχει βορά και δύση στο παράδειγμά μας), θα τίθεται ίσος με MPI\_PROC\_NULL.

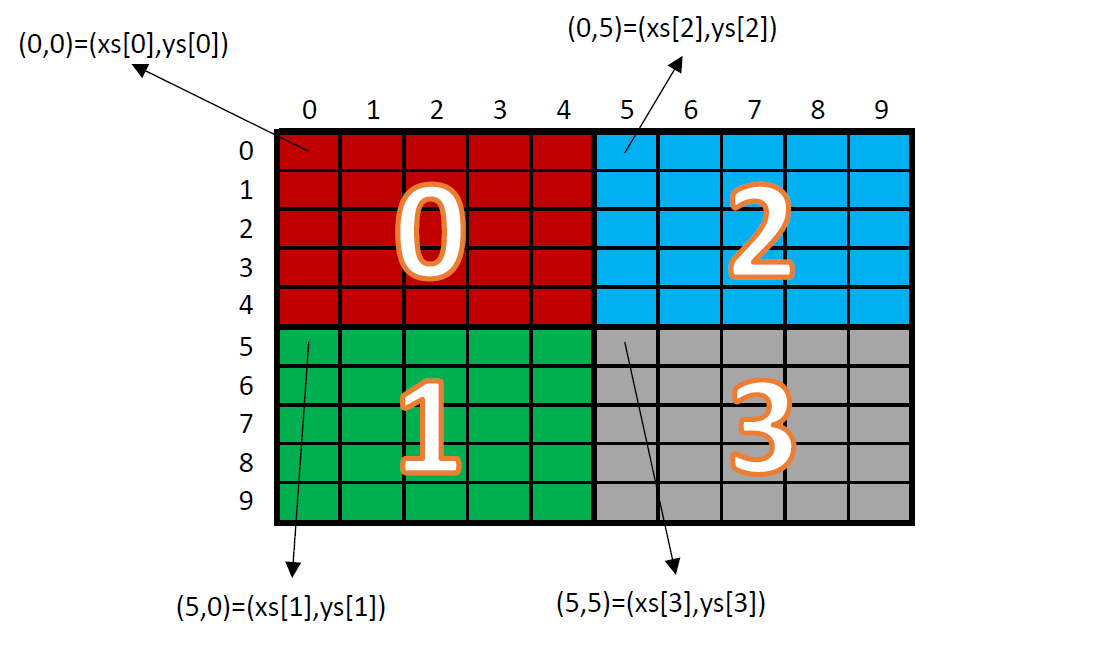


Εικόνα 3: Δισδιάστατο καρτεσιανό πλέγμα διεργασιών

Ο πίνακας πάνω στον οποίο δουλεύουμε είναι ο u, όπως και στο δοσμένο πρόγραμμα. Ο u είναι ένας τρισδιάστατος πίνακας. Ουσιαστικά είναι δύο δισδιάστατοι πίνακες ενωμένοι σε έναν τρισδιάστατο για να μπορούμε να ανταλλάσουμε τους δείκτες τους. Το μέγεθος καθενός από τους δύο δισδιάστατους πίνακες ισούται με το κομμάτι που του αντιστοιχεί απο τον ολικό πίνακα του προβλήματος που είναι ίσος με . Σε κάθε υποπίνακα προσθέτουμε 2 γραμμές και 2 στήλες για τα halo points. Συνεπώς, το μέγεθος του καθενός από τους δύο δισδιάστατους πίνακες του u ειναι .

Παρόλο που ο μεγάλος πίνακας , δεν εμφανίζεται πουθενά στο πρόγραμμα, θα πρέπει να προσδιορίσουμε τη θέση κάθε υποπίνακα που χειρίζεται κάθε διεργασία στον ολικό πίνακα, ώστε αφενός να μπορούμε να αρχικοποιήσουμε τους υποπίνακες και αφετέρου να μπορέσουμε να παράξουμε τα αρχεία initial.dat και final.dat. Για αυτόν τον προσδιορισμό θα υπολογίσουμε και θα μοιράσουμε σε όλες τις διεργασίες τους πίνακες xs και ys, στη θέση i των οποίων βρίσκεται η συνεταγμένη x και y αντίστοιχα του πάνω αριστερού σημείου της i-οστής διεργασίας, όπως αυτή καθορίζεται από τη συνάρτηση MPI\_Comm\_rank.

Η Εικόνα 4 ξεδιαλύνει όσα αναλυτικά εξηγήθηκαν στις προηγούμενες δύο παραγράφους. Ο πίνακας ειναι και χωρίζεται σε 4 blocks, 2 σε κάθε διάσταση, καθένα εκ των οποίων έχει διάσταση . Έτσι, κάθε πίνακας u που αναλαμβάνει μια διεργασία είναι μεγέθους (.



Εικόνα 4: Πίνακας u και διαχωρισμός του σε blocks

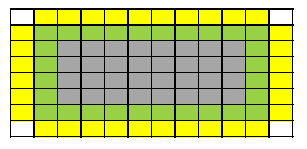
Για την ανταλλαγή γραμμών και στηλών μεταξύ των διεργασιών δημιουργούμε τους δύο MPI τύπους row και column, ώστε να στέλνουμε γραμμές και στήλες αντίστοιχα. Για τον τύπο row χρησιμοποιούμε τη συνάρτηση MPI\_Type\_contiguous που επιτρέπει την αντιγραφή συνεχόμενων δεδομένων ενός τύπου. Για τον τύπο column χρησιμοποιούμε τη συνάρτηση MPI\_Type\_vector που είναι ένας πιο γενικός datatype constructor που επιτρέπει την αντιγραφή blocks δεδομένων που ισαπέχουν μεταξύ τους. Τα στοιχεία μιας στήλης ενός υποπίνακα απέχουν μεταξύ τους κατά το μέγεθος μίας γραμμής του πίνακα u που είναι , όσες δηλαδή και οι στήλες του. Το μέγεθος του block δεδομένων είναι 1, αφού χρειαζόμαστε ένα στοιχείο από κάθε γραμμή, ενώ χρειαζόμαστε τέτοια blocks, όσες δηλαδή και οι γραμμές του κάθε υποπίνακα.

Κάθε διεργασία αρχικοποιεί το δικό της πίνακα σαν να είναι τμήμα του μεγαλύτερου πίνακα σύμφωνα με τον κώδικα της initdat από το αρχικό πρόγραμμα. Όπως προαναφέρθηκε, η κάθε διεργασία μπορεί να προσδιορίζει τα όρια της μέσω των πινάκων xs και ys.

Η κεντρική επανάληψη που χρονομετρούμε και συγκρίνουμε στα πλαίσια της άσκησης βασίζεται στη λογική του ψευδοκώδικα που δόθηκε στις οδηγίες σχεδιασμού, ο οποίος ακολουθεί:

MPI\_Barrier (συγχρονισμός διεργασιών πριν τις μετρήσεις)  
Start MPI\_Wtime  
For #επαναλήψεων  
 Isend (SRequest) X 4 (Β,Ν,Δ,Α)  
 Irecv (RRequest) X 4 (Β,Ν,Δ,Α)  
 Υπολογισμός εσωτερικών στοιχείων «μετά» (γκρι στοιχεία στο σχήμα)  
 Wait (RRequest) Χ 4 (Β,Ν,Δ,Α) ή WaitAll (array of RRequests)  
 Υπολογισμός εξωτερικών στοιχείων «μετά> (πράσινα στο σχήμα)  
 Πριν Πίνακας = Μετά Πίνακας  
 (reduce για σύγκλιση εδώ)  
 Wait(SRequest) X 4 (Β,Ν,Δ,Α) ή WaitAll (array of SRequests)  
End for  
End MPI\_Wtime

Στην Εικόνα 5 φαίνονται τα χρωματιστά σημεία που αναφέρονται στον ψευδοκώδικα και αντιστοιχούν σε έναν υποπίνακα που αναλαμβάνει μια διεργασία. Τα κίτρινα σημεία είναι τα halo points όπου αποθηκεύονται οι γραμμές και οι στήλες που στέλνονται από τους γείτονες. Τα πράσινα σημεία είναι τα εξωτερικά στοιχεία του υποπίνακα που έχουν κάποιον γείτονα που ανήκει σε άλλη διεργασία, οπότε ο υπολογισμός της νέας τους τιμής πρέπει να γίνει μετά τη λήψη στοιχείων από τους γείτονες. Τα γκρι σημεία είναι τα εσωτερικά στοιχεία του υποπίνακα που όλοι τους οι γείτονες ανήκουν στην ίδια διεργασία και ο υπολογισμός της νέας τους τιμής είναι ανεξάρτητες από ανταλλαγές με γείτονες. Τα λευκά σημεία δεν παίζουν κανέναν ρόλο γιατί δεν είναι γείτονες κανενός σημείου που ανήκει στον υποπίνακα.



Εικόνα 5: Εσωτερικά, εξωτερικά και αλώ σημεία υποπίνακα

Εν γένει, η επικοινωνία είναι σχετικά αργή εξαιτίας του δικτύου επικοινωνίας και της εξάρτησης από τη διεργασία-παραλήπτη. Με την κλασική blocking επικοινωνία (MPI\_Send, MPI\_Recv), η διεργασία πρέπει να περιμένει αδρανής όσο διαρκεί η επικοινωνία. Με την non-blocking επικοινωνία, το MPI παρέχει τη δυνατότητα επικάλυψης της επικοινωνίας με πράξεις, αν και απαιτείται ο επιπρόσθετος έλεγχος για το αν έχει ολοκληρωθεί η επικοινωνία. Κάθε φορά που διενεργείται μια επικοινωνία δημιουργείται ένα αντικείμενο επικοινωνίας που εκτός από το ίδιο το μήνυμα περιέχει και άλλες πληροφορίες, όπως ο παραλήπτης, ο τύπος του μηνύματος, ο παραλήπτης, το tag κλπ. Όταν μια διεργασία έχει σταθερούς γείτονες, τα πεδία αυτά είναι τα ίδια και το μόνο που αλλάζει είναι το μήνυμα καθαυτό, δεν είναι αναγκαίο να δημιουργούμε συνέχεια το ίδιο αντικείμενο με τα ίδια πεδία εκτός από το μήνυμα, αλλά να δημιουργούμε ένα ανά γείτονα και να το επαναχρησιμοποιούμε όσες φορές χρειαστεί. Αυτό το πλεονέκτημα μας το δίνει η persistent επικοινωνία, η λογική της είναι σαν να δημιουργούμε ένα μόνιμο μικρό «μισό» κανάλι επικοινωνίας στο οποίο μπορούμε να στέλνουμε πάντα τον ίδιο τύπο μηνύματος στο ίδιο παραλήπτη. Το κανάλι αυτό χαρακτηρίστηκε ως μισό, διότι ούτε η διεγασία δε λαμβάνει από το γείτονα της στο ίδιο κανάλι ούτε είναι απαραίτητο ότι ο γείτονας της θα χρησιμοποιεί persistent επικοινωνία.

Η πρώτη σημαντική διαφοροποίηση σε σχέση με τον ψευδοκώδικα είναι ότι αντί για non-blocking επικοινωνία (MPI\_Isend, MPI\_Irecv) χρησιμοποιούμε persistent επικοινωνία (MPI\_Send\_init, MPI\_Recv\_init, MPI\_Startall). Στον κώδικα δημιουργούμε ένα κανάλι επικοινωνίας για send και ένα για receive για κάθε έναν από τους τέσσερις γείτονες και με τις εντολές MPI\_Startall και MPI\_Waitall χειριζόμαστε πλέον την persistent επικοινωνία. Επειδή, όμως, στην κεντρική επανάληψη χειριζόμαστε εναλλάξ τις u[0] και u[1], χρειαζόμαστε ακριβώς τα διπλάσια κανάλια.

Όσον αφορά το κομμάτι της σύγκλισης, η σχεδιαστική επιλογή ήταν να γίνει ο υπολογισμός των τετραγωνικών διαφορών μέσα σε διπλό for γιατί πειραματικά αποδείχτηκε καλύτερη λύση από τον υπολογισμό στα ήδη υπάρχοντα for. Για τον υπολογισμό της ολικής τετραγωνικής διαφοράς γίνεται χρήση της MPI\_Allreduce γιατί όλες οι διεργασίες χρειάζονται το αποτέλεσμά της.

O χρόνος εκτέλεσης του προγράμματος δίνεται από το μέγιστο χρόνο που θα κάνει μια διεργασία να ολοκληρώσει τους υπολογισμούς της. Η ίδια λογική έχει χρησιμοποιηθεί και στη μέτρηση του αρχικού προγράμματος, ώστε η σύγκριση που θα γίνει να είναι ισότιμη.

Τέλος, έχει χρησιμοποιηθεί parallel I/O για την εκτύπωση των αρχείων initial.dat και final.dat, που περιέχουν τα δεδομένα στην αρχή και μετά το πέρας του προγράμματος. Το αρχείο που παράγεται γράφοντας ταυτόχρονα όλες οι διεργασίες είναι σε binary μοφή, οπότε ο master το ξαναδιαβάζει και το γράφει τους float αριθμούς που καταλαβαίνουμε.

Σχεδιασμός υβριδικού προγράμματος

Για την υλοποίηση του υβριδικού κώδικα MPI και OpenMP χρησιμοποιήθηκε ως βάση ο ίδιος κώδικας με το MPI. Κάθε MPI διεργασία για τον υπολογισμό του δικού της μέρους του πίνακα, θα χρησιμοποιεί OpenMP. Η προσέγγιση που χρησιμοποιούμε είναι η παραλληλοποίηση στα δύο ακολουθιακά τμήματα στην κεντρική επανάληψη, αυτό του υπολογισμού των νέων εσωτερικών σημείων και αυτό του υπολογισμού των νέων εξωτερικών σημείων.

Στην περίπτωση του ελέγχου σύγκλισης, χρησιμοποιούμε openmp-reduce μεταξύ των νημάτων για τον υπολογισμό τον τοπικών τετραγωνικών διαφορών και στη συνέχεια MPI\_Allreduce για να υπολογίσουμε την ολική τετραγωνική διαφορά.

Για να γλιτώσουμε τη δημιουργία και καταστροφή των threads σε κάθε επανάληψη, επιλέξαμε να τα δημιουργήσουμε μια φορά έξω από την επανάληψη και να τρέχουμε παράλληλα τα for με τα ήδη δημιουργημένα threads. Στα διπλά for χρησιμοποιούμε το collapse(2), που συγχωνεύει τις δύο επαναλήψεις σε μία μεγάλη και τη διαχωρίζει ανάλογα με το schedule που έχει δηλωθεί.

Σημαντικό είναι να αναφερθεί ότι ενώ οι επεξεργαστές έχουν κοινή μνήμη δεν έχουν κοινή cache. Αυτό σημαίνει ότι πρέπει να υπάρξει κατάλληλος χειρισμός ώστε κάθε μπλοκ μνήμης που μεταβαίνει στην cache και εγγράφεται, δεν πρέπει να βρίσκεται σε άλλο επεξεργαστή καθώς θα υπάρχει η ανάγκη ενημέρωσης όλων των cache με κάθε αλλαγή το οποίο προκαλεί καθυστέρηση στην εκτέλεση. Άρα ενώ για τον πίνακα από τον οποίο διαβάζονται τα δεδομένα δεν υπάρχει πρόβλημα, για τον δεύτερο όπου εγγράφονται τα αποτελέσματα μπορεί να υπάρξει πρόβλημα. Οπότε το τρέχον στοιχείο του πίνακα στο οποίο γίνεται επεξεργασία από έναν επεξεργαστή, πρέπει να μην βρίσκεται σε μπλοκ μνήμης όπου εκεί βρίσκεται το τρέχον στοιχείο ενός άλλου πίνακα. Άρα αρκεί ο διαμοιρασμός να έχει ελάχιστη απόσταση ενός block μνήμης.

# ΜΕΤΡΗΣΕΙΣ

Οι μετρήσεις αφορούν 1000 επαναλήψεις του προγράμματος. Στις περιπτώσεις που τρέχουμε με έλεγχο σύγκλισης, αυτός γίνεται κάθε 20 επαναλήψεις. Τα μεγέθη των δεδομένων και ο συνδυασμός υπολογιστικών κόμβων και των διεργασιών είναι προκαθορισμένα από την εκφώνηση της άσκησης. Συγκεκριμένα, όσον αφορά τα μεγέθη των δεδομένων, ξεκινάμε από 80x64 και κάθε φορά διπλασιάζουμε και τις δύο πλευρές σε κάθε επόμενο πείραμα, σταματώντας εκεί που μας επιτρέπει το marie.hellasgrid.gr. Το ανώτατο μέγεθος δεδομένων που μπορέσαμε να τρέξουμε επιτυχώς το πρόγραμμα ήταν 2560x2048. Όσον αφορά το συνδυασμό υπολογιστικών κόμβων και διεργασιών, χρησιμοποιήσαμε αυτούς που δίνονται στην εκφώνηση και διαιρούν ακριβώς τις πλευρές. Ούτως ή άλλως ήταν ήδη πολλά σε πλήθος τα πειράματα για να χρησιμοποιήσουμε και τα μεγέθη που δε διαιρούν ακριβώς την πλευρά. Όλοι οι χρόνοι μετρώνται σε δευτερόλεπτα και οι μετρήσεις έχουν γίνει στο marie.hellasgrid.gr.

MPI πρόγραμμα χωρίς έλεγχο σύγκλισης

Στους Πίνακες 1, 2 και 3 ακολουθούν οι μετρήσεις για χρόνο, επιτάχυνση και αποδοτικότητα του MPI προγράμματος που σχεδιάστηκε στα πλαίσια της άσκησης χωρίς έλεγχο σύγκλισης.

Πίνακας 1: Μετρήσεις χρόνου MPI προγράμματος χωρίς έλεγχο σύγκλισης

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nodes / Tasks (sec) | | | | | | |
|  | **1/1** | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 2.53E-02 | 9.30E-03 | 1.24E-01 | 1.19E-01 | 2.82E-01 | 2.12E-01 |
| 160x128 | 9.87E-02 | 2.91E-02 | 2.19E-01 | 2.18E-01 | 2.04E-01 | 2.63E-01 |
| 320x256 | 7.52E-01 | 1.04E-01 | 2.33E-01 | 3.33E-01 | 3.14E-01 | 2.18E-01 |
| 640x512 | 3.01E+00 | 7.63E-01 | 2.57E-01 | 3.26E-01 | 2.52E-01 | 2.13E-01 |
| 1280x1024 | 1.27E+01 | 3.02E+00 | 9.05E-01 | 5.80E-01 | 4.12E-01 | 2.52E-01 |
| 2560x2048 | 5.09E+01 | 1.20E+01 | 3.19E+00 | 9.65E-01 | 1.54E+00 | 5.18E-01 |

Πίνακας 2: Επιτάχυνση MPI προγράμματος χωρίς έλεγχο σύγκλισης

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nodes / Tasks | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 2.72 | 0.20 | 0.21 | 0.09 | 0.12 |
| 160x128 | 3.39 | 0.45 | 0.45 | 0.48 | 0.38 |
| 320x256 | 7.23 | 3.23 | 2.26 | 2.39 | 3.45 |
| 640x512 | 3.94 | 11.71 | 9.23 | 11.94 | 14.13 |
| 1280x1024 | 4.21 | 14.03 | 21.90 | 30.83 | 50.40 |
| 2560x2048 | 4.24 | 15.96 | 52.75 | 33.05 | 98.26 |

Πίνακας 3: Αποδοτικότητα MPI προγράμματος χωρίς έλεγχο σύγκλισης

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nodes / Tasks | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 0.680 | 0.013 | 0.003 | 0.001 | 0.001 |
| 160x128 | 0.848 | 0.028 | 0.007 | 0.004 | 0.002 |
| 320x256 | 1.808 | 0.202 | 0.035 | 0.019 | 0.022 |
| 640x512 | 0.986 | 0.732 | 0.144 | 0.093 | 0.088 |
| 1280x1024 | 1.051 | 0.877 | 0.342 | 0.241 | 0.315 |
| 2560x2048 | 1.060 | 0.997 | 0.824 | 0.258 | 0.614 |

MPI πρόγραμμα με έλεγχο σύγκλισης κάθε ν επαναλήψεις

Στους Πίνακες 4, 5 και 6 ακολουθούν οι μετρήσεις για χρόνο, επιτάχυνση και αποδοτικότητα του MPI προγράμματος που σχεδιάστηκε στα πλαίσια της άσκησης με έλεγχο σύγκλισης κάθε 20 επαναλήψεις.

Πίνακας 4: Μετρήσεις χρόνου MPI προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nodes / Tasks (sec) | | | | | | |
|  | **1/1** | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 3.33E-02 | 1.30E-02 | 4.26E-01 | 2.54E-01 | 3.79E-01 | 2.06E-01 |
| 160x128 | 1.24E-01 | 3.97E-02 | 2.92E-01 | 6.44E-01 | 2.31E-01 | 2.49E-01 |
| 320x256 | 8.51E-01 | 1.33E-01 | 3.13E-01 | 5.65E-01 | 1.76E+00 | 2.29E-01 |
| 640x512 | 3.39E+00 | 8.66E-01 | 7.29E-01 | 6.72E-01 | 8.68E-01 | 2.42E-01 |
| 1280x1024 | 1.58E+01 | 3.43E+00 | 1.24E+00 | 6.61E-01 | 7.30E-01 | 2.63E-01 |
| 2560x2048 | 6.29E+01 | 1.95E+01 | 4.54E+00 | 1.53E+00 | 1.12E+00 | 4.80E-01 |

Πίνακας 5: Επιτάχυνση MPI προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nodes / Tasks | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 2.56 | 0.08 | 0.13 | 0.09 | 0.16 |
| 160x128 | 3.12 | 0.42 | 0.19 | 0.54 | 0.50 |
| 320x256 | 6.40 | 2.72 | 1.51 | 0.48 | 3.72 |
| 640x512 | 3.91 | 4.65 | 5.04 | 3.91 | 14.01 |
| 1280x1024 | 4.61 | 12.74 | 23.90 | 21.64 | 60.08 |
| 2560x2048 | 3.23 | 13.85 | 41.11 | 56.16 | 131.04 |

Πίνακας 6: Αποδοτικότητα MPI προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nodes / Tasks | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 0.640 | 0.005 | 0.002 | 0.001 | 0.001 |
| 160x128 | 0.781 | 0.027 | 0.003 | 0.004 | 0.003 |
| 320x256 | 1.600 | 0.170 | 0.024 | 0.004 | 0.023 |
| 640x512 | 0.979 | 0.291 | 0.079 | 0.031 | 0.088 |
| 1280x1024 | 1.152 | 0.796 | 0.373 | 0.169 | 0.375 |
| 2560x2048 | 0.806 | 0.866 | 0.642 | 0.439 | 0.819 |

Υβριδικό MPI με OpenMP

Στους Πίνακες 7, 8 και 9 ακολουθούν οι μετρήσεις για χρόνο, επιτάχυνση και αποδοτικότητα του υβριδικού MPI με OpenMP προγράμματος που σχεδιάστηκε στα πλαίσια της άσκησης με έλεγχο σύγκλισης κάθε 20 επαναλήψεις.

Πίνακας 7: Μετρήσεις χρόνου υβριδικού προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nodes / Tasks / Threads (sec) | | | | | | | | | | | | | | | |
|  | **1/2/2** | **1/1/4** | **2/8/2** | **2/8/4** | **2/4/4** | **2/2/8** | **8/32/2** | **8/16/4** | **8/8/8** | **16/64/2** | **16/32/4** | **16/16/8** | **20/80/2** | **20/40/4** | **20/20/8** |
| 80x64 | 0.05 | 0.17 | 0.09 | 3.48 | 0.09 | 0.21 | 0.27 | 0.13 | 0.17 | 0.2 | 0.14 | 0.19 | 0.16 | 0.13 | 0.2 |
| 160x128 | 0.15 | 0.57 | 0.12 | 3.29 | 0.1 | 0.61 | 0.13 | 0.12 | 0.24 | 0.71 | 0.13 | 0.25 | 0.18 | 0.14 | 0.26 |
| 320x256 | 1.16 | 4.6 | 0.15 | 3.5 | 0.22 | 4.63 | 0.15 | 0.13 | 0.64 | 0.16 | 0.12 | 0.46 | 0.32 | 0.15 | 0.46 |
| 640x512 | 4.58 | 18.3 | 0.41 | 4.44 | 1.33 | 18.3 | 0.2 | 0.25 | 4.66 | 0.18 | 0.19 | 1.43 | 0.21 | 0.17 | 1.14 |
| 1280x1024 | 18.6 | 73.5 | 2.63 | 8.16 | 5.09 | 73.8 | 0.72 | 1.38 | 18.4 | 0.27 | 0.73 | 9.3 | 0.32 | 0.61 | 4.67 |
| 2560x2048 | 74.2 | 294 | 9.77 | 22.8 | 19.97 | 294.5 | 2.61 | 5.19 | - | 1.34 | 2.72 | - | 1.1 | 2.11 | 17.2 |

Πίνακας 8: Επιτάχυνση υβριδικού προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nodes / Tasks / Threads | | | | | | | | | | | | | | | |
|  | **1/2/2** | **1/1/4** | **2/8/2** | **2/8/4** | **2/4/4** | **2/2/8** | **8/32/2** | **8/16/4** | **8/8/8** | **16/64/2** | **16/32/4** | **16/16/8** | **20/80/2** | **20/40/4** | **20/20/8** |
| 80x64 | 0.62 | 0.19 | 0.35 | 0.01 | 0.35 | 0.16 | 0.12 | 0.26 | 0.20 | 0.17 | 0.25 | 0.18 | 0.21 | 0.25 | 0.16 |
| 160x128 | 0.81 | 0.22 | 1.03 | 0.04 | 1.22 | 0.20 | 0.98 | 1.02 | 0.51 | 0.18 | 0.93 | 0.51 | 0.71 | 0.91 | 0.48 |
| 320x256 | 0.73 | 0.19 | 5.75 | 0.24 | 3.94 | 0.18 | 5.83 | 6.45 | 1.33 | 5.39 | 6.75 | 1.86 | 2.66 | 5.87 | 1.87 |
| 640x512 | 0.74 | 0.19 | 8.25 | 0.76 | 2.55 | 0.19 | 17.21 | 13.84 | 0.73 | 19.26 | 18.23 | 2.37 | 16.54 | 20.18 | 2.97 |
| 1280x1024 | 0.85 | 0.21 | 6.01 | 1.94 | 3.10 | 0.21 | 21.94 | 11.45 | 0.86 | 58.30 | 21.56 | 1.70 | 50.16 | 26.07 | 3.38 |
| 2560x2048 | 0.85 | 0.21 | 6.44 | 2.76 | 3.15 | 0.21 | 24.10 | 12.12 | - | 46.94 | 23.13 | - | 57.18 | 29.81 | 3.65 |

Πίνακας 9: Αποδοτικότητα υβριδικού προγράμματος με έλεγχο σύγκλισης κάθε ν=20 επαναλήψεις

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nodes / Tasks / Threads | | | | | | | | | | | | | | | |
|  | **1/2/2** | **1/1/4** | **2/8/2** | **2/8/4** | **2/4/4** | **2/2/8** | **8/32/2** | **8/16/4** | **8/8/8** | **16/64/2** | **16/32/4** | **16/16/8** | **20/80/2** | **20/40/4** | **20/20/8** |
| 80x64 | 0.16 | 0.05 | 0.02 | 0.0003 | 0.02 | 0.01 | 0.002 | 0.004 | 0.003 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 |
| 160x128 | 0.20 | 0.06 | 0.07 | 0.001 | 0.08 | 0.01 | 0.02 | 0.02 | 0.008 | 0.001 | 0.007 | 0.004 | 0.004 | 0.006 | 0.003 |
| 320x256 | 0.18 | 0.05 | 0.36 | 0.008 | 0.25 | 0.01 | 0.09 | 0.10 | 0.021 | 0.042 | 0.053 | 0.015 | 0.02 | 0.04 | 0.012 |
| 640x512 | 0.19 | 0.05 | 0.52 | 0.02 | 0.16 | 0.01 | 0.27 | 0.22 | 0.011 | 0.150 | 0.142 | 0.019 | 0.10 | 0.13 | 0.019 |
| 1280x1024 | 0.21 | 0.05 | 0.38 | 0.06 | 0.19 | 0.01 | 0.34 | 0.18 | 0.013 | 0.455 | 0.168 | 0.013 | 0.31 | 0.16 | 0.021 |
| 2560x2048 | 0.22 | 0.05 | 0.40 | 0.09 | 0.20 | 0.01 | 0.38 | 0.19 | - | 0.367 | 0.181 | - | 0.36 | 0.19 | 0.023 |

# ΣΥΓΚΡΙΣΗ ΜΕ ΑΡΧΙΚΟ ΠΡΟΓΡΑΜΜΑ

Στην ενότητα αυτή θα συγκρίνουμε την MPI υλοποίησή μας με εκείνη που μας δόθηκε σε σχέση με το χρόνο, την επιτάχυνση και την αποδοτικότητα τους.

Σύγκριση πραγματικών υπολογισμών

Στον Πίνακα 10 βρίσκονται οι χρόνοι που μετρήθηκαν εκτελώντας το δοσμένο πρόγραμμα. Η σύγκριση γίνεται με βάση τα περιεχόμενα του Πίνακα 1, ο οποίος έχει το χρόνους του δικού μας MPI προγράμματος χωρίς έλεγχο σύγκλισης.

Πίνακας 10: Μετρήσεις χρόνου αρχικού προγράμματος προγράμματος χωρίς έλεγχο σύγκλισης

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Processes / Threads (sec) | | | | | | |
|  | **1/1** | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 0.07 | 0.02 | 0.15 | 0.94 | 1.25 | 1.49 |
| 160x128 | 0.29 | 0.09 | 0.25 | 0.62 | 1.07 | 1.36 |
| 320x256 | 1.16 | 0.30 | 0.23 | 0.54 | 1.05 | 1.24 |
| 640x512 | 4.70 | 1.18 | 0.62 | 0.74 | 1.03 | 1.30 |

Στον Πίνακα 11 ακολουθεί η επιτάχυνση του νέου προγράμματος έναντι του αρχικού. Υπολογίστηκε με τον τύπο , όπου : ο χρόνος εκτέλεσης του αρχικού προγράμματος που δόθηκε και : ο χρόνος εκτέλεσης του δικού μας προγράμματος.

Πίνακας 11: Επιτάχυνση υλοποίησή μας σε σχέση με το αρχικό πρόγραμμα

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Processes / Threads | | | | | | |
|  | **1/1** | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 2.76 | 2.26 | 1.22 | 7.89 | 4.43 | 7.03 |
| 160x128 | 2.93 | 3.00 | 1.15 | 2.84 | 5.25 | 5.17 |
| 320x256 | 1.54 | 2.89 | 1.00 | 1.63 | 3.34 | 5.69 |
| 640x512 | 1.56 | 1.55 | 2.41 | 2.26 | 4.09 | 6.10 |

Παρατηρούμε σημαντική επιτάχυνση του δικού μας προγράμματος έναντι του αρχικού σε κάθε περίπτωση. Αξιοσημείωτο είναι ότι η επιτάχυνση κλιμακώνει στις πολλές διεργασίες που είναι ένα από τους στόχους της υλοποίησής μας.

Σύγκριση αναλυτικών μετρήσεων

Στους Πίνακες 12, 13 και 14 ακολουθούν οι θεωρητικές μετρήσεις για το χρόνο, την επιτάχυνση και την αποδοτικότητα για τη λογική του δοσμένου προγράμματος που είναι διαχωρισμός σε λωρίδες. Βλέπουμε ότι οι μετρήσεις είναι σημαντικά πιο γρήγορες σε σχέση με τις πραγματικές, γεγονός που δείχνει ότι οι υπολογισμοί δεν έχουν γίνει κατά το βέλτιστο τρόπο.

Πίνακας 12: Θεωρητική μέτρηση χρόνου για διαχωρισμό σε λωρίδες

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Processes / Threads (sec) | | | | | | |
|  | **1/1** | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 4.17E-04 | 3.21E-04 | 2.97E-04 | 2.91E-04 | 2.89E-04 | 2.89E-04 |
| 160x128 | 1.09E-03 | 7.05E-04 | 6.09E-04 | 5.85E-04 | 5.81E-04 | 5.77E-04 |
| 320x256 | 3.20E-03 | 1.67E-03 | 1.28E-03 | 1.19E-03 | 1.17E-03 | 1.16E-03 |
| 640x512 | 1.05E-02 | 4.35E-03 | 2.82E-03 | 2.43E-03 | 2.37E-03 | 2.35E-03 |
| 1280x1024 | 3.74E-02 | 1.28E-02 | 6.66E-03 | 5.12E-03 | 4.87E-03 | 4.80E-03 |
| 2560x2048 | 1.40E-01 | 4.20E-02 | 1.74E-02 | 1.13E-02 | 1.02E-02 | 9.99E-03 |

Πίνακας 13: Θεωρητική μέτρηση επιτάχυνσης για διαχωρισμό σε λωρίδες

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processes / Threads | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 1.30 | 1.40 | 1.43 | 1.44 | 1.44 |
| 160x128 | 1.54 | 1.79 | 1.86 | 1.87 | 1.89 |
| 320x256 | 1.92 | 2.50 | 2.70 | 2.74 | 2.76 |
| 640x512 | 2.41 | 3.73 | 4.31 | 4.43 | 4.46 |
| 1280x1024 | 2.92 | 5.61 | 7.30 | 7.68 | 7.78 |
| 2560x2048 | 3.34 | 8.06 | 12.45 | 13.70 | 14.05 |

Πίνακας 14: Θεωρητική μέτρηση αποδοτικότητας για διαχωρισμό σε λωρίδες

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processes / Threads | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 0.325 | 0.088 | 0.022 | 0.011 | 0.009 |
| 160x128 | 0.386 | 0.112 | 0.029 | 0.015 | 0.012 |
| 320x256 | 0.481 | 0.156 | 0.042 | 0.021 | 0.017 |
| 640x512 | 0.603 | 0.233 | 0.067 | 0.035 | 0.028 |
| 1280x1024 | 0.730 | 0.351 | 0.114 | 0.060 | 0.049 |
| 2560x2048 | 0.835 | 0.504 | 0.195 | 0.107 | 0.088 |

Στους Πίνακες 15, 16 και 17 ακολουθούν οι θεωρητικές μετρήσεις για το χρόνο, την επιτάχυνση και την αποδοτικότητα για τη λογική του δικού μας προγράμματος που είναι διαχωρισμός σε μπλοκ. Σε αυτήν την περίπτωση παρατηρούμε πιο λογικές διαφορές. Είναι αναμενόμενο να έχουμε διαφορές μεταξύ ενός θεωρητικού μοντέλου και της πρακτικής του υλοποίησης.

Πίνακας 15: Θεωρητική μέτρηση χρόνου για διαχωρισμό σε μπλοκ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Processes / Threads (sec) | | | | | | |
|  | **1/1** | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 1.91E-02 | 4.91E-03 | 1.30E-03 | 3.61E-04 | 1.76E-04 | 1.55E-04 |
| 160x128 | 7.53E-02 | 1.91E-02 | 4.91E-03 | 1.30E-03 | 6.55E-04 | 5.21E-04 |
| 320x256 | 2.99E-01 | 7.53E-02 | 1.91E-02 | 4.91E-03 | 2.42E-03 | 1.98E-03 |
| 640x512 | 1.19E+00 | 2.99E-01 | 7.53E-02 | 1.91E-02 | 9.51E-03 | 7.58E-03 |
| 1280x1024 | 4.76E+00 | 1.19E+00 | 2.99E-01 | 7.53E-02 | 3.76E-02 | 3.00E-02 |
| 2560x2048 | 1.90E+01 | 4.76E+00 | 1.19E+00 | 2.99E-01 | 1.50E-01 | 1.19E-01 |

Πίνακας 16: Θεωρητική μέτρηση επιτάχυνσης για διαχωρισμό σε μπλοκ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processes / Threads | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 3.89 | 14.72 | 52.84 | 108.21 | 123.43 |
| 160x128 | 3.94 | 15.34 | 58.07 | 114.96 | 144.57 |
| 320x256 | 3.97 | 15.67 | 60.94 | 123.56 | 150.91 |
| 640x512 | 3.99 | 15.83 | 62.45 | 125.38 | 157.21 |
| 1280x1024 | 3.99 | 15.92 | 63.22 | 126.56 | 158.91 |
| 2560x2048 | 4.00 | 15.96 | 63.61 | 127.01 | 159.56 |

Πίνακας 17: Θεωρητική μέτρηση αποδοτικότητας για διαχωρισμό σε μπλοκ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processes / Threads | | | | | |
|  | **1/4** | **2/16** | **8/64** | **16/128** | **20/160** |
| 80x64 | 0.972 | 0.920 | 0.826 | 0.845 | 0.771 |
| 160x128 | 0.986 | 0.959 | 0.907 | 0.898 | 0.904 |
| 320x256 | 0.993 | 0.979 | 0.952 | 0.965 | 0.943 |
| 640x512 | 0.996 | 0.990 | 0.976 | 0.980 | 0.983 |
| 1280x1024 | 0.998 | 0.995 | 0.988 | 0.989 | 0.993 |
| 2560x2048 | 0.999 | 0.997 | 0.994 | 0.992 | 0.997 |

Η σύγκριση των θεωρητικών μετρήσεων δείχνει ότι σε αυτά τα μεγέθη δεδομένων ίσως είναι λίγο ταχύτερη από άποψη χρόνου η προσέγγιση σε λωρίδες, αλλά η προσέγγιση σε μπλοκ κλιμακώνει εμφατικά καλύτερα, πράγμα που φαίνεται και από τους πίνακες για επιτάχυνση και αποδοτικότητα. Τα μεγέθη είναι μικρά γιατί μας περιορίζει το σύστημα στο οποίο μετρήσαμε. Είναι βέβαιο ότι αν μπορούσαμε να δοκιμάσουμε μεγαλύτερα μεγέθη δεδομένων, η εικόνα του πίνακα χρόνων θα ήταν αντίστοιχη.

# ΣΥΜΠΕΡΑΣΜΑΤΑ – ΠΑΡΑΤΗΡΗΣΕΙΣ – ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

Το γενικό συμπέρασμα είναι πως η ενσωμάτωση μεθόδων παραλληλίας στις σημερινές εφαρμογές κρίνεται αναγκαία, δεδομένης της συνεχούς αύξησης του μεγέθους των δεδομένων προς επεξεργασία, και της πολυπλοκότητας των λειτουργιών που χρειάζεται να πραγματοποιηθούν πάνω σε αυτά.

Παρατηρούμε ότι οι διαφορές μεταξύ των χρόνων εκτέλεσης των δυο υλοποιήσεων (MPI και υβριδικό MPI με OpenMP) είναι πολύ μικρές. Άλλες φορές εμφανίζεται να βρίσκεται μπροστά για λίγο η MPI υλοποίηση και άλλες η υβριδική. Άρα αυτό που μπορούμε να συμπεράνουμε είναι ότι η επιπρόσθετη προσπάθεια για την εισαγωγή του OpenMP κώδικα δεν προσέφερε στην ουσία κάτι.

Θα είχε ενδιαφέρον να υπήρχαν να μπορούσαμε να τεστάρουμε το πρόγραμμά μας με μεγαλύτερα δεδομένα. Τα 2Gb ως όριο μνήμης δεν είναι επαρκές, αν υπολογίσουμε ότι χρειαζόμαστε επιπρόσθετες γραμμές και στήλες ως halo points, πίνακα που να κρατάει τα όρια του κάθε block, δεσμεύσεις τύπων row και column για επικοινωνία μεταξυ των blocks, δεσμεύσεις τύπου subarray για parallel I/O, πίνακες για να κρατάμε persistent communication κ.α., γεγονός που μας περιορίζει το μέγεθος των δεδομένων που μπορούμε να δεσμεύσουμε. Παράλληλα, όσο μεγαλύτερα είναι τα δεδομένα, τόσο καλύτερα κλιμακώνει το πρόγραμά μας και τόσο καλύτερα μπορούν να μας βγουν τα αποτελέσματα αποδοτικότητας και επιτάχυνσης. Εδώ δεν είναι ιδανικά, κάτι που οφείλεται σε μεγάλο βαθμό στο μέγεθος των δεδομένων.

# ΕΞΟΔΟΣ mpiP

Η έξοδος που ακολουθεί αφορά το MPI πρόγραμμα με είσοδο μεγέθους χωρίς έλεγχο σύγκλισης για 1000 επαναλήψεις. Το πρόγραμμα τρέχει σε 2 κόμβους με 8 διεργασίες έκαστος. Η εντολή για compile είναι mpicc -O3 -g -Wall mpi\_heat.c -L$MPIPROOT/lib -lmpiP -lbfd -liberty -lunwind -o mpi\_heat.x. Επειδή η έξοδος είναι πολύ μεγάλη, παραθέτω ένα τμήμα της. Ολόκληρο το αρχείο με όνομα mpi\_heat.mpiP υπάρχει στον φάκελο που παραδίδεται.

@ mpiP

@ Command : grad1612\_mpi\_heat.x -c

@ Version : 3.4.1

@ MPIP Build date : Feb 6 2018, 18:24:04

@ Start time : 2019 02 20 12:46:02

@ Stop time : 2019 02 20 12:46:03

@ Timer Used : PMPI\_Wtime

@ MPIP env var : [null]

@ Collector Rank : 0

@ Collector PID : 15678

@ Final Output Dir : .

@ Report generation : Collective

@ MPI Task Assignment : 0 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 1 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 2 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 3 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 4 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 5 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 6 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 7 wn006.marie.hellasgrid.gr

@ MPI Task Assignment : 8 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 9 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 10 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 11 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 12 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 13 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 14 wn007.marie.hellasgrid.gr

@ MPI Task Assignment : 15 wn007.marie.hellasgrid.gr

---------------------------------------------------------------------------

@--- MPI Time (seconds) ---------------------------------------------------

---------------------------------------------------------------------------

Task AppTime MPITime MPI%

0 0.994 0.892 89.82

1 0.994 0.892 89.82

2 0.994 0.892 89.82

3 0.993 0.853 85.82

4 0.994 0.892 89.78

5 0.994 0.892 89.81

6 0.993 0.892 89.79

7 1.01 0.904 89.90

8 0.993 0.887 89.28

9 0.993 0.889 89.52

10 0.993 0.889 89.51

11 0.993 0.889 89.50

12 0.993 0.887 89.32

13 0.993 0.888 89.44

14 0.993 0.89 89.60

15 1 0.899 89.60

\* 15.9 14.2 89.40

---------------------------------------------------------------------------

@--- Callsites: 40 --------------------------------------------------------

---------------------------------------------------------------------------

ID Lev File/Address Line Parent\_Funct MPI\_Call

1 0 grad1612\_mpi\_heat.c 76 main Cart\_shift

2 0 grad1612\_mpi\_heat.c 181 main File\_set\_view

3 0 grad1612\_mpi\_heat.c 280 main File\_open

4 0 grad1612\_mpi\_heat.c 222 main Recv\_init

5 0 grad1612\_mpi\_heat.c 221 main Recv\_init

6 0 grad1612\_mpi\_heat.c 203 main Barrier

7 0 grad1612\_mpi\_heat.c 217 main Recv\_init

8 0 grad1612\_mpi\_heat.c 211 main Send\_init

9 0 grad1612\_mpi\_heat.c 207 main Send\_init

10 0 grad1612\_mpi\_heat.c 78 main Cart\_shift

11 0 grad1612\_mpi\_heat.c 180 main Type\_commit

12 0 grad1612\_mpi\_heat.c 175 main File\_open

13 0 grad1612\_mpi\_heat.c 224 main Recv\_init

14 0 grad1612\_mpi\_heat.c 73 main Cart\_create

15 0 grad1612\_mpi\_heat.c 271 main Waitall

16 0 grad1612\_mpi\_heat.c 220 main Recv\_init

17 0 grad1612\_mpi\_heat.c 277 main Reduce

18 0 grad1612\_mpi\_heat.c 214 main Send\_init

19 0 grad1612\_mpi\_heat.c 210 main Send\_init

20 0 grad1612\_mpi\_heat.c 282 main File\_close

21 0 grad1612\_mpi\_heat.c 308 main Type\_free

22 0 grad1612\_mpi\_heat.c 309 main Type\_free

23 0 grad1612\_mpi\_heat.c 187 main File\_close

24 0 grad1612\_mpi\_heat.c 219 main Recv\_init

25 0 grad1612\_mpi\_heat.c 141 main Type\_commit

26 0 grad1612\_mpi\_heat.c 213 main Send\_init

27 0 grad1612\_mpi\_heat.c 209 main Send\_init

28 0 grad1612\_mpi\_heat.c 232 main Startall

29 0 grad1612\_mpi\_heat.c 230 main Startall

30 0 grad1612\_mpi\_heat.c 185 main Type\_commit

31 0 grad1612\_mpi\_heat.c 223 main Recv\_init

32 0 grad1612\_mpi\_heat.c 143 main Bcast

33 0 grad1612\_mpi\_heat.c 218 main Recv\_init

34 0 grad1612\_mpi\_heat.c 306 main Type\_free

35 0 grad1612\_mpi\_heat.c 208 main Send\_init

36 0 grad1612\_mpi\_heat.c 138 main Type\_commit

37 0 grad1612\_mpi\_heat.c 241 main Waitall

38 0 grad1612\_mpi\_heat.c 144 main Bcast

39 0 grad1612\_mpi\_heat.c 307 main Type\_free

40 0 grad1612\_mpi\_heat.c 212 main Send\_init

---------------------------------------------------------------------------

@--- Aggregate Time (top twenty, descending, milliseconds) ----------------

---------------------------------------------------------------------------

Call Site Time App% MPI% COV

File\_open 12 4.65e+03 29.21 32.68 0.03

Waitall 37 3.29e+03 20.64 23.09 0.06

File\_open 3 2.27e+03 14.23 15.92 0.07

File\_close 20 1.17e+03 7.37 8.24 0.15

File\_close 23 998 6.27 7.01 0.09

Cart\_create 14 645 4.05 4.53 0.00

File\_set\_view 2 515 3.24 3.62 0.27

Barrier 6 263 1.65 1.85 0.36

Startall 29 229 1.44 1.61 0.48

Startall 28 94.8 0.60 0.67 0.16

Waitall 15 74.4 0.47 0.52 0.03

Send\_init 9 10.9 0.07 0.08 3.87

Bcast 32 9.19 0.06 0.06 0.25

Send\_init 8 8.83 0.06 0.06 3.91

Reduce 17 2.55 0.02 0.02 1.00

Send\_init 26 0.644 0.00 0.00 1.88

Send\_init 19 0.564 0.00 0.00 0.92

Type\_commit 36 0.451 0.00 0.00 0.15

Send\_init 40 0.435 0.00 0.00 1.14

Recv\_init 5 0.37 0.00 0.00 1.02

---------------------------------------------------------------------------

@--- Aggregate Sent Message Size (top twenty, descending, bytes) ----------

---------------------------------------------------------------------------

Call Site Count Total Avrg Sent%

Bcast 32 16 1.02e+03 64 47.06

Bcast 38 16 1.02e+03 64 47.06

Reduce 17 16 128 8 5.88

……….

---------------------------------------------------------------------------

@--- End of Report --------------------------------------------------------

---------------------------------------------------------------------------