Συστήματα Παράλληλης Επεξεργασίας

Ομάδα 6 Βασίλειος Βρεττός - el18126, Ανδρέας Βατίστας - el18020, Αλέξανδρος Τσάφος - el18211

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο

1 Άσκηση 1 - Εξοικείωση με το περιβάλλον προγραμματισμού

Η συγκεκριμένη άσκηση είναι εισαγωγική με σκοπό την εξοικείωση μας με τα μηχανήματα του cslab στα οποία εκτελούμε τις εργασίες. Για να δοκιμάσουμε την κατανόησή μας, ζητήθηκε να παραλληλοποιήσουμε το Conway's Game of Life, ένα απλό παιχνίδι το οποίο "παίζεται" πάνω σε ένα ταμπλό (στην δική μας περίπτωση, διαστάσεων N x N). Το παιχνίδι τρέχει για Κ γύρους (χρονικά διαστήματα). Το κάθε σημείο του ταμπλό έχει 2 καταστάσεις, είτε ζωντανό ή νεκρό. Οι 2 κύριοι κανόνες του παιχνιδιού είναι:

- 1. Αν ένα ζωντανό σημείο έχει περισσότερους από 3 γείτονες ή λιγότερους από 2, τότε γίνεται νεκρό στον επόμενο γύρο. Αν έχει ακριβώς 2 ή 3, παραμένει ζωντανό.
 - 2. Αν ένα νεκρό σημείο έχει ακριβώς 3 γείτονες, γίνεται ζωντανό στον επόμενο γύρο

Τρέχουμε το παραπάνω σενάριο για 1000 γενιές (γύρους) σε πίναχες διαστάσεων 64x64, 1024x1024 και 4096x4096.

Ο κώδικας ο οποίος ζητείται να παραλληλοποιηθεί είναι ο εξής.

```
current[i][j]=0;
}

#ifdef OUTPUT
print_to_pgm(current, N, t+1);
#endif
//Swap current array with previous array
swap=current;
current=previous;
previous=swap;
}
```

Το πρώτο for-loop, το οποίο είναι η αλλαγή των γενιών, δεν έχει νόημα να παραλληλοποιηθεί αφού χρειαζόμαστε την τιμή της προηγούμενης γενιάς για να υπολογίσουμε τις καταστάσεις των κελιών για την επόμενη γενιά.

Η δική μας προσθήκη στον κώδικα είναι η γραμμή.

```
#pragma omp parallel for shared (previous, current) private (i,j,nbrs)
```

Το συγκεκριμένο compiler directive (pragma) είναι χαρακτηριστικό του OpenMP. Είναι μια "οδηγία" προς τον μεταγλωττιστή η οποία του ζητάει να παραλληλοποιηθούν τα εσωτερικά for-loops με νήματα ορισμένα από ένα environmental variable. Στο συγκεκριμένο directive, επίσης ζητάμε:

- 1. Οι δείχτες previous, current να μοιράζονται μεταξύ των νημάτων. Η διάσταση N μοιράζεται by default.
 - 2. Οι μεταβλητές i,j,nbrs να είναι ξεχωριστές για κάθε νήμα.

Την απόδοση του πολυνηματισμού θα εξετάσουμε με χρήση την μετρική του χρόνου, και κατά συνέπεια, του speedup $S=\frac{T_s}{T_c}$.

```
GameOfLife: Size 64 Steps 1000 Time 0.023134 Threads: 1
GameOfLife: Size 64 Steps 1000 Time 0.013553 Threads: 2
GameOfLife: Size 64 Steps 1000 Time 0.010247 Threads: 4
GameOfLife: Size 64 Steps 1000 Time 0.009112 Threads: 6
GameOfLife: Size 64 Steps 1000 Time 0.009294 Threads: 8
GameOfLife: Size 1024 Steps 1000 Time 10.969249 Threads: 1
GameOfLife: Size 1024 Steps 1000 Time 5.452860 Threads: 2
GameOfLife: Size 1024 Steps 1000 Time 5.452860 Threads: 4
GameOfLife: Size 1024 Steps 1000 Time 1.828835 Threads: 6
GameOfLife: Size 1024 Steps 1000 Time 1.376873 Threads: 8
GameOfLife: Size 4096 Steps 1000 Time 175.911689 Threads: 1
GameOfLife: Size 4096 Steps 1000 Time 88.237651 Threads: 2
GameOfLife: Size 4096 Steps 1000 Time 44.518161 Threads: 4
GameOfLife: Size 4096 Steps 1000 Time 37.291715 Threads: 6
GameOfLife: Size 4096 Steps 1000 Time 37.291715 Threads: 6
```

Figure 1: Έξοδος Game of Life

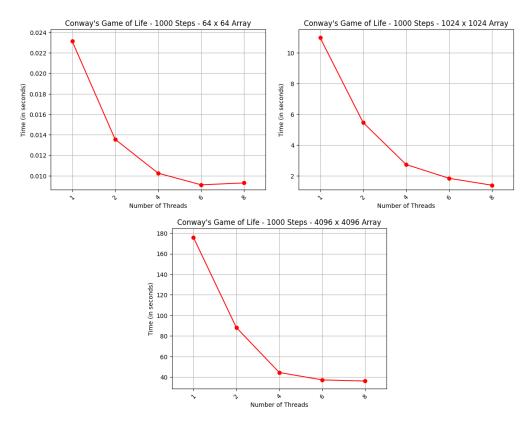


Figure 2: Χρόνος Εκτέλεσης Game of Life

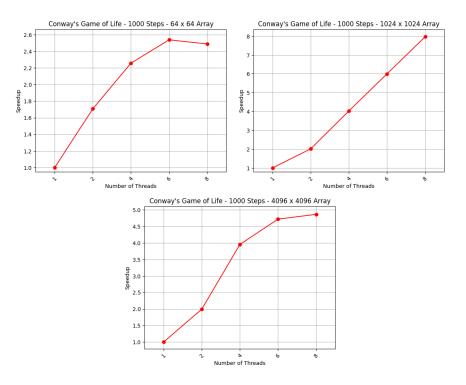


Figure 3: Speedup Game of Life $\,$

Συμπεράσματα

Για διαστάσεις 64x64

Παρατηρούμε ότι η μείωση του χρόνου δεν είναι αχριβώς ανάλογη των αριθμών των νημάτων. Από 1 σε 2 νήματα ή από 2 σε 4 νήματα βλέπουμε σημαντική βελτίωση της απόδοσης. Η εναλλαγή από 4 σε 6 νήματα προσφέρει πολύ μικρότερη αύξηση απόδοσης. Τέλος, από 6 σε 8 νήματα παρατηρούμε **ΑΥΞΗΣΗ** του χρόνου εκτέλεσης (μηδαμινή).

Η συγκεκριμένη αύξηση οφείλεται στο overhead που υπάρχει με την χρήση πολυνηματισμού σε μια διεργασία (γέννηση νημάτων, επικοινωνία κ.λ.π.). Ο συγκεκριμένος πίνακας είναι τόσο μικρός που η παραλληλοποίησή του περαιτέρω δεν βγάζει νόημα. Αν είχαμε την δυνατότητα να εξετάσουμε μεγαλύτερο αριθμό νημάτων, πολύ πιθανό να βλέπαμε περαιτέρω αύξηση του χρόνου εκτέλεσης

Για διαστάσεις 1024x1024

Παρατηρούμε ότι το speedup τείνει να είναι πλήρως γραμμικό. Ο χρόνος υποδιπλασιάζεται με κάθε διπλασιασμό νημάτων. Το συγκεκριμένο ταμπλό φαίνεται ιδανικό για παραλληλοποίηση. Η επικοινωνία μεταξύ νημάτων δεν περιορίζει την απόδοσή τους

Για διαστάσεις 4096x4096

Στους πρώτους 2 διπλασιασμούς των threads, βλέπουμε γραμμική μείωση του χρόνου (υποδιπλασιασμός) και κατά συνέπεια γραμμική αύξηση του speedup. Η λογική αυτή σταματάει να ισχύει για περαιτέρω αύξηση των νημάτων. Το συγκεκριμένο φαινόμενο δικαιολογείται από την ύπαρξη μεγάλου φόρτου (συμφόρησης) στον διάδρομο μνήμης. Υπάρχει συχνός διαμοιρασμός δεδομένων μεταξύ νημάτων καθώς αυξάνονται οι τιμές που πρέπει να ελεγχθούν για να υπολογιστούν οι γείτονες.

2 Άσκηση 2 - Παραλληλοποίηση και βελτιστοποίηση αλγορίθμων σε αρχιτεκτονικές κοινής μνήμης

Στην συγχεχριμένη άσχηση, σχοπός είναι η παραλληλοποίηση 2 διαφορετικών εκδόσεων των αλγορίθμων K-means και Floyd-Warshall.

2.1 Αλγόριθμος K-means

Ο αλγόριθμος k-means διαχωρίζει N αντιχείμενα σε k μη επιχαλυπτόμενες ομάδες. Ο παραχάτω ψευδοχώδιχας περιγράφει τον αλγόριθμο.

```
until convergence (or fixed loops)
  for each object
    find nearest cluster
  for each cluster
    calculate new cluster center coordinates.
```

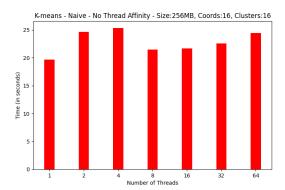
Στο συγκεκριμένο πρόβλημα μελετάμε 2 διαφορετικές εκδόσεις. Στην 1η, ο πίνακας των συντεταγμένων μοιράζεται από τα νήματα (shared cluster). Στην 2η υλοποίηση, δίνουμε στο κάθε νήμα μια τοπική έκδοση του πίνακα. Έτσι, τα νήματα μπορούν να κάνουν πράξεις χωρίς να υπάρχουν προβλήματα διαμοιρασμού (cache invalidation e.t.c.). Στο τέλος, κάθε νήμα ενημερώνει το master thread με τα δικα του αποτελέσματα (reduction).

2.1.1 K-means - Shared Clusters

Το μειονέκτημα της συγκεκριμένης υλοποίησης είναι ότι εφόσον λειτουργούμε πάνω σε διαμοιραζόμενα δεδομένα, υπάρχει ανάγκη η ανανέωση των shared variables **newClusterSize** και **newClusters** να γίνεται ατομικά.

Τα κλειδώματα για την ανανέωση των μεταλητών έγινε με την χρήση του **atomic** pragma. Το πρόγραμμα λήγει για συγκεκριμένα iterations ή μέχρι να συγκλίνει. Τα παρακάτω δεδομένα είναι για τον συνδυασμό {Size, Coords, Clusters, Loops} = {256, 16, 16, 10}.

```
#pragma omp parallel for default(shared) shared(newClusterSize, newClusters)
    private(i, j, index)
for (i=0; i<numObjs; i++){
/*
    *
    */
#pragma omp atomic
newClusterSize[index]++;
/*
    *
    */
#pragma omp atomic
newClusters[index*numCoords + j] += objects[i*numCoords + j];
}</pre>
```



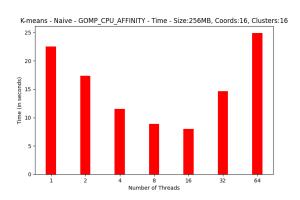
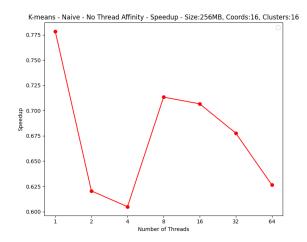


Figure 4: Χρόνος Εκτέλεσης K-Means Naive



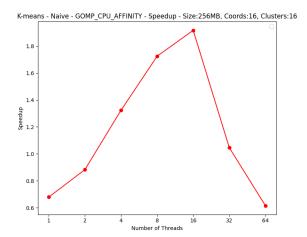


Figure 5: Speedup K-Means Naive

Σημείωση: Για το speedup θεωρούμε σειριακό χρόνο εκτέλεσης την έκδοση $X\Omega PI\Sigma$ κλειδώματα. Εδώ το speedup με 1 νήμα είναι μικρότερο της μονάδας γιατί έχουμε κλειδώματα.

Το άμεσο συμπέρασμα είναι ότι το πρόγραμμα δεν βελτιώνεται με χρήση παραπάνω νημάτων. Συγκεκριμένα, η αύξηση των διαθέσιμων νημάτων αυξάνει τον χρόνο εκτέλεσης.

Σαν προσπάθεια να βελτιώσουμε τα αποτελέσματα, χρησιμοποιήσαμε την μεταβλητή περιβάλλοντος GOMP_CPU_AFFINITY. Η λειτουργία αυτής της μεταβλήτης είναι ότι προσδένει τα "λογικά" νήματα σε hardware νήματα των επεξεργαστών βάσει μιας ακολουθίας. Γνωρίζοντας ότι οι επεξεργαστές του Sandman θεωρούν ότι π.χ. στο Physical Core 0 υπάρχουν τα Logical Cores 0,32, επιλέξαμε τα νήματα να προσδένονται ακολουθιακά στους φυσικούς πυρήνες των επεξεργαστών.

Παρατηρήσαμε βελτίωση των χρόνων για τα πρώτα 32 νήματα. Φυσικό ήταν να μην βελτιωθεί η κατάσταση για τα 64 νήματα.

2.1.2 K-means - Copied Clusters and Reduce

2.2 Αλγόριθμος Floyd-Warshall

Έστω πίναχας "γειτνίασης" Α. Ο κώδικας σε C για τον αλγόριθμο Floyd-Warshall είναι ο εξής:

```
for (k=0; k<N; k++)
  for (i=0; i<N; i++)
    for (j=0; j<N; j++)
        A[i][j] = min(A[i][j], A[i][k]+A[k][j]);</pre>
```

Το χομμάτι το οποίο μπορεί να παραλληλοποιηθεί είναι οι 2 εσωτερικά-φωλιασμένοι βρόγχοι.

2.2.1 Floyd-Warshall - Standard Edition

Η standard έχδοση του αλγορίθμου (σε C με OpenMP) είναι ως εξής:

```
for(k=0;k<N;k++)
    #pragma omp parallel for shared(A) private(i,j)
    for(i=0; i<N; i++)
        for(j=0; j<N; j++)
        A[i][j]=min(A[i][j], A[i][k] + A[k][j]);</pre>
```

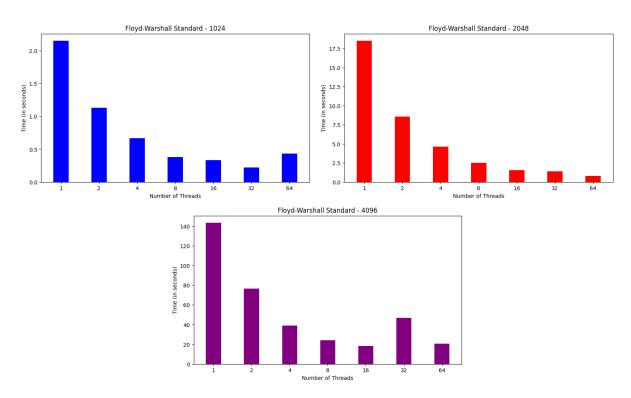


Figure 6: Χρόνος Εκτέλεσης FW Standard

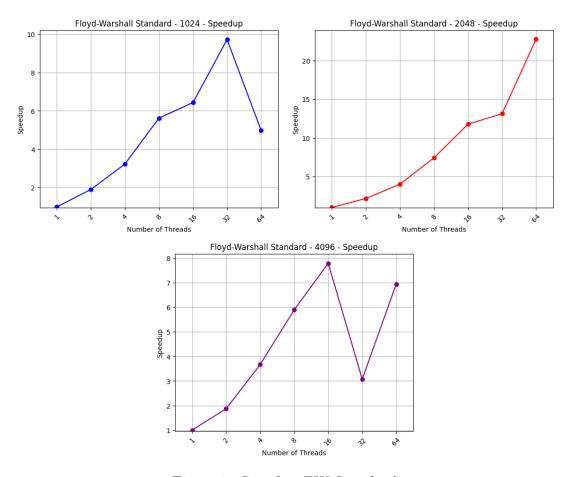


Figure 7: Speedup FW Standard

Best Times - Standard

Size 1024: 0.2213s - 32 Threads Size 2048: 0.8121s - 64 Threads Size 4096: 18.4082s - 16 Threads

Παρατηρούμε ότι τα συμπεράσματα πέρι κλιμάκωσης είναι ολόιδια με το πρόβλημα του Game Of Life. Για μικρούς πίνακες, υπάρχει μεγάλη δυνατότητα για κλιμάκωση, αλλά αυξάνοτας τα νήματα, το overhead που εισάγεται λόγω της επικοινωνίας μεταξύ τους περιορίζει την αύξηση της απόδοσης.

Για μεσαίους πίνακες, οι οποίοι χωράνε στην cache των πυρήνων, έχουμε ακόμα μεγαλύτερη δυνατότητα για κλιμάκωση.

Για μεγάλους πίναχες, το πρόβλημα γίνεται πιο περίπλοχο χαθώς πλέον δεν χωράει ο πίναχας στις cache μνήμες του επεξεργαστή. Αυξάνονται οι ανάγχες για μεταφορά cache lines. Άρα πλέον το πρόβλημα είναι memory bound.

2.2.2 Floyd-Warshall - Recursive/Task Based

Το νόημα της συγκεκριμένης έκδοσης του αλγορίθμου είναι να τρέχει αναδρομικά μέχρι να συναντήσει πίνακα μεγέθους, επιλεγμένο από εμας, βολικό για τον επεξεργαστή. Το μέγεθος αυτό ονομάζουμε block size. Δηλαδή, προσπαθούμε να χωρέσουμε ολόκληρο τον πίνακα στην cache. Παρακάτω παρουσιάζουμε το task graph της συνάρτησης αυτής για την πρώτη φορά που καλείται.

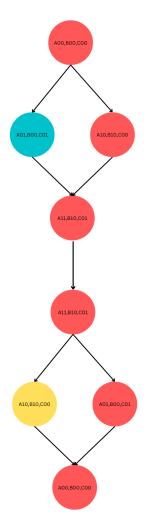
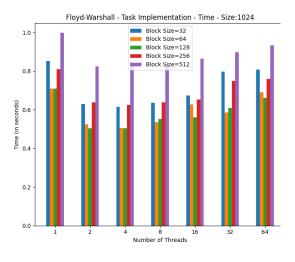
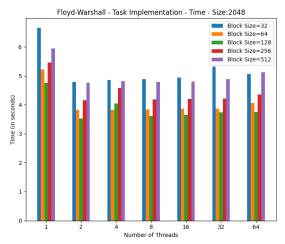


Figure 8: FW Recursive Task Graph - 1st iteration

Η παραλληλοποίηση αυτής της έκδοσης γίνεται με χρήση task, ώστε να μπορεί να είναι αναδρομική. Η OpenMP προσφέρει αυτή την δυνατότητα με την χρήση του

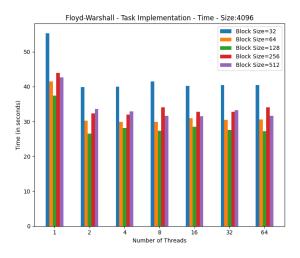
#pragma omp task





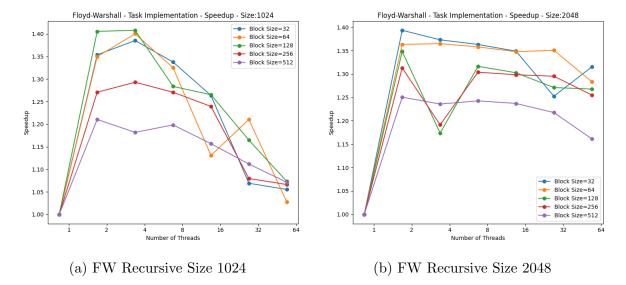
(a) FW Recursive Size 1024

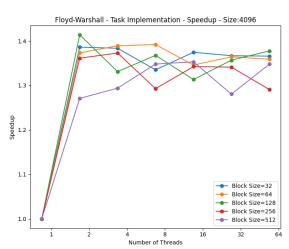
(b) FW Recursive Size 2048



(c) FW Recursive Size 4096

Figure 9: Χρόνος Εκτέλεσης FW Recursive





(c) FW Recursive - Size 4096 - Speedup

Figure 10: Speedup FW Recursive

Best Times - Task - No Nested

Size 1024: 0.5034s - 4 Threads - Block Size 128 Size 2048: 3.5248s - 4 Threads - Block Size 128 Size 4096: 26.4761s - 2 Threads - Block Size 128

Η πρώτη παρατήρηση είναι ότι το speedup δεν ξεπερνάει το 1.4. Σύμφωνα και με το task graph, η μέγιστη παραλληλοποίησή γίνεται με 2 νήματα σε κάθε επίπεδο αναδρομής. Ως προεπιλογή, το OpenMP ΔΕΝ επιτρέπει φωλιασμένη παραλληλοποίηση. Όταν ένα νήμα εισέρχεται σε παράλληλη περιοχή μέσα σε άλλη παράλληλη περιοχή, τότε δεν καλεί παραπάνω νήματα για την εκτέλεση αυτής της περιοχής. Άρα υπάρχουν περιοχές στο πρόγραμμα που εκτελούνται με το πολύ 1 εργάτη.

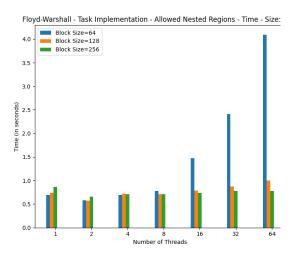
Η λύση για το παραπάνω είναι να αλλάξουμε την συμπεριφορά του OpenMP. Γίνεται να προσπεράσουμε αυτόν τον περιορισμό θέτοντας ένα environmental variable, το $\mathbf{OMP_NESTED}$ σε \mathbf{TRUE} .

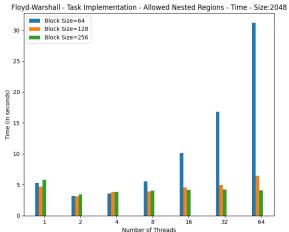
Η λύση αυτή δεν επαρχεί καθώς μετά μπορεί να επιφέρει αντίθετα αποτελέσματα από τα επιθυμητά. Επιτρέποντας την εχτέλεση φωλιασμένων παράλληλων περιοχών, υπάρχει ο χίνδυνος να καλέσουμε πολλά νήματα για την δημιουργία των task αλλά να μην υπάρχουν διαθέσιμα νήματα για την ίδια την εχτέλεση των πράξεων.

Χρειάζονται περαιτέρω περιορισμοί. Αυτοί δίνονται από το OpenMP με την μορφή μεταβλητών περιβάλλοντος.

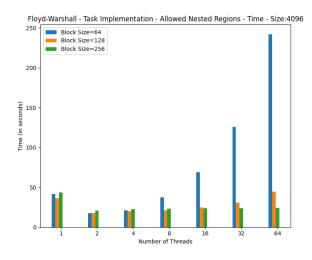
Οι SUNW_MP_MAX_POOL_THREADS και SUNW_MP_MAX_NESTED_LEVELS περιορίζουν τον αριθμό των διαθέσιμων νημάτων/δούλων (δηλαδή όλων εκτός του νήματος αφέντη) και το μέγιστο βάθος που επιτρέπονται από τα νήματα να καλούν άλλα νήματα αντίστοιχα.

Εφαρμόζοντας τα παραπάνω, βγάζουμε τα παρακάτω αποτελέσματα:



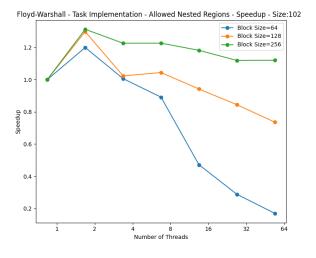


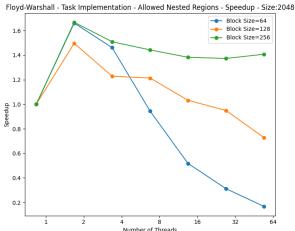
- (a) FW Recursive + Nested Size 1024
- (b) FW Recursive + Nested Size 2048



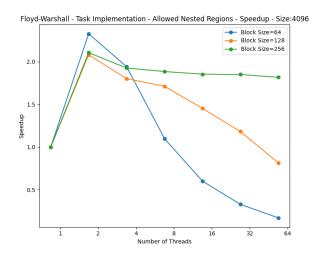
(c) FW Recursive + Nested Size 4096

Figure 11: Χρόνος Εκτέλεσης FW Recursive/Nested





- (a) FW Recursive Size 1024 Speedup
- (b) FW Recursive Size 2048 Speedup



(c) FW Recursive - Size 4096 - Speedup

Figure 12: Speedup FW Recursive/Nested

Best Times - Task - With Nesting

Size 1024: 0.5688s - 2 Threads - Block Size 128 Size 2048: 3.1347s - 4 Threads - Block Size 128 Size 4096: 17.3402s - 2 Threads - Block Size 128

Σημείωση: Για το speedup θεωρούμε σειριακό χρόνο εκτέλεσης για κάθε έκδοση τον χρόνο εκτέλεσης με ένα νήμα.

Και οι 2 εκδόσεις του recursive αλγορίθμου δεν φτάνουν σε απόδοση τον standard αλγόριθμο, σε ζήτημα scaling. Χρειάζεται περισσότερη μελέτη από το μέρος μας για να φτάσουμε καλύτερες επιδόσεις, κυρίως με καλύτερη χρήση των μεταβλητών περιβάλλοντος της OpenMP.

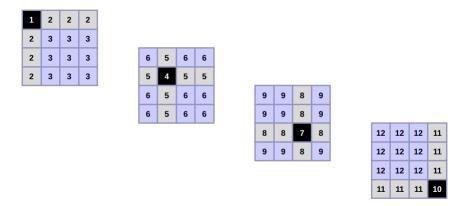
Η κατάλληλη επιλογή block size εξαρτάται κάθε φορά από τον επεξεργαστή που χρησι-

μοποιούμε. Σημαντικό να μην χρησιμοποιηθούν μικρά block sizes στην έκδοση που επιτρέπονται οι φωλιασμένες παράλληλες περιοχές γιατί γρήγορα υπερφορτώνεται το pool απο task όμως δεν υπάρχουν διαθέσιμα νήματα να δεχθούν task.

2.2.3 Floyd-Warshall - Tiled

Στην συγκεκριμένη έκδοση, εκμεταλλευόμαστε το παρακάτω σχήμα/ακολουθία πράξεων.

Figure 13: Βήματα Παραλληλισμού FW Tiled



Σε κάθε βήμα, υπολογίζουμε πρώτα το διαγώνιο στοιχείο του πίνακα (k-οστο), μετά υπολογίζουμε τα στοιχεία της ίδιας γραμμής και στήλης και μετά τα υπόλοιπα στοιχεία του πίνακα. Αυτές οι 3 ενέργειες πρέπει να γίνουν διαδοχικά μεταξύ τους με αυτή την σειρά. Αυτό σημαίνει ότι οι 2 τελευταίες ενέργειες μπορούν να γίνουν ταυτόχρονα (μεταξύ τους).

Υλοποιήσαμε 2 διαφορετικές εκδόσεις, μία με παραλληλισμό βρόγχων και μια με παραλληλισμό βρόγχων και tasks.

Τα αποτελέσματα που παρουσιάζουμε είναι από την έκδοση με τον παραλληλισμό βρόγχων.

Κώδιχας FW Tiled

```
for(k=0;k<N;k+=B){

FW(A,k,k,k,B);
  #pragma omp parallel shared(A,k,B)
{

    #pragma omp for nowait
    for(i=0; i<k; i+=B)
        FW(A,k,i,k,B);

    #pragma omp for nowait
    for(i=k+B; i<N; i+=B)
        FW(A,k,i,k,B);</pre>
```

```
#pragma omp for nowait
      for(j=0; j<k; j+=B)</pre>
         FW(A,k,k,j,B);
      #pragma omp for nowait
      for(j=k+B; j<N; j+=B)</pre>
         FW(A,k,k,j,B);
   }
   #pragma omp parallel shared(A,k,B)
      #pragma omp for collapse(2) nowait
      for(i=0; i<k; i+=B)</pre>
         for(j=0; j<k; j+=B)</pre>
            FW(A,k,i,j,B);
      #pragma omp for collapse(2) nowait
      for(i=0; i<k; i+=B)</pre>
         for(j=k+B; j<N; j+=B)</pre>
            FW(A,k,i,j,B);
      #pragma omp for collapse(2) nowait
      for(i=k+B; i<N; i+=B)</pre>
         for(j=0; j<k; j+=B)</pre>
            FW(A,k,i,j,B);
      #pragma omp for collapse(2) nowait
      for(i=k+B; i<N; i+=B)</pre>
         for(j=k+B; j<N; j+=B)</pre>
            FW(A,k,i,j,B);
   }
}
```

Η επιλογή nowait χρησιμοποιείται αφού μεταξύ τους οι βρόγχοι της κάθε παράλληλης περιοχής μπορούν να τρέξουν ταυτόχρονα.

Η επιλογή collapse(2) χρησιμοποιείται καθώς χωρίς αυτήν, ο 2ος βρόγχος θα έτρεχε μονονηματικά, από το νήμα που εκτελεί το for loop. Με το συγκεκριμένο, μπορούν να χρησιμοποιηθούν όλα τα νήματα του συστήματος για να παραλληλοποιηθεί ο τέλεια φωλιασμένος βρόγχος.

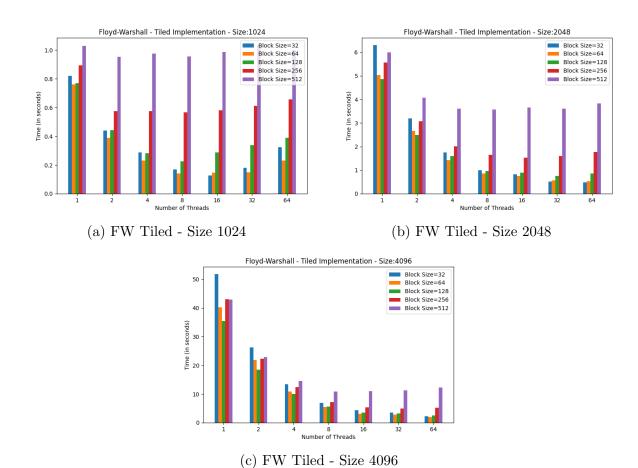
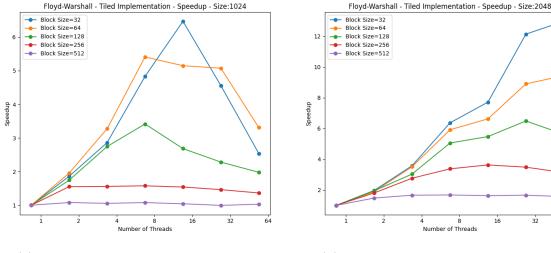
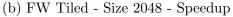
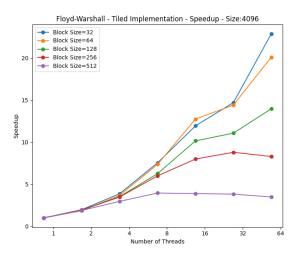


Figure 14: Χρόνος Εκτέλεσης FW Tiled



(a) FW Tiled - Size 1024 - Speedup





(c) FW Tiled - Size 4096 - Speedup

Figure 15: Speedup FW Tiled

Best Times - Tiled - ParFor

Size 1024: 0.1267s - 16 Threads - Block Size 32 Size 2048: 0.4858s - 64 Threads - Block Size 32 Size 4096: 1.9965s - 64 Threads - Block Size 64

Οι χρόνοι εκτέλεσης είναι πλέον πολύ καλύτεροι και από τις 2 εκδόσεις. Συγκριτικά με την αρχική έκδοση, όπου ο καλύτερος χρόνος ήταν 20.6s (με 64 νήματα) για τον πίνακα μεγέθους 4096, πλέον βέλτιστος χρόνος είναι 1.99s για block size=64.

Η Tiled έκοδση φαίνεται να κλιμακώνει πολύ καλύτερα από την standard για μεγάλους πίνακες. Για μεσαίους και μικρούς, τα αποτελέσματα είναι αρκετά παρόμοια. Είναι καλύτερη επιλογή από άποψη αξιοποίησης πόρων καθώς ακόμα και με μικρότερο αριθμό νημάτων, πετυχαί-

νουμε πολύ καλύτερους χρόνους.