

Práctica 2: Computación paralela con planetas

Índice

1.	Ejercicio 1.....	2
2.	Ejercicio 2.....	2
2.1	Planets2i - primer <i>bucle i</i>	2
2.2	Planets2j.....	4
2.3	Ambas versiones - último <i>bucle i</i>	5
3.	Ejercicio 3.....	6
4.	Ejercicio 4.....	7
4.1	Ejecuciones en Kahan	8
4.2	Planificaciones parciales - planets2i con distintas planificaciones para solo un bucle i	9
4.2.1	Planets2i con distinta planificación para el primer bucle i	9
4.2.2	Planets2i con distintas planificaciones para el segundo bucle i	11
4.3	Planificaciones para todo el programa	14
4.3.1	Planets2i completamente planificado.....	14
4.3.2	Planets2j completamente planificado.....	17
4.4	Conclusiones	20
5.	Ejercicio 5.....	22

Integrantes del equipo:

- **David Arnal García**
- **Alejandro Sánchez Melero**

1. Ejercicio 1

En el inicio del programa principal (*main*) se han declarado dos variables, las variables $t1$ y $t2$, de tipo doble (*double*). A continuación, se imprime por pantalla el tiempo de ejecución, siendo este la resta de $t2 - t1$, así como el número de planetas, siendo este último el valor de una variable auxiliar n que toma el valor de N (*i.e.* $n = N$).

```
generate_data(n, p);

t1 = omp_get_wtime();
k = process_data(n, p, &cm);
t2 = omp_get_wtime();

printf("The center of mass is in (%g, %g, %g).\n",
      cm.x, cm.y, cm.z);
printf("The planet with most neighbors is %d (%d neighbors).\n",
      k, p[k].ngb);

printf("Spent time: %f s.\n", t2 - t1);
printf("Number of planets: %d\n", n);

free(p);
return 0;
```

Ilustración 1

2. Ejercicio 2

En este ejercicio, explicaremos cada bucle por separado y, por último, el *bucle i final* que comparten ambas versiones. También es importante mencionar que ha sido posible realizar la paralelización de todos los bucles debido a que no hay ninguna dependencia de datos entre los bucles.

2.1 Planets2i – primer *bucle i*

En este archivo se ha paralelizado el bucle externo y, por tanto, se consigue aquí la paralelización óptima. Al paralelizar el programa, se deben crear unas nuevas variables de tipo *double* que se sustituirán por las variables del *struct Point3D*, *i.e.* $cm.x$, $cm.y$, $cm.z$, ya que, al paralelizar este programa, *OpenMP*

no puede acceder a las estructuras de datos, por lo que se sustituyen todos los accesos a la estructura *Point3D* por las variables anteriormente mencionadas. El fragmento de código empleado aquí es:

```
int i;
int j;
int ngbi;
int k;
int max;
double dx;
double dy;
double dz;
double d;
double m;
Point3D cm; // center of mass
double cm_X = 0;
double cm_Y = 0;
double cm_Z = 0;
m = 0;
```

Ilustración 2

Posteriormente, se debe actualizar correctamente fuera del bucle la estructura *Point3D* con sus valores correspondientes. El código de este fragmento es:

```
cm.x = cm_X;
cm.y = cm_Y;
cm.z = cm_Z;
```

Ilustración 3

Ahora mencionaremos las directivas para conseguir paralelización en el primer bucle que aparece en el programa. Se han incluido como privadas las variables *j*, *d*, *dx*, *dy*, *dz*, *ngbi*, ya que cada hilo toma sus propios valores en cada iteración. Además, las variables *m*, *cm_x*, *cm_y*, *cm_z*, al contribuir al estado global del programa, son de tipo reducción (*reduction*).

```
#pragma omp parallel for private(j, dx, dy, dz, d, ngbi) reduction(+ \
: m, cm_X, cm_Y, cm_Z)
for (i = 0; i < n; i++)
{
```

Ilustración 4

2.2 Planets2j

En esta versión del bucle se paraleliza el bucle interno, es decir, el bucle j . Esta versión será menos óptima que la anterior y, como se podrá observar, posteriormente, en otro ejercicio, se obtendrá una mayor complejidad temporal que en el bucle externo debido a la mayor sobrecarga por la activación y desactivación de hilos. En este bucle se han declarado como privadas las variables d, dx, dy, dz , ya que, en cada iteración, cada hilo tendrá un valor propio, y no compartido.

Además, como intento de optimización de código y aseguramiento de un correcto valor para la variable $ngbi$, se ha incluido dentro de una cláusula reducción.

```
#pragma omp parallel for private(dx, dy, dz, d) reduction(+ \
: ngbi)
```

Ilustración 5

En esta versión, es importante mencionar que al realizar la instrucción *if (...)* no se producen condiciones de carrera, ya que cada hilo mantiene su propio valor para d , pero sí que es importante incluir una directiva *atomic* al vector p en la posición j , ya que todos los accesos y escrituras en las variables se producen de forma concurrente y j , en este caso, no es una variable privada.

```
if (d < NGB_DST)
{
    // Planets i and j are neighbors. Update number of neighbors
    // for both planets
    ngbi++;
#pragma omp atomic
    p[j].ngb++;
}
p[i].ngb += ngbi;
}
```

Ilustración 6

2.3 Ambas versiones – último *bucle i*

El último *bucle i* que aparece en ambas versiones se puede paralelizar. En este bucle no se ha declarado ninguna variable como privada ni de tipo reducción, por lo que todas las variables son compartidas. Sin embargo, en este bucle sí que se produce una condición de carrera a la hora de acceder a la escritura en las variables k , max , por lo que se deberá declarar una sección crítica y duplicar la sentencia *if*, además de incorporar en ella el comparador \geq , ya que, si un valor de planetas vecinos coincide con el máximo, se deberá seleccionar el planeta de menor índice. Por este motivo, se introduce otro *if* (...) donde se calcula el mínimo entre los dos planetas con una función propia, llamada *minimum*, con el siguiente código:

```
int minimum(int num1, int num2)
{
    if (num1 <= num2)
    {
        return num1;
    }
    else
    {
        return num2;
    }
}
```

Ilustración 7

Siendo el código completo del último *bucle i* el siguiente:

```
// Find the planet with most neighbors
k = 0;
max = -1;

#pragma omp parallel for
for (i = 0; i < n; i++)
{
    if (p[i].ngb >= max)
    {
#pragma omp critical
        if (p[i].ngb >= max)
        {
            if (p[i].ngb == max)
            {
                k = minimum(k, i);
            }
            else
            {
                max = p[i].ngb;
                k = i;
            }
        }
    }
}
```

Ilustración 8

3. Ejercicio 3

En el siguiente ejercicio mostraremos explícitamente distintas planificaciones para los programas. En este momento, comenzaremos analizando qué sucede si solo planificamos un solo bucle *i* en *planets2i*. Si, por ejemplo, se planifica el primer bucle *i*, **no encontraríamos diferencias significativas de rendimiento entre las distintas planificaciones**. En otro caso, si únicamente se planifica el segundo bucle *i*, **tampoco**

tendría por qué haber diferencias significativas en cuanto a las prestaciones, tal y como se podrá comprobar en el siguiente ejercicio.

Por otra parte, si hacemos un análisis más general, **no se encontrará ninguna planificación óptima** (es decir, mejor que las demás) **si se planifica el programa *planets2i* completamente**, es decir, ambos bucles *i*, ya que la asignación de los hilos siempre se realizará desde el inicio de los bucles y no habrá problemas de rendimiento por ello, aunque podría ser que la planificación *guided* diera peores resultados debido a una peor gestión de hilos, ya que no son tareas cortas y sencillas, y podría no ser óptimo el tiempo de ejecución para este caso.

Sin embargo, si se planifica el programa *planets2j* completamente, sí habrá diferencias notables, siendo **la planificación *static con chunk* la mejor de todas ellas**, y cuanto más grande el *chunk*, mejores prestaciones para todas las planificaciones, ya que habrá menos desactivaciones y reactivaciones de hilos. **Además, la peor planificación, en este caso, sería la planificación dinámica**, que sufriría bastante en cuanto a prestaciones debido a las continuas activaciones y desactivaciones de los hilos que se realizan con esta planificación, y destacar que, a mayor *chunk*, las prestaciones mejorarían, aunque no a un nivel óptimo, debido a que aquí la planificación dinámica no es adecuada, en general.

Todo lo anterior comentado se analizará y se podrá observar detalladamente en el posterior ejercicio.

4. Ejercicio 4

En este ejercicio, explicaremos cómo hemos lanzado los programas en *Kahan*, a continuación, analizaremos unas planificaciones parciales del programa *planets2i* y, finalmente, analizaremos los programas *planets2i* y *planets2j* completamente planificados.

4.1 Ejecuciones en *Kahan*

Antes de analizar las planificaciones gráficamente, vamos a explicar cómo lanzamos los programas en *Kahan*. Para esto, hemos creado unos archivos de extensión *.sh* llamados *workopenmp2.sh* y *workopenmp3.sh* en los que hemos ido haciendo las ejecuciones de *planets2i.c* y *planets2j.c*, respectivamente.

En estos ficheros, se han definido las propiedades del trabajo sobre el sistema de colas, en el que elegimos trabajar con un único nodo de los seis disponibles en *Kahan*, ya que se trabaja en una máquina de memoria compartida donde todas las ejecuciones se realizan sobre el mismo nodo, y no se puede acceder a la memoria de otros nodos. Por tanto, decir, en este momento, que, se ha decidido ejecutarlo hasta con 32 hilos y no con más, ya que los programas se ejecutan solo en un nodo de *Kahan*, que tiene 2 procesadores y cada uno de ellos 16 núcleos, por lo que un nodo únicamente admite hasta 32 hilos, y no tendría sentido ejecutarlo con más hilos.

Por otra parte, junto al número de nodos elegidos, incluimos un *walltime* con un tiempo máximo de duración para nuestras ejecuciones; En nuestro caso, hemos decidido utilizar un tiempo límite de ejecución de 10 minutos.

Además, siendo de la asignatura CPA, escogemos trabajar en la cola de dicha asignatura con el comando *-q* y, por último, utilizamos la opción *-d* indicando que nuestro fichero estará en el mismo directorio desde donde se ejecutará *Kahan*.

Por último, en ambos ficheros, utilizamos diferentes líneas de ejecuciones para ejecutar ambos programas con diferentes números de hilos y planificaciones. Por ejemplo, para el archivo *planets2i.c*, un fragmento de código de *workopenmp2.sh* es:


```

#!/bin/sh
#PBS -l nodes=1,walltime=00:10:00
#PBS -q cpa
#PBS -d .

# OMP_NUM_THREADS=2 OMP_SCHEDULE=static ./planets2i
# OMP_NUM_THREADS=4 OMP_SCHEDULE=static ./planets2i
# OMP_NUM_THREADS=8 OMP_SCHEDULE=static ./planets2i
# OMP_NUM_THREADS=16 OMP_SCHEDULE=static ./planets2i
# OMP_NUM_THREADS=32 OMP_SCHEDULE=static ./planets2i

# OMP_NUM_THREADS=2 OMP_SCHEDULE=static,1 ./planets2i
# OMP_NUM_THREADS=4 OMP_SCHEDULE=static,1 ./planets2i
# OMP_NUM_THREADS=8 OMP_SCHEDULE=static,1 ./planets2i
# OMP_NUM_THREADS=16 OMP_SCHEDULE=static,1 ./planets2i
# OMP_NUM_THREADS=32 OMP_SCHEDULE=static,1 ./planets2i

# OMP_NUM_THREADS=2 OMP_SCHEDULE=static,2 ./planets2i
# OMP_NUM_THREADS=4 OMP_SCHEDULE=static,2 ./planets2i
# OMP_NUM_THREADS=8 OMP_SCHEDULE=static,2 ./planets2i
# OMP_NUM_THREADS=16 OMP_SCHEDULE=static,2 ./planets2i
# OMP_NUM_THREADS=32 OMP_SCHEDULE=static,2 ./planets2i

```

Ilustración 9

Y el archivo continuaría con ejecuciones con las planificaciones y número de hilos que deseáramos. Lo mismo sucedería para el archivo *planets2j.c*, cambiando únicamente en el archivo *.sh ./planets2i* por *./planets2j*.

En la entrega, adjuntaremos los archivos *workopenmp2.sh* y *workopenmp3.sh*.

4.2 Planificaciones parciales - planets2i con distintas planificaciones para solo un bucle *i*

4.2.1 Planets2i con distinta planificación para el primer bucle *i*

Por ejemplo, en el histograma *Planets2i con distintas planificaciones para el primer bucle *i**, la planificación del segundo bucle *j* se mantiene por defecto. Debajo de cada histograma se sacarán sus correspondientes aceleraciones y eficiencias. Además, **todos los tiempos se han tomado con un número de iteraciones $N = 10.000 \cdot 8 = 80.000$ iteraciones.**

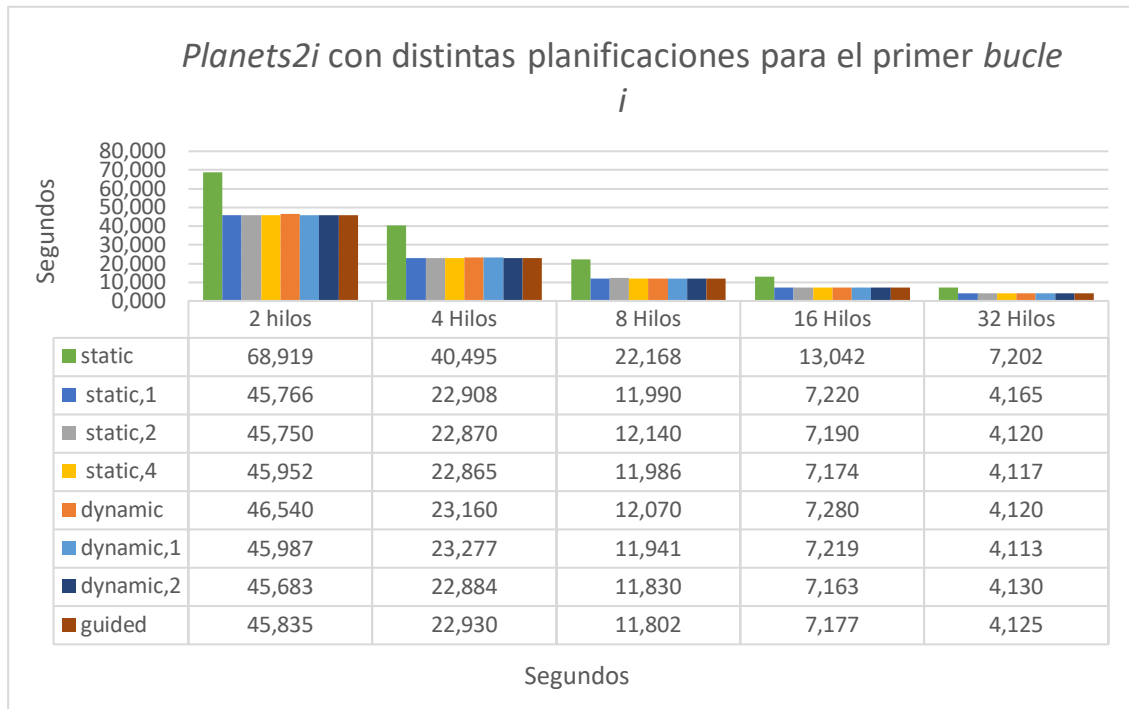


Ilustración 10

Para tomar las aceleraciones, se ha ejecutado *planets1.c* y ha dado como $S(n) = 84.700411$ s. Se utilizará este valor para todas las aceleraciones.

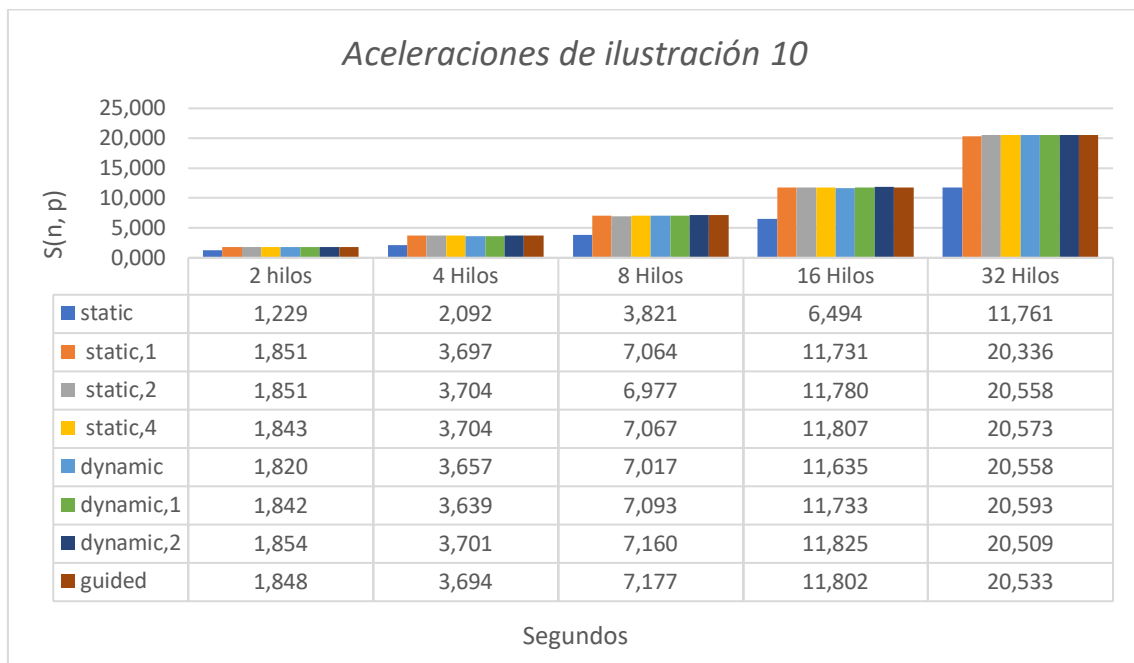


Ilustración 11

Como se ha podido observar en las ejecuciones anteriores, la planificación estática *sin chunk* es la que mayor tiempo ha dado, por lo que es evidente que es la que tiene

una menor aceleración en esta gráfica. Respecto a todas las planificaciones que hemos ejecutado, es la que se puede destacar, ya que las demás obtienen una aceleración muy parecida.

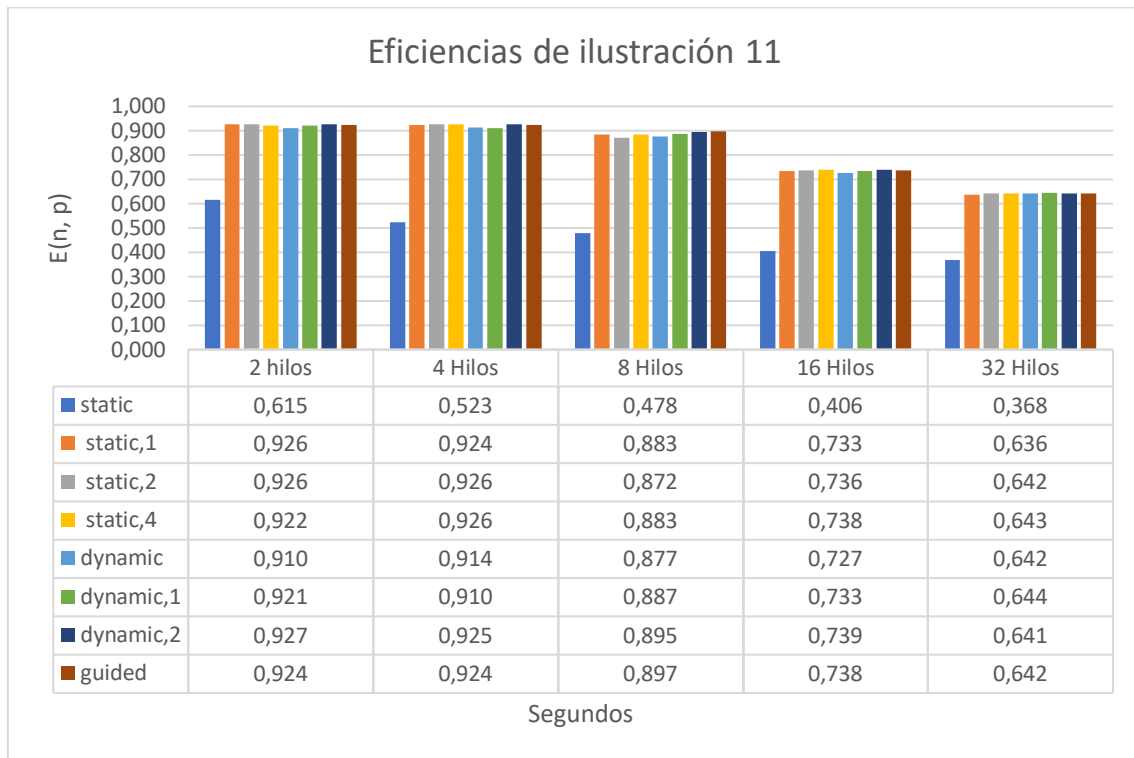


Ilustración 12

Por tanto, todas las eficiencias en esta gráfica son muy parecidas, excepto para la planificación *static* sin *chunk*, que obtiene una eficiencia menor.

4.2.2 Planets2i con distintas planificaciones para el segundo bucle i

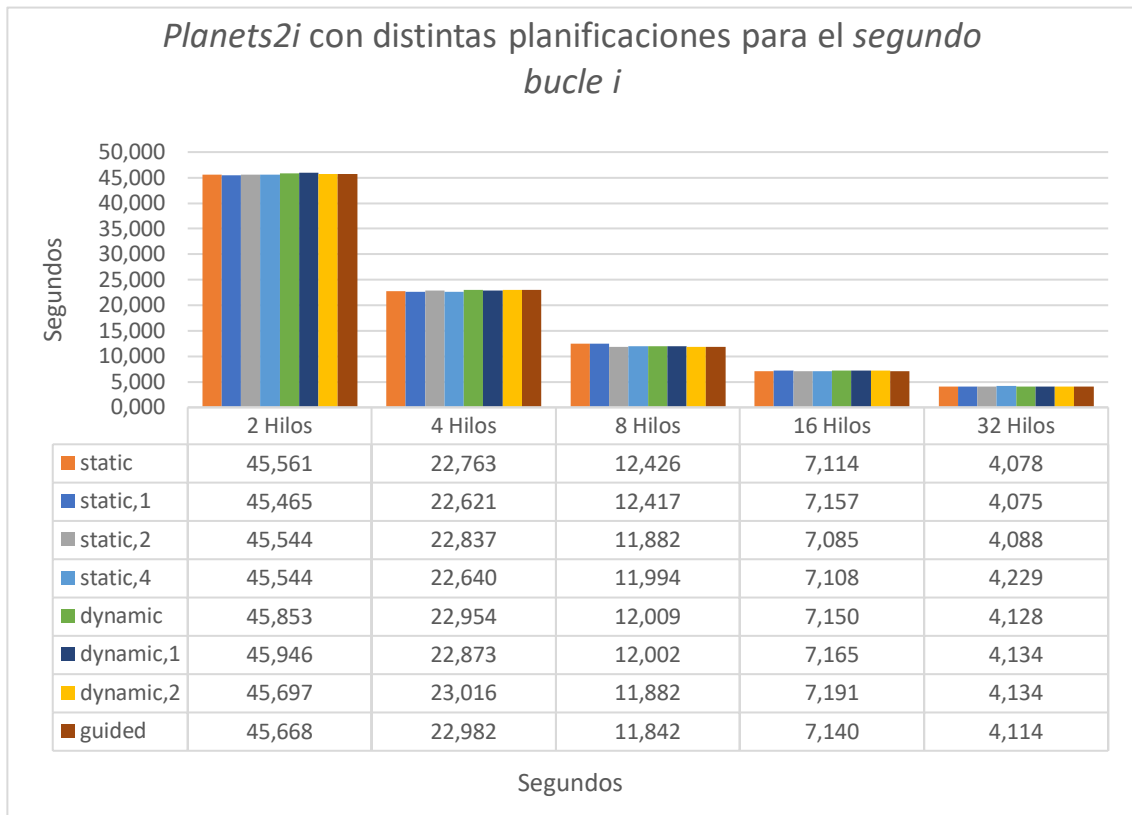
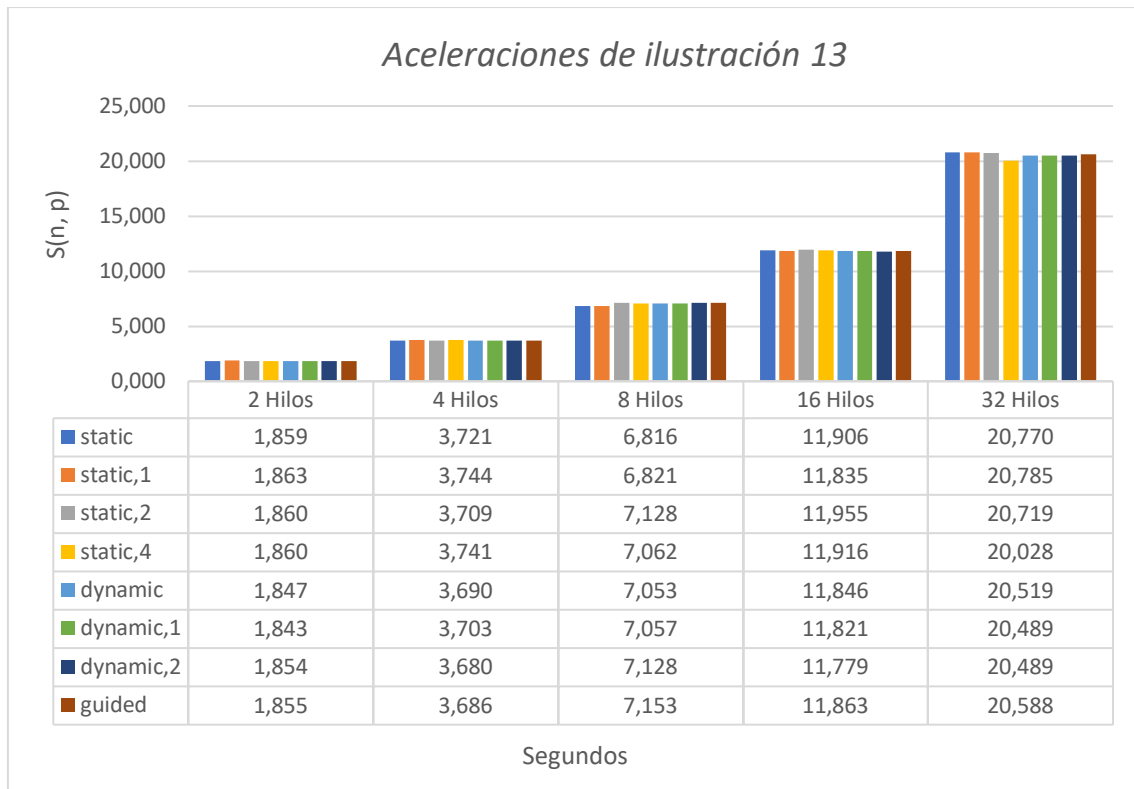


Ilustración 13

Si ejecutamos *planets2i.c* únicamente planificando el segundo bucle *i*, no encontraremos ninguna diferencia significativa respecto a planificaciones.

*Ilustración 14*

Por esto, todas las aceleraciones serán muy parecidas.

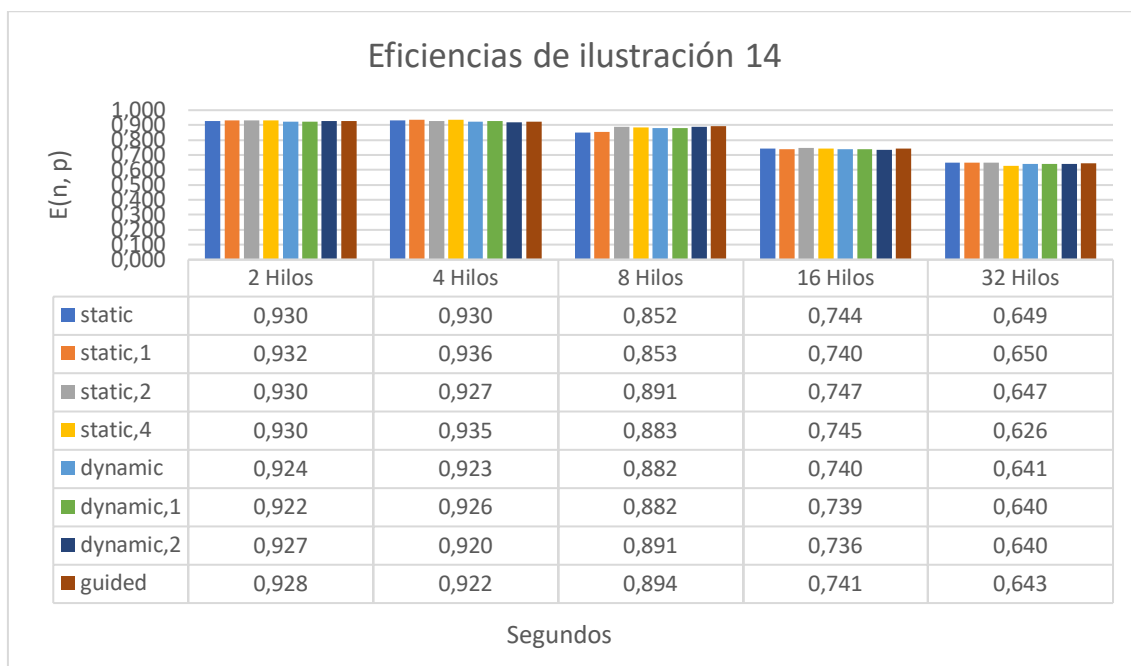


Ilustración 15

Así como sus eficiencias.

Por otra parte, se han ejecutado los programas completos con determinadas planificaciones, pudiendo observarse algunos casos particulares en los que el coste aumenta considerablemente.

4.3 Planificaciones para todo el programa

4.3.1 Planets2i completamente planificado

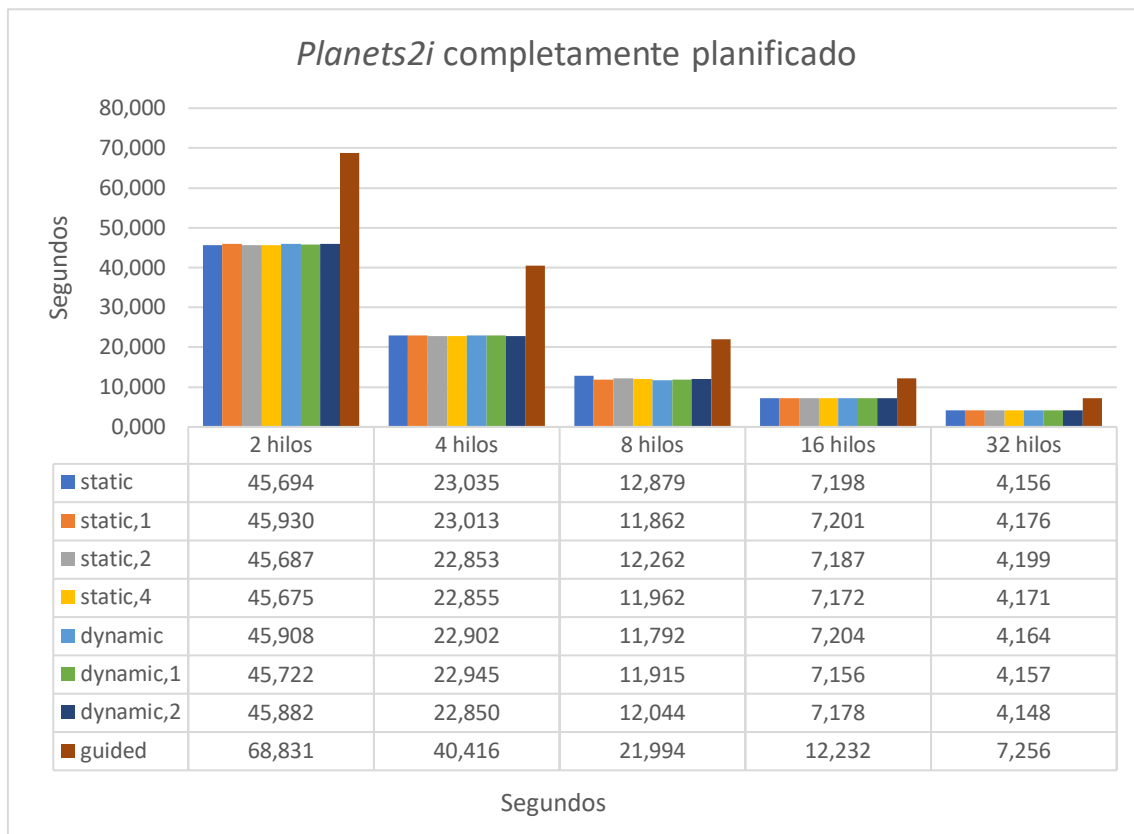


Ilustración 16

En este apartado comenzamos analizando qué sucede si se planifica el programa *planets2i.c* completamente con una planificación determinada. En este caso, el tiempo de ejecución únicamente es mayor en el caso de la planificación *guided*, siendo mayor debido a una mala gestión de la desactivación y reactivación de los hilos. En cuanto a las demás, no hay nada importante que decir, debido a que todas ellas tienen tiempos similares.

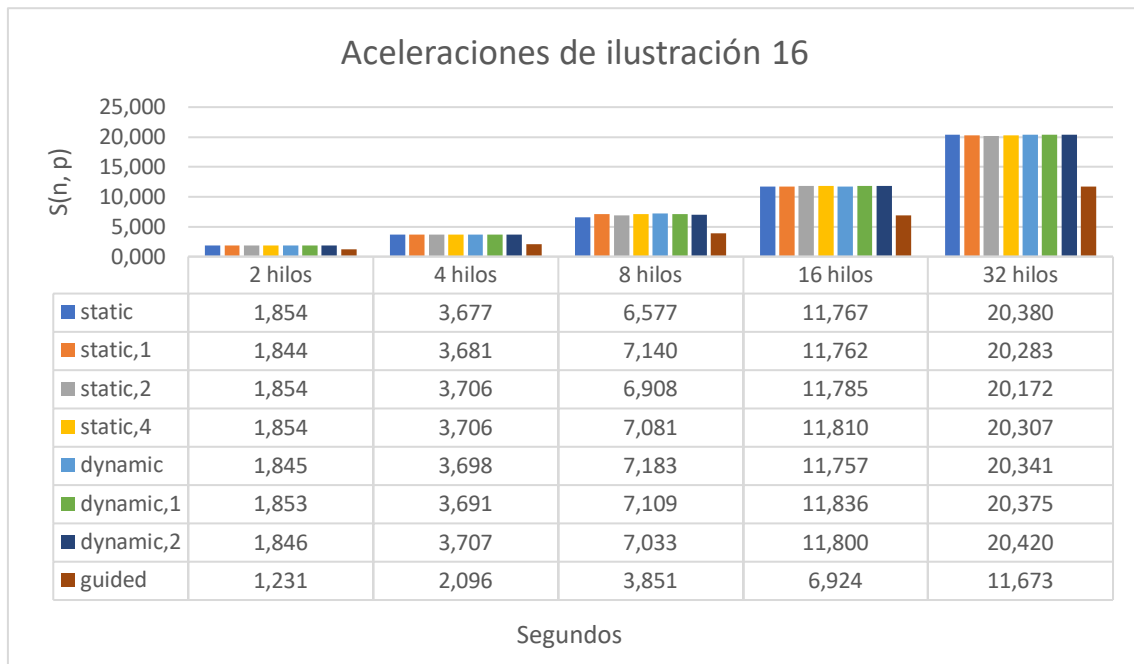


Ilustración 17

Por lo anterior mencionado, la planificación *guided* aquí es la que tiene una menor aceleración.

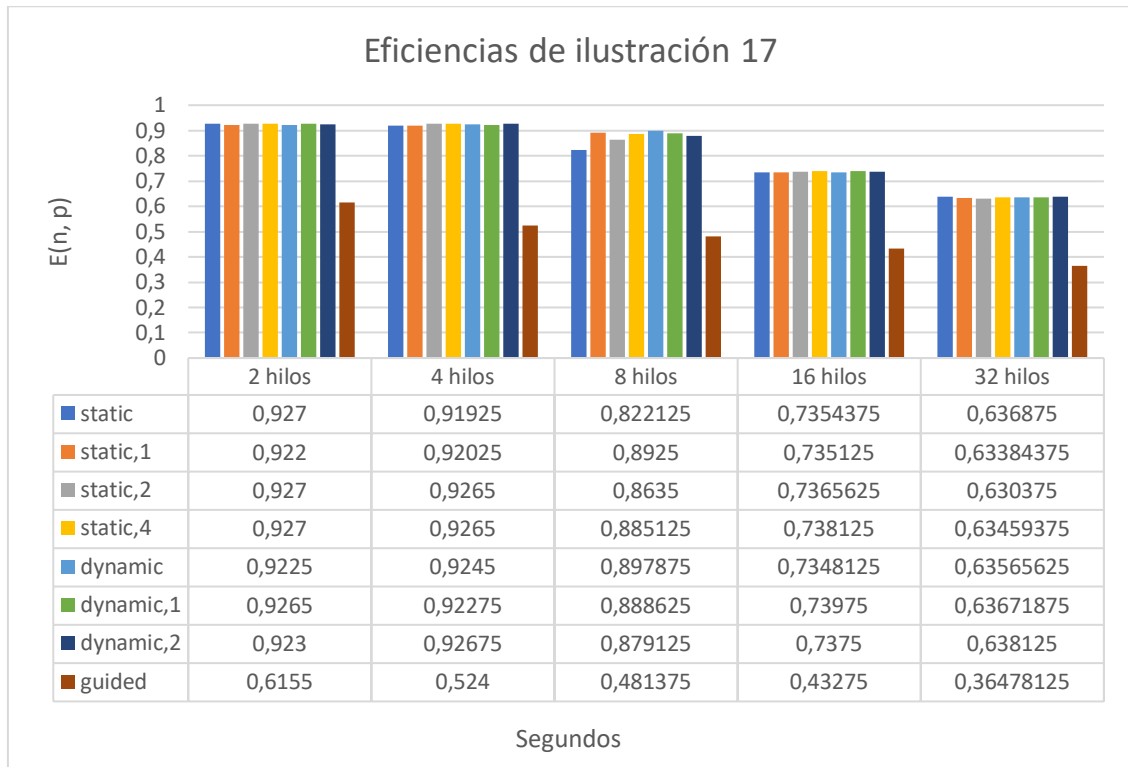


Ilustración 18

Y, de la misma forma, es la que cuenta con una menor eficiencia.

4.3.2 Planets2j completamente planificado

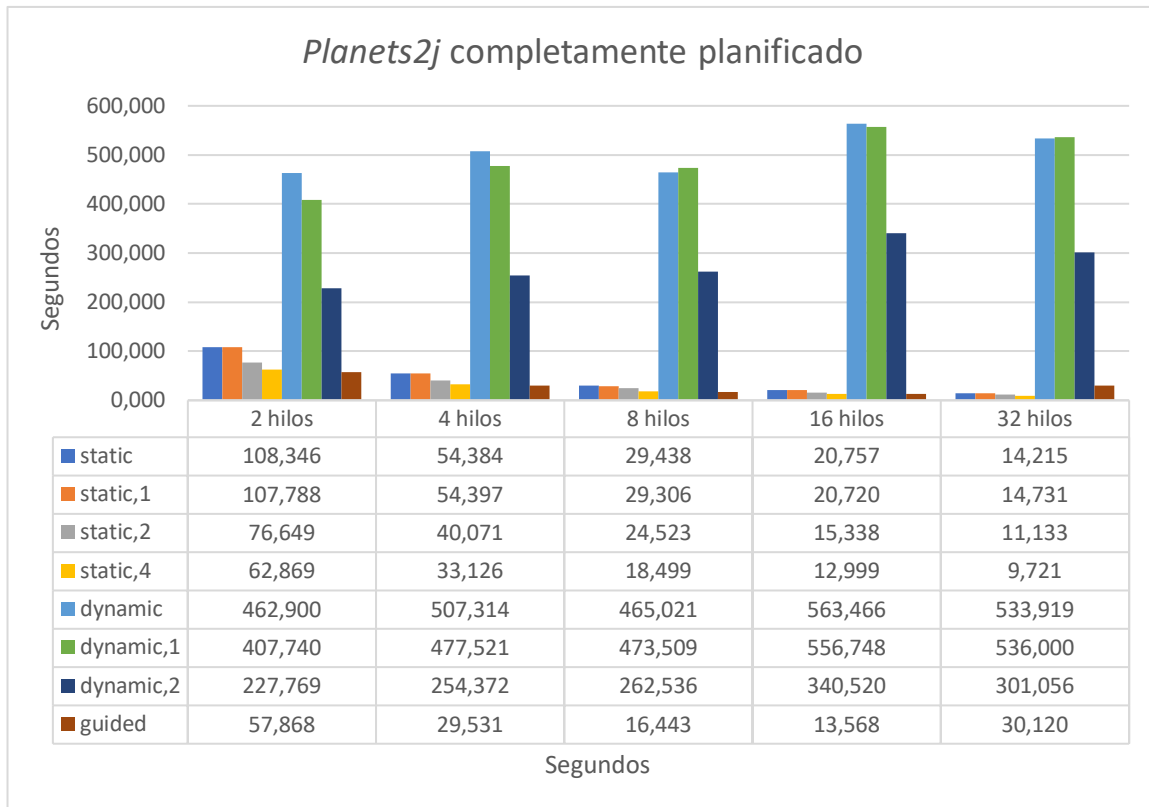


Ilustración 19

Aquí se puede destacar, para empezar, el mal rendimiento de la planificación dinámica, debido a las continuas desactivaciones y reactivaciones de hilos que realiza, comparado con las planificaciones estáticas, por ejemplo.

Aunque la planificación *guided* obtendría un rendimiento parecido a la planificación estática con varios *chunks* y con pocos hilos, se dispara cuando se ejecuta con muchos hilos, por ejemplo, con 32 hilos, por lo que no sería tan buena su gestión con las activaciones y desactivaciones de hilos.

Además, destacar aquí que cuanto mayor *chunk* en cualquier planificación, mejor será el rendimiento, debido a que se harán menos desactivaciones y activaciones de hilos, siendo como mejor planificación, para este caso, la estática (con mayor *chunk*).

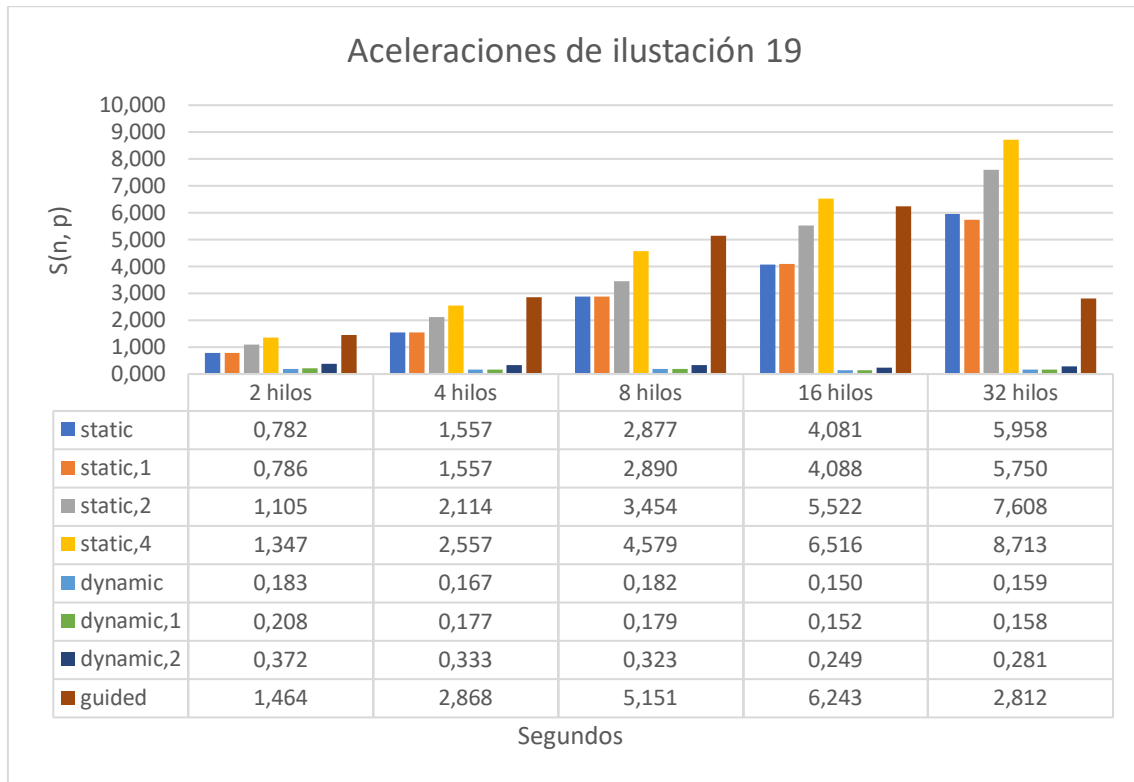


Ilustración 20

Por tanto, la mayor aceleración será para el *static* con mayor *chunk* (i.e. *static, 4*), seguido del *guided* hasta 16 hilos, y de los demás *static* con menor *chunk*, para el caso general. Por otra parte, las planificaciones dinámicas serán la que peor rendimiento obtengan debido a lo anterior mencionado.

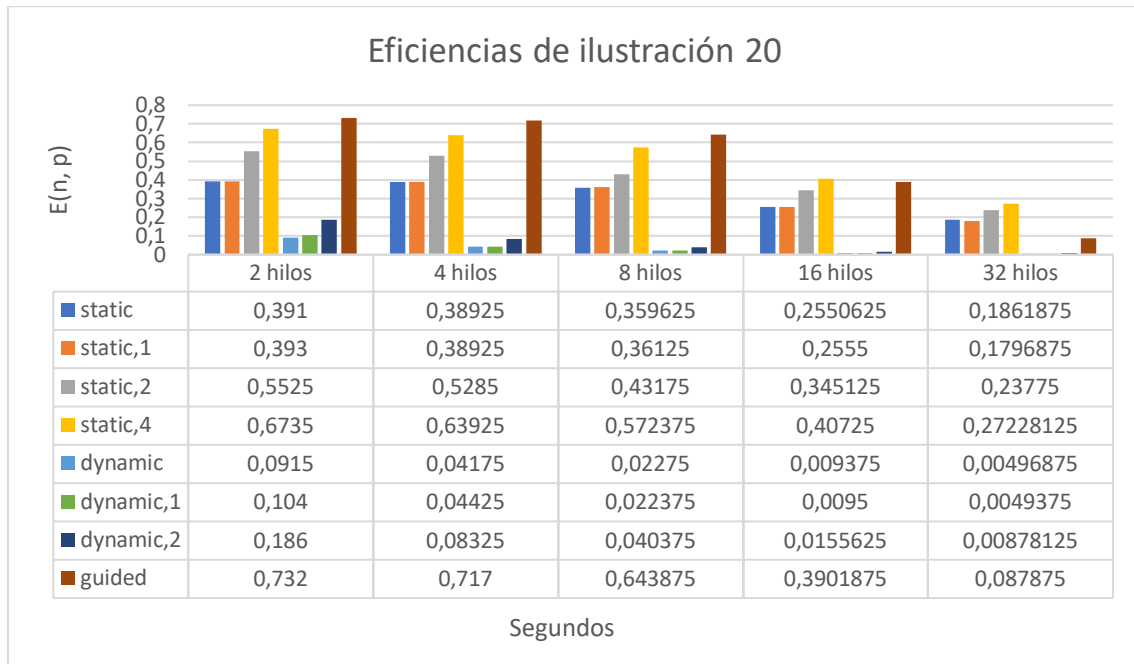


Ilustración 21

Y las eficiencias tendrán la misma proporción que las aceleraciones anteriores.

4.4 Conclusiones

Como conclusión, podemos acabar comparando la mejor planificación para *planets2i* completamente planificado con la mejor planificación para *planets2j* completamente planificado.

Haciendo esto, saldrían las siguientes gráficas:

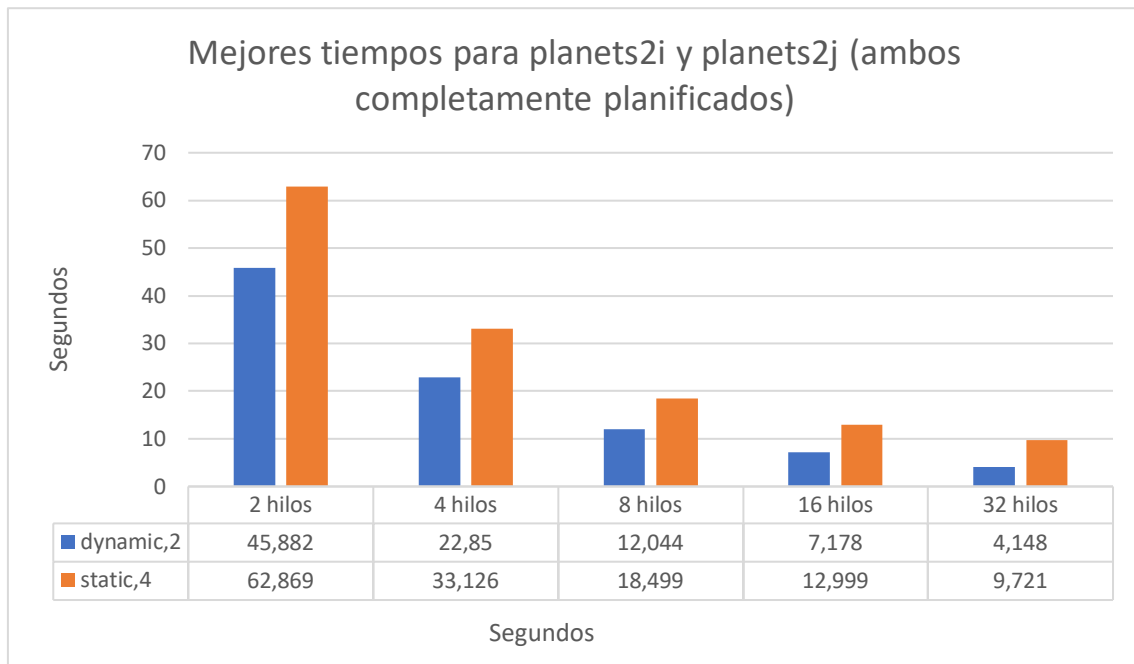


Ilustración 22

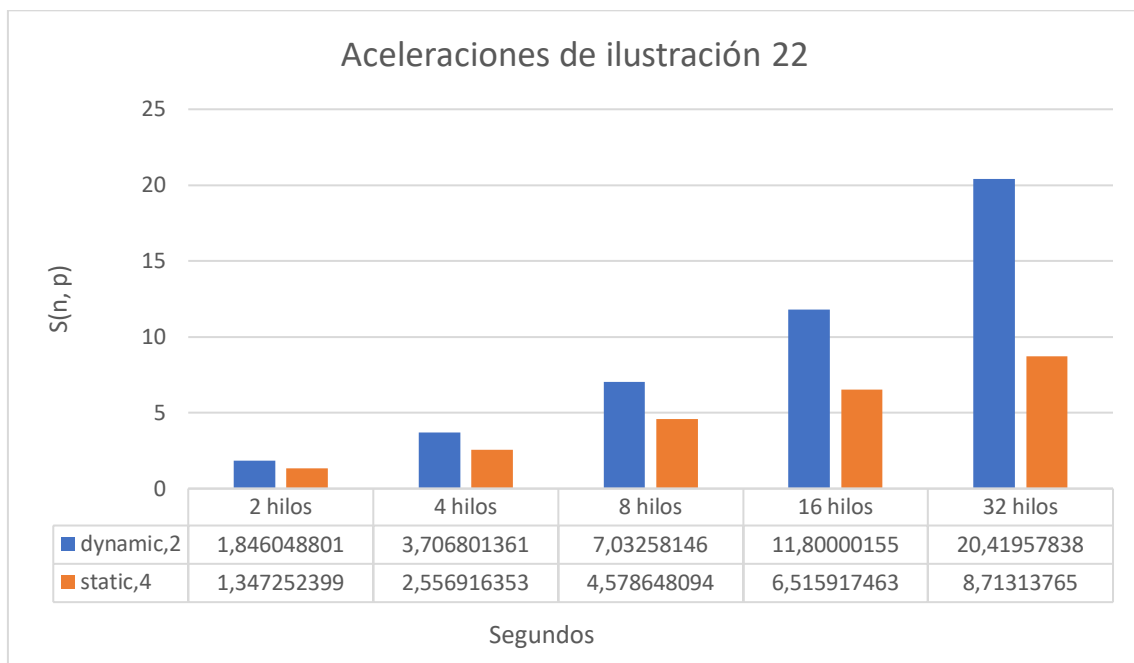


Ilustración 23

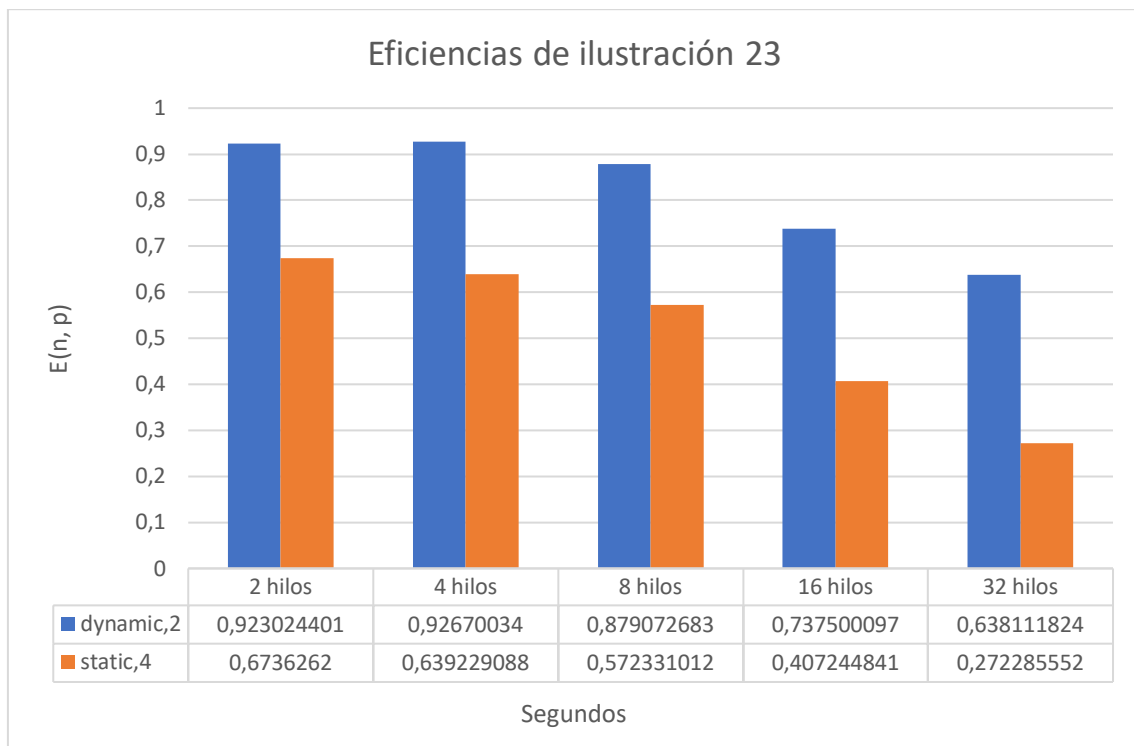


Ilustración 24

Para concluir, se puede observar, a través de las aceleraciones y las eficiencias, que *planets2i* es el programa más rápido para todos los casos (en comparación con *planets2j*), teniendo sentido, ya que, cuanto más externo sea el bucle paralelizado, más eficiente debería ser y, con este último gráfico, observamos que con 2 y 4 hilos es donde más eficiencia hay, y cuantos más hilos se utilicen, más eficiencia se pierde en ambos programas.

5. Ejercicio 5

En este ejercicio hemos creado varias variables auxiliares para poder mostrar las iteraciones procesadas por cada hilo, así como el centro de masa de los planetas que ha procesado cada hilo, además de una región paralela anterior al primer bucle *for*.

A continuación, indicamos en la siguiente fotografía las variables auxiliares necesarias (están incluidas las variables cm_x , cm_y , cm_z , a pesar de no tener relevancia).

```

int nP = 0;
double cm_X = 0;
double cm_Y = 0;
double cm_Z = 0;
m = 0;

double cm_X_1 = 0;
double cm_Y_1 = 0;
double cm_Z_1 = 0;
double m1 = 0;

```

Ilustración 25

La variable **nP** es el contador de planetas que procesa cada hilo, las variables **cm_{X1}**, **cm_{Y1}**, **cm_{Z1}**, son las variables que contendrán el centro de masa en las 3 dimensiones para cada hilo, y **m1** la suma de las masas de todos los planetas que procesa un hilo.

Para poder desarrollar este ejercicio hemos creado una región paralela como hemos mencionado antes, la cual abarca todo el bucle y acaba antes de las asignaciones del último *bucle i*.

```

#pragma omp parallel private(nP, cm_X_1, cm_Y_1, cm_Z_1, m1)
{
    nP = 0;

```

Ilustración 26

A continuación, mostramos cómo asignamos el valor a estas variables:

```

for (i = 0; i < n; i++)
{
    // Accumulate mass and weighted position to compute the center of mass
    m += p[i].m;
    cm_X += p[i].m * p[i].p.x;
    cm_Y += p[i].m * p[i].p.y;
    cm_Z += p[i].m * p[i].p.z;
    cm_X_1 = cm_X;
    cm_Y_1 = cm_Y;
    cm_Z_1 = cm_Z;
    m1 = m;

```

Ilustración 27

Todas las variables son privadas, ya que cada hilo mantiene un valor para cada variable, y ponemos la variable ***nP*** en cada iteración para que cada hilo comience con el valor asignado a 0, como mostramos en el siguiente fragmento de código:

```

    if (d < NGB_DST)
    {
        // Planets i and j are neighbors. Update number of neighbors
        // for both planets
        ngbi++;
        p[j].ngb++;
    }
}
p[i].ngb += ngbi;
nP++;
}

cm.x = cm_X;
cm.y = cm_Y;
cm.z = cm_Z;

// Finish computation of the center of mass
if (m > EPS)
{
    cm.x /= m;
    cm.y /= m;
    cm.z /= m;
}
if (m1 > EPS)
{
    cm_X_1 /= m1;
    cm_Y_1 /= m1;
    cm_Z_1 /= m1;
}
printf("The number of planets processed by Thread %d are: %d \n", omp_get_thread_num(), nP);
printf("The center of mass for Thread %d is: (%f, %f, %f)\n\n", omp_get_thread_num(), cm_X_1, cm_Y_1, cm_Z_1);
}

```

Ilustración 28

En cada interacción del bucle *i* se incrementa el contador de número de planetas para ese hilo. Luego hacemos la misma condición que para *m* con *m1* y dividimos las variables que contienen el centro de masa en las tres dimensiones por la masa de los planetas. Por último, mostramos por pantalla el número de planetas procesados por cada hilo, así como el centro de masa de los planetas procesados por dicho hilo.