Arquitectura de Computadores (AC)

2° curso / 2° cuatr. Grado Ing. Inform. Cuaderno de prácticas. Bloque Práctico 4. Optimización de código

Estudiante (nombre y apellidos): Javier Victoria Mohamed Grupo de prácticas y profesor de prácticas:

Fecha de entrega:

Fecha evaluación en clase:

Antes de comenzar a realizar el trabajo de este cuaderno consultar el fichero con los normas de prácticas que se encuentra en SWAD

Denominación de marca del chip de procesamiento o procesador (se encuentra en /proc/cpuinfo): (respuesta)

Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz

Sistema operativo utilizado: (respuesta)

Ubuntu 18.04.2 LTS

Versión de gcc utilizada: (respuesta) gcc (Ubuntu 7.4.0-1ubuntu1~18.04) 7.4.0

Volcado de pantalla que muestre lo que devuelve 1scpu en la máquina en la que ha tomado las medidas



- 1. Para el núcleo que se muestra en el Figura 1, y para un programa que implemente la multiplicación de matrices con datos flotantes en doble precisión (use variables globales):
 - 1.1 Modifique el código C para reducir el tiempo de ejecución (evalúe el tiempo y modifique sólo el trozo que hace la multiplicación y el trozo que se muestra en la Figura 1). Justifique los tiempos obtenidos (use -O2) a partir de la modificación realizada. Incorpore los códigos modificados en el cuaderno.
 - 1.2 Genere los códigos en ensamblador con -O2 para el original y dos códigos modificados obtenidos en el punto anterior (incluido el que supone menor tiempo de ejecución) e incorpórelos al cuaderno de prácticas. Destaque las diferencias entre ellos en el código ensamblador.
 - 1.3 (Ejercicio EXTRA) Intente mejorar los resultados obtenidos transformando el código ensamblador del programa para el que se han conseguido las mejores prestaciones de tiempo

Figura 1. Código C++ que suma dos vectores

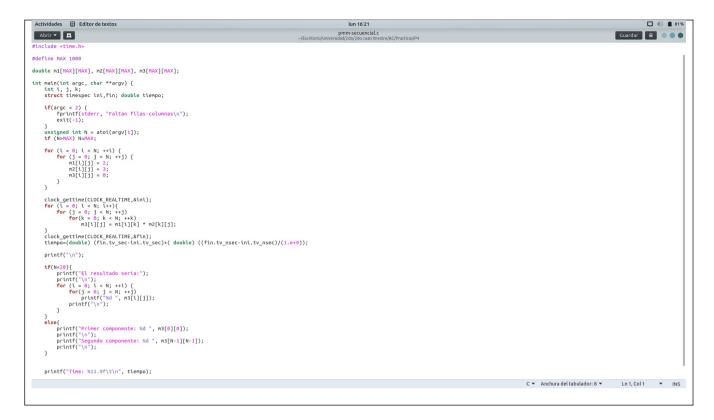
```
struct {
    int a;
    int b;
} s[5000];

main()
{
    ...
    for (ii=0; ii<40000;ii++) {
        X1=0; X2=0;
        for(i=0; i<5000;i++) X1+=2*s[i].a+ii;
        for(i=0; i<5000;i++) X2+=3*s[i].b-ii;

    if (X1<X2) R[ii]=X1 else R[ii]=X2;
    }
    ...
}</pre>
```

A) MULTIPLICACIÓN DE MATRICES:

CAPTURA CÓDIGO FUENTE: pmm-secuencial.c



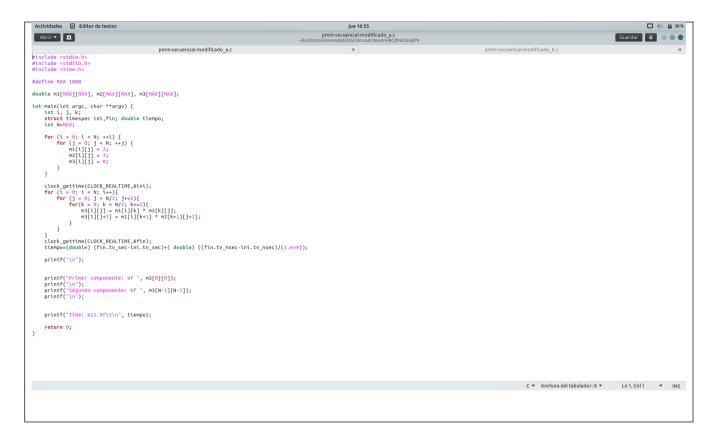
1.1. MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación a) –explicación-: Hacemos varios cálculos en cada iteración.

Modificación b) —**explicación-:** Invertimos los bucles j y k, con esto accedemos a los elementos de la matriz 2 más rápido porque sería de forma consecutiva en memoria pero en las otras 2 matrices se producen muchas más faltas de caché.

1.1. CÓDIGOS FUENTE MODIFICACIONES

a) Captura de pmm-secuencial-modificado_a.c



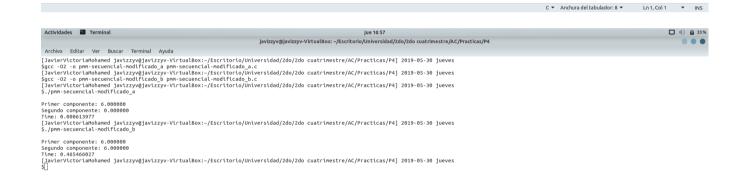
Capturas de pantalla (que muestren la compilación y que el resultado es correcto):

```
Archivo Editar Ver Buscar Terminal Ayuda

| Javiervitctort-alkohaned javitzzyvgjavitzzyv-VirtualBox:-/Escritorio/Universidad/Zdo/Zdo cuatrinestre/Ac/Practicas/P4| 2019-05-30 jueves
Sgcc - 0.2 - op pm-secuenctal-nodificado a. c
| Javiervitctort-alkohaned javitzzyvgjavitzzyv-VirtualBox:-/Escritorio/Universidad/Zdo/Zdo cuatrinestre/Ac/Practicas/P4| 2019-05-30 jueves
Sgcc - 0.2 - op pm-secuenctal-nodificado b. c
| Javiervitctort-alkohaned javitzzyvgjavitzzyv-VirtualBox:-/Escritorio/Universidad/Zdo/Zdo cuatrinestre/Ac/Practicas/P4| 2019-05-30 jueves
Sgcc - 0.2 - op pm-secuenctal-nodificado b. c
| Javiervitctort-alkohaned javitzzyvgjavitzzyv-VirtualBox:-/Escritorio/Universidad/Zdo/Zdo cuatrinestre/Ac/Practicas/P4| 2019-05-30 jueves
S-pm-secuenctal-nodificado a

Primer componente: 0.000000
Segundo componente: 0.000000
Segun
```

b) ...



1.1. TIEMPOS:

1.1. 11EWH 05.		
Modificación	-O2	
Sin modificar		0,01-0,02s
Modificación a)		0,0006-0,0007s
Modificación b)		0,41-0,52s

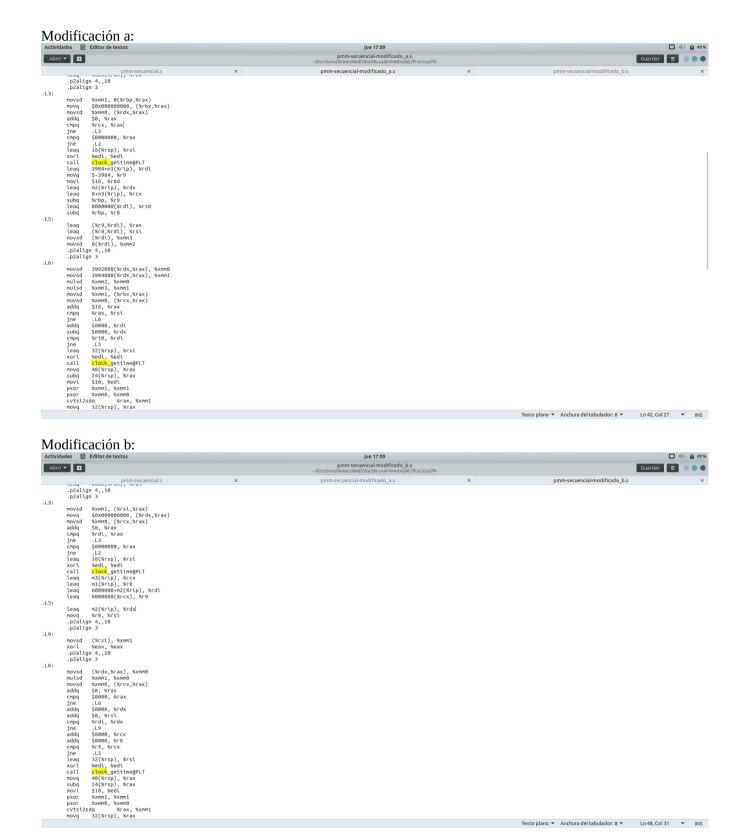
1.1. COMENTARIOS SOBRE LOS RESULTADOS Y JUSTIFICACIÓN DE LAS MEJORAS EN TIEMPO:

Envidentemente la modificación b no nos compensa por lo ya comentado anteriormente, pero la modificación a sí que nos compensa y mucho ya que nos saltamos la mitad de la iteraciones de 2 de los bucles y eso se nota un montón en el tiempo obtenido.

1.2. CÓDIGOS ENSAMBLADOR:

Original:

Activitation | Getter features | Description | Getter features | Get

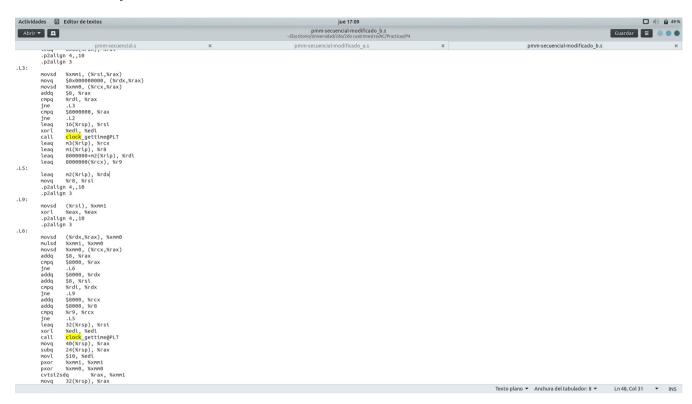


Vemos como del original y el a al b se añaden y sustituyen 2 etiquetas que harían el trabajo de los bucles intercambiados mencionados antes. En cuanto a la diferencia entre el original y el a es simplemente el tamaño de la etiqueta L3 debido a que ahora el bucle hace más operaciones, pero ahorramos posibles fallos por predicción de saltos.

Texto plano ▼ Anchura del tabulador: 8 ▼ Ln 48, Col 31 ▼ INS

1.3. MEJORA EN ENSAMBLADOR:

Partimos de la mejora a:



Aumentamos más aún el número de iteraciones y las variables que guardamos en pila para poder hacer eso mismo, como consecuencia obtenemos un tiempo promedio de 0,0003 mucho mejor que la mejora a.

B) CÓDIGO FIGURA 1:

CAPTURA CÓDIGO FUENTE: figura1-original.c



1.1. MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación a) –explicación-: El más obvio sería agrupar los 2 bucles para ahorrar tiempo y no tener que hacer dos bucles iguales para cosas diferentes.

Modificación b) -explicación-: Hacer el doble de acciones cada iteración.

1.1. CÓDIGOS FUENTE MODIFICACIONES

a) Captura figura1-modificado_a.c

```
| Anthon of the state of the st
```

Capturas de pantalla (que muestren la compilación y que el resultado es correcto):



b) ...



1.1. TIEMPOS:

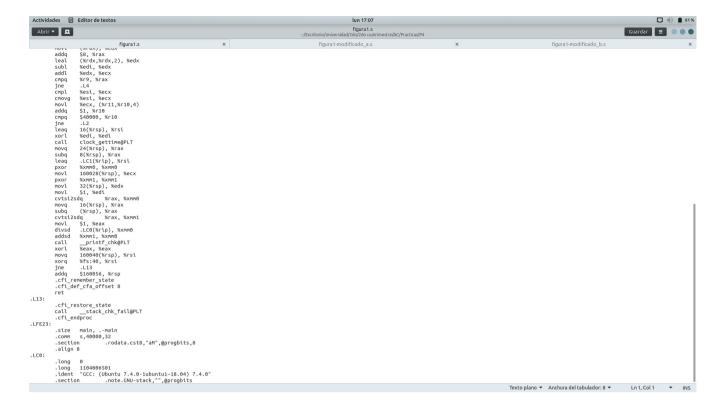
Modificación	Breve descripción de las modificaciones	-O2
Sin modificar	Original	0,31–0,32s
Modificación a)	Bucles fusionados	0,19–0,20s
Modificación b)	Dos cálculos por bucle	0,18-0,19s
•••		

1.1. COMENTARIOS SOBRE LOS RESULTADOS Y JUSTIFICACIÓN DE LAS MEJORAS EN TIEMPO:

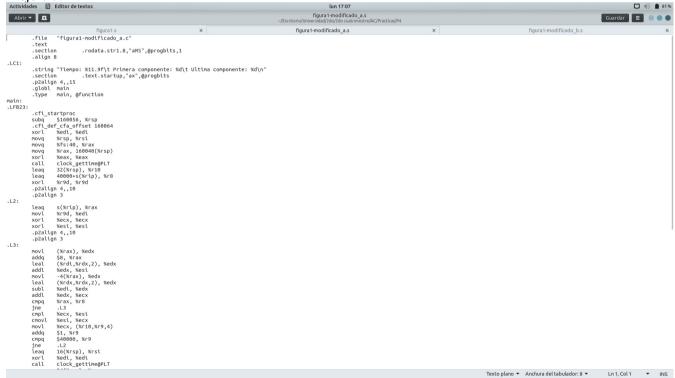
Vemos como en la primera mejora se nota un montón la bajada en el tiempo ya que nos estamos ahorrando un buecle completo y que esto se mejora un poco en la segunda debido a que hacemos la mitad de iteraciones, sin embargo a más cálculos por iteración menos tiempos tendríamos seguramente (dentro de ciertas cotas)

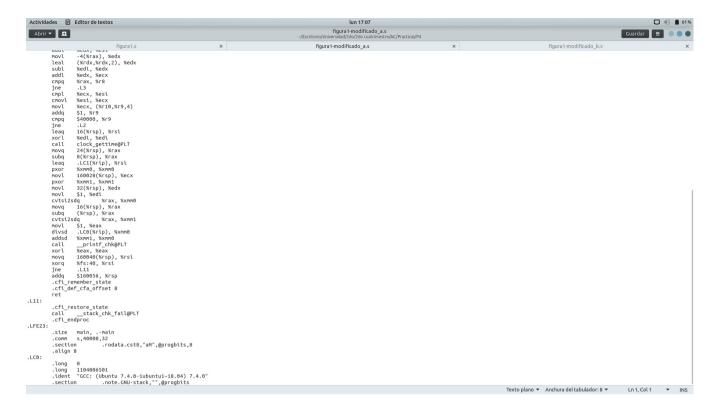
1.2. CÓDIGOS ENSAMBLADOR

Original:

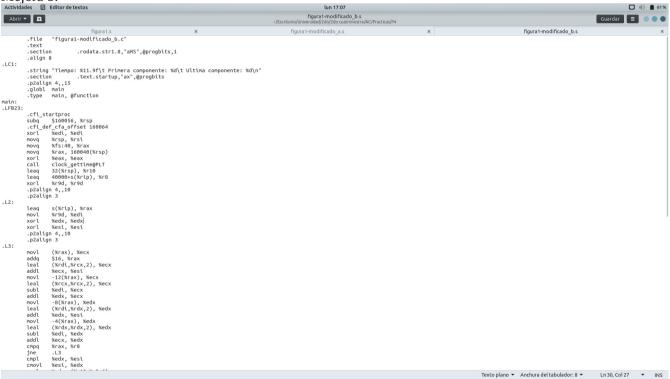


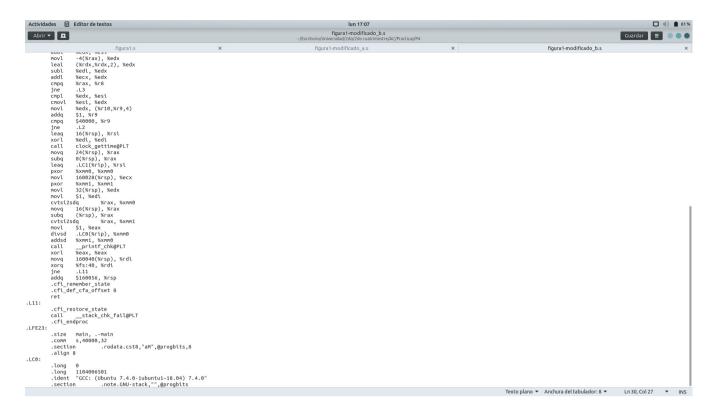
Mejora a:





Mejora b:





Respuesta:

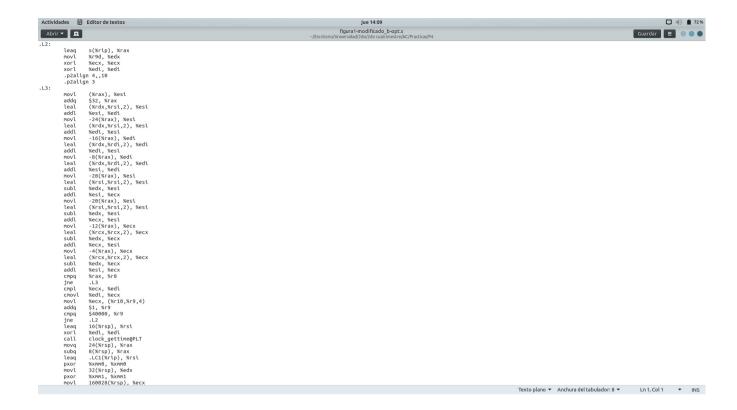
La diferencia entre la original y la primera mejora es básicamente que la etiqueta L3 y L4 se han fusionado y no hay dos saltos parecidos, esto mejora la eficiencia. En cuanto a la diferencia entre la primera mejora y la segunda sería que la etiqueta L3 es más grande ya que hace lo que haría para 2 iteraciones seguidas pero saltándose saltos innecesarios.

1.3. MEJORA EN ENSAMBLADOR

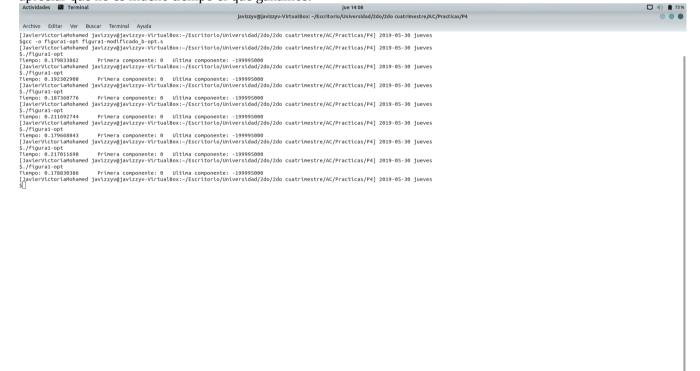
Partiremos de la versión 'b'.

En vez de saltos podríamos usar instrucciones predicadas sin embargo la sintaxis es muy ambigua y depende del compilador así que haremos una optimización menos eficiente pero mejor que la anterior.

Para ello en la parte de declaración de variables declararemos el doble (en pila habrá hasta 32 bits de variables cuando antes teníamos 16) y gracias a esto podremos hacer el doble de iteraciones antes de llegar al salto por lo que perderemos menos ciclos en caso de que la predicción de saltos sea errónea, el código ensamblador que se ha cambiado quedaría tal que así:



De esta forma no perderemos tantos ciclos por los saltos especulativos y ganaremos eficiencia aunque podemos apreciar que no es mucho tiempo el que ganamos.



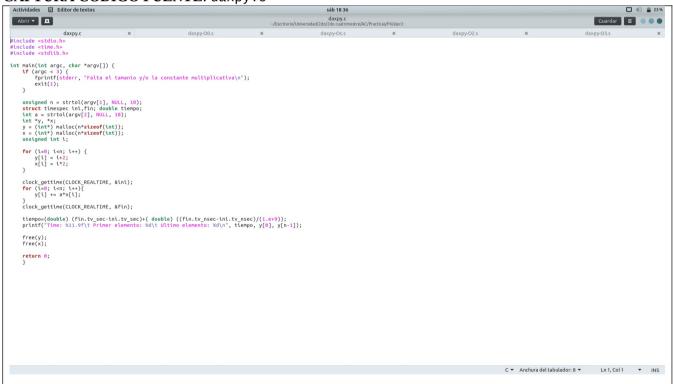
Vemos como ahora obtenemos una media de 0,17 segundos lo cual sería un poco más eficiente.

2. El benchmark Linpack ha sido uno de los programas más ampliamente utilizados para evaluar las prestaciones de los computadores. De hecho, se utiliza como base en la lista de los 500 computadores más rápidos del mundo (el Top500 Report). El núcleo de este programa es una rutina que opera con flotantes de doble precisión denominada DAXPY (*Double precision- real Alpha X Plus Y*) que multiplica un vector por una constante y los suma a otro vector (Lección 3/Tema 1):

for
$$(i=1; i \le N, i++)$$
 $y[i] = a*x[i] + y[i];$

- 2.1. Genere los programas en ensamblador para cada una de las siguientes opciones de optimización del compilador: -O0, -Os, -O2, -O3. Explique las diferencias que se observan en el código justificando al mismo tiempo las mejoras en velocidad que acarrean. Incorpore los códigos al cuaderno de prácticas y destaque las diferencias entre ellos. Sólo se debe evaluar el tiempo del núcleo DAXPY
- 2.2. (Ejercicio EXTRA) Para la mejor de las opciones, obtenga los tiempos de ejecución con distintos valores de N y determine para su sistema los valores de Rmax (valor máximo del número de operaciones en coma flotante por unidad de tiempo), Nmax (valor de N para el que se consigue Rmax), y N1/2 (valor de N para el que se obtiene Rmax/2). Estime el valor de la velocidad pico (Rpico) del procesador y compárela con el valor obtenido para Rmax. -Consulte la Lección 3 del Tema 1.

CAPTURA CÓDIGO FUENTE: daxpy.c



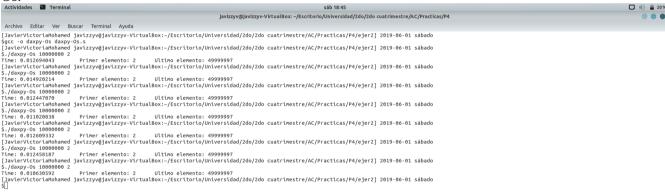
	-O0	-Os	-O2	-O3
Tiempos ejec.	0,038-0,039s	0,010-	0,010-	0,0075-
		0,014s	0,011s	0,008s

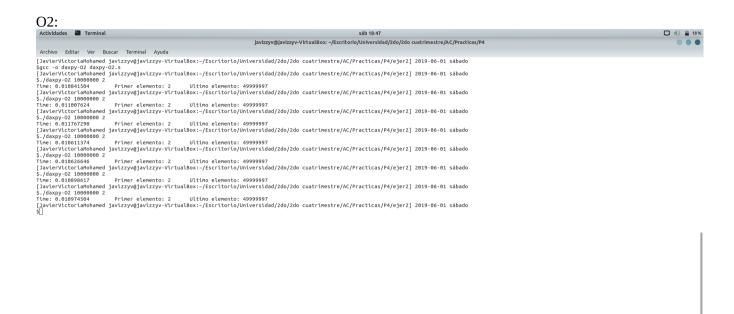
CAPTURAS DE PANTALLA (que muestren la compilación y que el resultado es correcto):

O0:



Os:







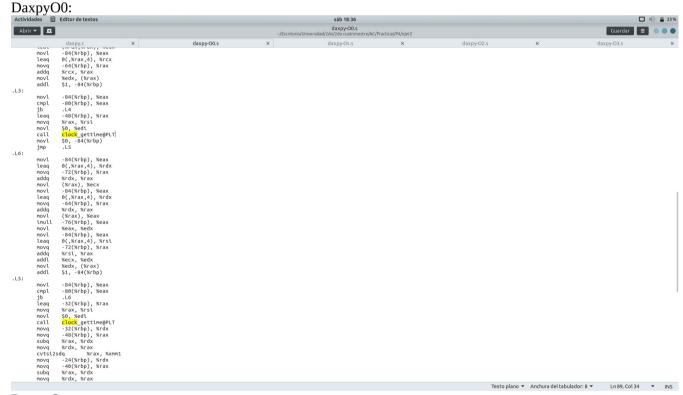
COMENTARIOS QUE EXPLIQUEN LAS DIFERENCIAS EN ENSAMBLADOR:

O0 usa la pila, rbp, para gestionar las variables mientras que el resto usa rsp, un registro de la arquitectura, esto hace que el resto gane eficiencia en cuanto a tiempo frente a O0. La diferencia fundamental entre Os y O2 es que Os optimiza en cuanto a espacio por lo que pierde algunas milésimas de segundo frente a O2. O3 realiza un desenrollado de bucles por lo que gana un montón de eficiencia como se puede ver en el tiempo (siendo cerca de ½ del tiempo de O2).

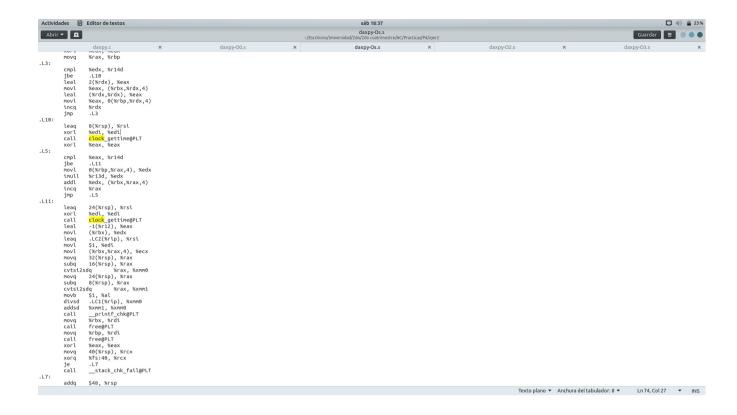
CÓDIGO EN ENSAMBLADOR (no es necesario introducir aquí el código como captura de pantalla, ajustar el tamaño de la letra para que una instrucción no ocupe más de un renglón):

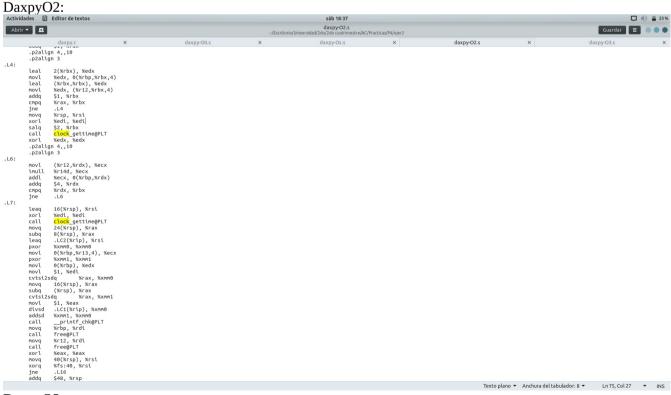
(PONER AQUÍ SÓLO LA ZONA DEL CÓDIGO ENSAMBLADOR DONDE ESTÁ EL CÓDIGO EVALUADO, USE COLORES PARA DESTACAR LAS DIFERENCIAS)

daxpy00.s	daxpy0s.s	daxpy02.s	daxpy03.s

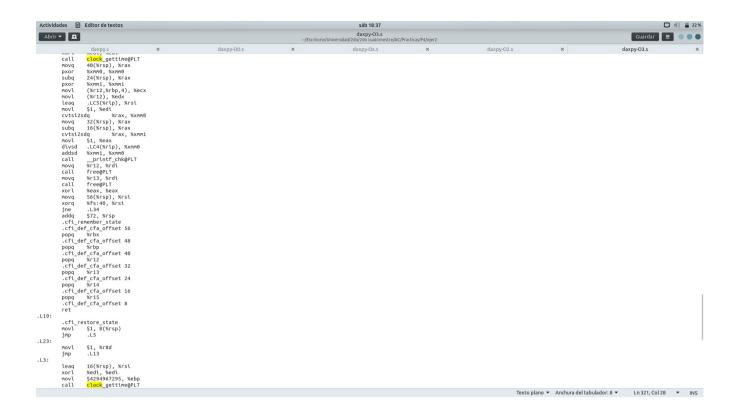


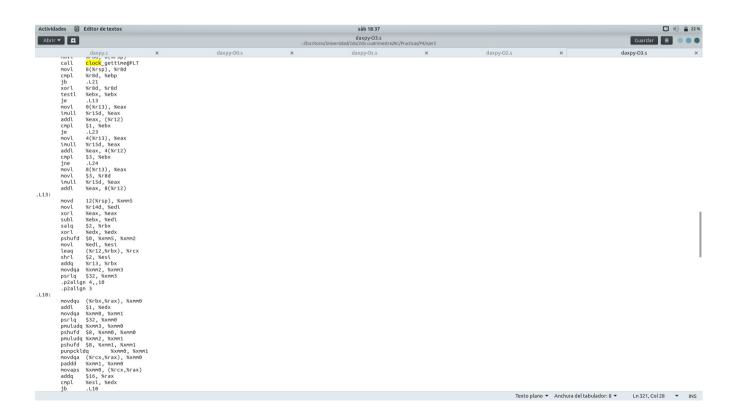
DaxpyOs:





DaxpyO3:





2.2.

Usaremos la versión de -O3.

Nmax = 100000 Rmax = 2*100000/0,0004*10⁶) = 500 MFLOPS

Nn/2 = 1000000

 $Rn/2 = 2*1000000/0,0007*10^{6}) = 285 MFLOPS$

Rpico=2073 (calculado gracias a la página hpl-calculator y las especificaciones del procesador se han mirado en userBenchmark).

Obtenemos que el Rpico es muy superior al Rmax para este programa por lo que aún podríamos exprimir el procesador muchos más.