2° curso / 2° cuatr.

Grado Ing. Inform.

Doble Grado Ing. Inform. y Mat.

### Arquitectura de Computadores (AC)

Cuaderno de prácticas. Bloque Práctico 4. Optimización de código

Estudiante (nombre y apellidos): César Muñoz Reinoso

Grupo de prácticas: Grupo 2

Fecha de entrega: 30/05 Fecha evaluación en clase:

**Denominación de marca del chip de procesamiento o procesador (se encuentra en /proc/cpuinfo):** Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz

Sistema operativo utilizado: Ubuntu 18.04

Versión de gcc utilizada: gcc 7.3.0

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 (use variables globales):
  - 1.1 Modifique el código C para reducir el tiempo de ejecución del mismo. 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 {

#### A) MULTIPLICACIÓN DE MATRICES:

CAPTURA CÓDIGO FUENTE: pmm-secuencial.c

```
#include <stdio.h>
#include <time.h>
int matrix1[N][N],matrix2[N][N],salida[N][N];
int main(int argc, char** argv){
        int suma=0;
        struct timespec cgt1, cgt2; double ncgt;
         //INICIALIZAMOS MATRIZ Y VECTOR
        for(int i = 0;i < N;i++){</pre>
                for(int j = 0; j < N; j++){
                                  matrix1[i][j]=i+j;
                                  matrix2[i][j]=i+j;
        clock gettime(CLOCK REALTIME,&cgt1);
        //MULTIPLICACIÓN
        for(int i = 0; i < N; i++){
                 for(int j = 0; j < N; j++){
    for(int z = 0; z < N; z++){</pre>
                                  salida[i][j] += matrix1[i][z]*matrix2[z][j];
        clock gettime(CLOCK REALTIME,&cgt2);
        ncgt=(double) (cgt2.tv_sec-cgt1.tv_sec) + (double) ((cgt2.tv_nsec-cgt1.tv_nsec)/(1.e+9));
         printf("\nTiempo(seg.): \$11.9f\n Tamaño: \$u\n salida[0][0] = \$d \ / \ / \ salida[\$d][\$d] = \$d \ \n", 
ncgt,N,salida[0][0],N-1,N-1,salida[N-1][N-1]);
        return 0:
}
```

#### 1.1. MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación a) –explicación-: Desenrollado de bucle

**Modificación b) –explicación-:** Transposición de la segunda matriz para acceder a ella de una forma secuencial en memoria.

#### 1.1. CÓDIGOS FUENTE MODIFICACIONES

a) Captura de pmm-secuencial-modificado\_a.c

```
int main(int argc, char** argv){
        int suma=0;
        struct timespec cgt1, cgt2; double ncgt;
        //INICIALIZAMOS MATRIZ Y VECTOR
        for(int i = 0; i < N; i++){
                for(int j = 0; j < N; j++){
                                 matrix1[i][j]=i+j;
                                 matrix2[i][j]=i+j;
                }
        }
        clock gettime(CLOCK REALTIME,&cgt1);
        //MULTIPLICACIÓN
int tmp0=0, tmp1=0, tmp2=0, tmp3=0;
        for(int i = 0; i < N; i++){</pre>
                for(int j = 0; j < N; j++){
                         for(int z = 0; z < N; z+=4){
                                 tmp0 += matrix1[i][z]*matrix2[z][j];
                                 tmp1 += matrix1[i][z+1]*matrix2[z+1][j];
                                 tmp2 += matrix1[i][z+2]*matrix2[z+2][j];
                                 tmp3 += matrix1[i][z+3]*matrix2[z+3][j];
                         salida[i][j]=tmp0+tmp1+tmp2+tmp3;
                         tmp0=tmp1=tmp2=tmp3=0;
        clock gettime(CLOCK REALTIME,&cgt2);
```

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

```
[César Muñoz Reinoso cesar@cesar-TM1701:-/Escritorio/UNIVERSIDAD/SEGUNDO/2° Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes 
5gcc -02 pmm-secuencial-modificado_a.c -o pmm-secuencial-modificado_a 
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2° Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes 
$./pmm-secuencial-modificado_a 
Tiempo(seg.):0.212987135 
Tamaño:512 
salida[0][0]=44608256 / / salida[511][511]=311996160
```

b) Captura de pmm-secuencial-modificado\_b.c

```
int main(int argc, char** argv){
        int suma=0;
        struct timespec cqt1, cqt2; double ncqt;
        //INICIALIZAMOS MATRIZ Y VECTOR
        for(int i = 0; i < N; i++){
                for(int j = 0; j < N; j++){
                                 matrix1[i][j]=i+j;
                                 matrix2[i][j]=i+j;
                }
        }
        clock gettime(CLOCK REALTIME,&cgt1);
        //TRASPUESTA
        for(int i=0;i<N;i++){</pre>
                for(int j=0; j<N; j++){</pre>
                         int aux=matrix2[i][j];
                                 matrix2[i][j]=matrix2[j][i];
                                 matrix2[j][i]=aux;
                }
        }
        //MULTIPLICACIÓN
        for(int i = 0; i < N; i++){</pre>
                for(int j = 0; j < N; j++){
                         for(int z = 0; z < N; z++){
                                 salida[i][j] += matrix1[i][z]*matrix2[j][z];
                }
        }
        clock gettime(CLOCK REALTIME,&cgt2);
```

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

```
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2° Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes 

$gcc -02 pmm-secuencial-modificado_b.c -o pmm-secuencial-modificado_b
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2° Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes
$./pmm-secuencial-modificado_b

Tiempo(seg.):0.070131735

Tamaño:512

salida[0][0]=44608256 / / salida[511][511]=311996160
```

#### **1.1. TIEMPOS:**

Modificación	-O2
Sin modificar	0.236325213
Modificación a)	0.212987135
Modificación b)	0.070131735

#### 1.1. COMENTARIOS SOBRE LOS RESULTADOS:

En cuanto a la ejecución de los diferentes programas, podemos observar que la primera modificación de desenrollado de bucles tan solo reduce el tiempo inicial en 0,02 segundos. En cambio, la segunda modificación de transposición de la segunda matriz es mucho más efectiva, reduciendo en 0,16 segundos la ejecución.

# 1.2. CÓDIGO EN ENSAMBLADOR DEL ORIGINAL Y DE DOS MODIFICACIONES : (PONER AQUÍ SÓLO LA ZONA DEL CÓDIGO ENSAMBLADOR EVALUADA, USE COLORES PARA DESTACAR LAS DIFERENCIAS)

pmm-secuencial.s	pmm-secuencial-modificado_a.s	pmm-secuencial-modificado_b.s
Leal (%rcx,%rax), %edx   mov1 edx, (%rs1,%rax,4)   mov1 edx, (%rd1,%rax,4)   addq   \$1, %rax   cmpq   \$512, %rax   jne   131, %ecx   add1   \$2048, %rsi   addg   \$2048, %rsi   addg   \$2048, %rsi   cmpq   \$512, %ecx   jne   L2   movq   %rsp, %rsi   xori   %edx   wettine@PLT   cmpq   for   for	Leal (%rcx,%rax), %edx	Lal

#### **B) CÓDIGO FIGURA 1:**

#### CAPTURA CÓDIGO FUENTE: figura1-original.c

```
#include <stdlib.h>
#include <stdio.h>
#include <time.h>
struct{
 int a:
 int b;
} s[5000];
int R[40000];
int main(int argc,char ** argv){
  int ii,i,X1,X2;
  struct timespec cgt1, cgt2; double ncgt;
  for(int i=0;i<5000;i++){</pre>
    s[i].a=i;
    s[i].b=i+1;
  clock gettime(CLOCK REALTIME,&cgt1);
  for(ii=0;ii<40000;ii++){</pre>
    X1=0; X2=0;
    for(i=0;i<5000;i++) X1+=2*s[i].a+ii;</pre>
    for(i=0;i<5000;i++) X2+=3*s[i].b-ii;
    if(X1<X2) R[ii]=X1; else R[ii]=X2;</pre>
  clock gettime(CLOCK REALTIME,&cgt2);
  ncgt=(double) (cgt2.tv_sec-cgt1.tv_sec) + (double) ((cgt2.tv_nsec-cgt1.tv_nsec)/(1.e+9));
  printf("\nTiempo(seg.):%11.9f\n R[0]=%d / / R[39999]=%d \n", ncgt,R[0],R[39999]);
  return 0;
```

#### 1.1. MODIFICACIONES REALIZADAS (al menos dos modificaciones):

**Modificación a) –explicación-:** Unimos los dos bucles ya que los datos se almacenan secuencialmente, teniendo en cada iteración un valor para a y otro para b unidos en memoria.

**Modificación b) –explicación-:** Desenrollamos los bucles para romper la dependencia de datos en cada iteración y así poder unificar el procesamiento de datos, reduciendo el número de iteraciones. También va incluida la modificación a.

#### 1.1. CÓDIGOS FUENTE MODIFICACIONES

a) Captura figura1-modificado\_a.c

```
#include <stdlib.h>
#include <stdio.h>
#include <time.h>
struct{
  int a;
  int b;
} s[5000];
int R[40000];
int main(int argc,char ** argv){
  int ii,i,X1,X2;
  struct timespec cgt1, cgt2; double ncgt;
  for(int i=0;i<5000;i++){</pre>
    s[i].a=i;
   s[i].b=i+1;
  clock gettime(CLOCK REALTIME,&cgt1);
  for(ii=0;ii<40000;ii++){</pre>
   X1=0; X2=0;
    for(i=0;i<5000;i++){
     X1+=2*s[i].a+ii;
     X2+=3*s[i].b-ii;
   if(X1<X2){
     R[ii]=X1;
    }else{
     R[ii]=X2;
  }
  clock_gettime(CLOCK_REALTIME,&cgt2);
  ncgt=(double) (cgt2.tv sec-cgt1.tv sec) + (double) ((cgt2.tv nsec-cgt1.tv nsec)/(1.e+9));
  printf("\nTiempo(seq.):%11.9f\n R[0]=%d / / R[39999]=%d \n", ncqt,R[0],R[39999]);
  return 0;
}
```

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

```
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes
Sgcc -02 figura1-modificado_a.c -o figura1-modificado_a
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes
S./figura1-modificado_a
Tiempo(seg.):0.143247732
R[0]=24995000 / / R[39999]=-162487500
```

#### b) Captura figura1-modificado\_b.c

```
int main(int argc,char ** argv){
  int ii,i,X1,X2,tmp0,tmp1,tmp2,tmp3;
  struct timespec cgt1, cgt2; double ncgt;
  for(int i=0;i<5000;i++){</pre>
    s[i].a=i:
   s[i].b=i+1;
  clock gettime(CLOCK REALTIME,&cgt1);
  for(ii=0;ii<40000;ii++){</pre>
    X1=0; X2=0, tmp0=0; tmp1=0; tmp2=0; tmp3=0;
    for(i=0;i<5000;i+=2){
      tmp0+=2*s[i].a+ii;
      tmp1+=2*s[i+1].a+ii;
      tmp2+=3*s[i].b-ii;
      tmp3+=3*s[i+1].b-ii;
    X1=tmp0+tmp1;
    X2=tmp2+tmp3;
    if(X1<X2){
     R[ii]=X1;
    }else{
      R[ii]=X2;
  clock gettime(CLOCK REALTIME,&cgt2);
  ncgt=(double) (cgt2.tv sec-cgt1.tv sec) + (double) ((cgt2.tv nsec-cgt1.tv nsec)/(1.e+9));
```

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

```
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -02 figura1-modificado_b [César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes $./ftgura1-modificado_b
Tiempo(seg.):0.122902374
R[0]=24995000 / / R[39999]=-162487500
```

#### 1.1. TIEMPOS:

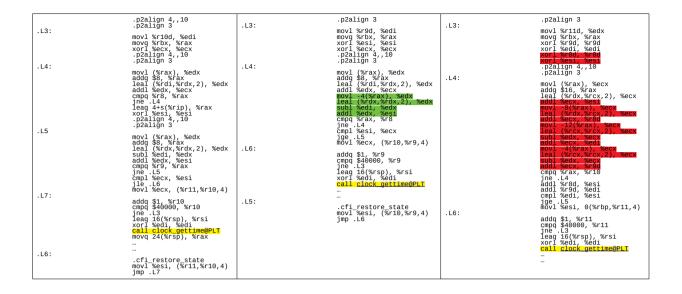
Modificación	-02	
Sin modificar	0.213954227	
Modificación a)	0.143247732	
Modificación b)	0.122902374	

#### 1.1. COMENTARIOS SOBRE LOS RESULTADOS:

En este caso tenemos un tiempo de referencia un poco menor. Con la primera modificación, unión de dos simples bucles, conseguimos reducir el tiempo de ejecución 0,7 segundos. Con la segunda modificación, desenrollando los bucles, conseguimos reducir el tiempo 0,9 segundos desde el inicial.

# 1.2. CÓDIGO EN ENSAMBLADOR DEL ORIGINAL Y DE DOS MODIFICACIONES: (PONER AQUÍ SÓLO LA ZONA DEL CÓDIGO ENSAMBLADOR EVALUADA, USE COLORES PARA DESTACAR LAS DIFERENCIAS)

figura1-original.s	figura1-modificado_a.s	figura1-modificado_b.s	
.L2:  mov1 %eax, (%rdx) add1 \$1, %eax add2 \$8, %rdx add3 \$8, %rdx mov1 %eax, -4(%rdx) cmp1 \$5000, %eax jne .L2 movg %rsp, %rsi xor1 %ed1, %edi call clock gettime0PLT lead 40004(%rbx), %r9 lead (%rsip), %fine xor1 %ri0d, %r1od, %r8	.L2:  movl %eax, (%rdx) addl \$1, %eax addg \$8, %rdx addg \$8, %rdx movl %eax, -4(%rdx) cml \$5000, %eax jne .L2 movg %rsp, %rsi xorl %edi, %edi call clock gettime@PLT leaa R(%rip), %ri0 lead 40000 (%rbx), %r8 xorl %r9d, %r9d .p2align 4,,10	.L2:  movl %eax (%rdx) addl \$1, %ax addl \$2, %rdx addl \$3, %rdx movl %eax -4(%rdx) cmpl \$5000, %eax jne .L2 movg %edi, %edi leaq R(%rip), %rsi xorl %edi, %edi leaq R(%rip), %rsp call clock gettime@PLT leaq 40000(%rbx), %ri0 xorl %riid, %riid .p2align 4,,10	



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 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.
- 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 (consulte en [4] el número de ciclos por instrucción punto flotante para la familia y modelo de procesador que está utilizando) y compárela con el valor obtenido para Rmax. -Consulte la Lección 3 del Tema 1.

CAPTURA CÓDIGO FUENTE: daxpy.c

```
#include <stdlib.h>
#include <stdio.h>
#include <time.h>
int main(int argc,char ** argv){
                000,a=1000;
  int N=520
  struct timespec cgt1, cgt2; double ncgt;
  double y[N];
  double x[N];
  for(int i=0; i<N;i++){</pre>
        y[i]=i+1;
        x[i]=i+2;
}
    clock_gettime(CLOCK_REALTIME,&cgt1);
for (int i=0;i<N;i++) y[i]= a*x[i] + y[i];</pre>
  clock_gettime(CLOCK_REALTIME,&cgt2);
  ncgt=(double) (cgt2.tv_sec-cgt1.tv_sec) + (double) ((cgt2.tv_nsec-cgt1.tv_nsec)/(1.e+9));
   printf("\nTamano:%d \nTiempo(seg.):%11.9f\n y[0]=%f / y[%d]=%f \n", N,ncgt,y[0],N,y[N-1]); 
  return 0;
```

Tiempos ejec.	-O0	-Os	-O2	-O3
	0.002784	0.00057	0.000706	0.000633
	702	0547	274	818

Tamaño: 520000

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

```
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -00 daxpy.c -o daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes S./daxpy

Tamano:520000
Tiempo(seg.):0.002784702
y[0]=2001.000000 / / [S20000]=520521000.000000
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -03 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes S./daxpy

Tamano:520000
Tiempo(seg.):0.0000570547
y[0]=2001.000000 / / [S20000]=520521000.000000
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -02 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -02 daxpy.c -0 daxpy

Tamano:520000
Tiempo(seg.):0.000706274
y[0]=2001.000000 / / [S20000]=520521000.000000
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -03 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -03 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -03 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -03 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escritorio/UNIVERSIDAD/SEGUNDO/2º Cuatrimestre/AC/Practicas/Archivos/BP4] 2018-05-25 viernes Sgcc -03 daxpy.c -0 daxpy
[César Muñoz Reinoso cesar@cesar-TM1701:~/Escri
```

#### COMENTARIOS QUE EXPLIQUEN LAS DIFERENCIAS EN ENSAMBLADOR:

Vemos que la mayor eficiencia del código no se obtiene con la optimización -O3 como seria lo normal, si no que se obtiene con la optimización -Os, ya que se obtiene un código ensamblador con menos instrucciones y por tanto de menor duración.

En daxpyOs.s vemos que el código es mucho menor que el original, donde tenemos un código anidado a una función, en el optimizado vemos que se resume en apenas 10 instrucciones.

Por otro lado vemos que el código daxpyO2.s es prácticamente igual que la optimización Os por eso lo tiempos son tan similares.

Finalmente podemos ver que la optimización O3 es la menos eficiente, ya que no esta definitivamente preparado para alcanzar ese nivel de optimización.

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)



