Arquitectura de Computadores (AC)

2° curso / 2° cuatr. Grado Ing. Inform. Cuaderno de prácticas. Bloque Práctico 5. Optimización de código

Estudiante (nombre y apellidos): Elena Ortiz Moreno

Grupo de prácticas y profesor de prácticas: B2 Niceto Luque

Fecha de entrega: 27/05/21

Fecha evaluación en clase: 28/05/21

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):

Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz

Sistema operativo utilizado: *Ubuntu 18.04.5 LTS*

Versión de gcc utilizada: gcc version 7.5.0 (Ubuntu 7.5.0-3ubuntu1~18.04)

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

```
elena@elena97om:/$ lscpu
Arquitectura:
Arquitectura:
modo(s) de operación de las CPUs:
Orden de los bytes:
                                                                                                                  32-bit, 64-bit
Little Endian
 Lista de la(s) CPU(s) en línea:
Hilo(s) de procesamiento por núcleo:
                                                                                                                  0-3
 Núcleo(s) por «socket»:
«Socket(s)»
 Modo(s) NUMA:
ID de fabricante:
Familia de CPU:
                                                                                                                  -
GenuineIntel
                                                                                                                  142
  lodelo:
   ombre del modelo:
                                                                                                                   Nombre del mo
Revisión:
CPU MHz:
CPU MHz máx.:
CPU MHz mín.:
                                                                                                                  3100,0000
 BogoMIPS:
Virtualización:
  aché Lld:
aché Lli:
aché L2:
   aché L3:
    PU(s) del nodo NUMA 0:
CPU(s) del nodo NUMA 0:

0-3
Indicadores:

fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse
36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art arch_perf
mon pebs bts rep_good nopl xtopology nonstop_tsc cpuid aperfmperf pni pclmulqdq dtes64 monitor ds_cpl vmx
est tm2 ssse3 sdbg fma cx16 xtpr pdcm pcid sse4 1 sse4 2 x2apic movbe popcnt tsc_deadline_timer aes xsave
avx f16c rdrand lahf_lm abm 3dnowprefetch cpuid_fault epb invpcid_single pti ssbd ibrs ibpb stibp tpr_shad
ow vnmi flexpriority ept vpid ept_ad fsgsbase tsc_adjust bmil avx2 smep bmi2 erms invpcid mpx rdseed adx s
map clflushopt intel_pt xsaveopt xsavec xgetbv1 xsaves dtherm ida arat pln pts hwp_notify hwp_act_wind
ow hwp_epp md_clear flush_lld
```

1. **(a)** Implementar un código secuencial que calcule la multiplicación de dos matrices cuadradas. Utilizar como base el código de suma de vectores de BPO. Los datos se deben generar de forma aleatoria para un número de filas mayor que 8, como en el ejemplo de BPO, se puede usar drand48()).

MULTIPLICACIÓN DE MATRICES:

CAPTURA CÓDIGO FUENTE: pmm-secuencial.c

```
#include <time.h>
#include <tof(1).hn-
#include <to
```

(b) Modificar el código (solo el trozo que calcula la multiplicación) para reducir el tiempo de ejecución. Justificar los tiempos obtenidos (usando siempre -O2) a partir de la modificación realizada. Incorporar los códigos modificados en el cuaderno.

MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación A) –**explicación-:** Desenrrollado de bulce: 4 operaciones por bucle.

Modificación B) – **explicación-:** Trasponer la matriz B para aprovechar la localidad de los accesos.

•••

CÓDIGOS FUENTE MODIFICACIONES

A) Captura de pmm-secuencial-modificado_A.c

```
printf("\nMatriz B:\n");
for(i = 0; i < n; i++){
    for(j = 0; j < n; j++){
        printf("\d", B[i][j]);
    }
    printf("\d", B[i][j]);
}

clock_gettime(CLOCK_REALTIME,&cgt1);

for(i = 0; i < n; i++) {
    for(j = 0; j < n; j++) {
        for(int k = 0; k < n; k += 4){
            C[i][j] += A[i][k] * B[k][j] + A[i][k+1] * B[k+1][j] + A[i][k+2] * B[k+2][j] + A[i][k+3] * B[k+3][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 matriz tamanio %d: %11.9f", n, ncgt);

if(n <= 11){
        printf("\nMatriz resultado:\n");
        for(j = 0; j < n; j++){
            printf("\d", c[i][j]);
        }
        printf("\d", c[i][j]);
    }
    printf("\n");
}

return 0;
}</pre>
```

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

```
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer1$ gcc -02 pmm-secuencial-modificado-a.c -o pmm-secuencial-modi
ficado-a
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer1$ ./pmm-secuencial-modificado-a 500
Tiempo matriz tamanio 500: 0.498274110elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer1$
```

B) Captura de pmm-secuencial-modificado_B.c

```
#define MAX 1024
 unsigned int n = atoi(argv[1]);
                                                                         printf("\nTiempo matriz tamanio %d: %11.9f", n, ncgt);
```

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

```
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejerl$ gcc -02 pmm-secuencial-modificado-b.c -o pmm-secuencial-modificado-b
ficado-b
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejerl$ ./pmm-secuencial-modificado-b 1000
Tiempo matriz tamanio 1000: 0.941819319elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejerl$ ■
```

TIEMPOS:

Modificación	Breve descripción de las modificaciones	-O2	
Sin modificar		0.700916746	
N = 500			
Modificación A)	Desenrrollado de bulce: 4 operaciones por bucle.	0.498274110	
N = 500			
Modificación B)	Trasponer la matriz B para aprovechar la localidad de	0.941819319	
N=1000	los accesos.		
•••			

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

Es mejor la modificación B que la A y la no modificada, debido a que solo con trasponer la matriz y por como se realiza la multiplicación de matrices aprovechamos mucho mejor la localidad de los accesos, mientras que en la primera modificación solo nos ahorramos un bucle.

2. **(a)** Usando como base el código de BPO, generar un programa para evaluar un código de la Figura 1. M y N deben ser parámetros de entrada al programa. Los datos se deben generar de forma aleatoria para valores de M y N mayores que 8, como en el ejemplo de BPO.

CÓDIGO FIGURA 1:

CAPTURA CÓDIGO FUENTE: figura1-original.c

```
#include <stdlib.h>
#define TAM R 40000
} s[TAM S];
int R[TAM_R];
  clock_gettime(CLOCK_REALTIME,&cgt1);
  for ( ii = 0; ii < TAM_R; ii++ ) {
    for(i=0; i < TAM S; i++) X1+=2*s[i].a+ii;
    for(i=0; i < TAM S; i++) X2+=3*s[i].b-ii;
  ncgt=(double) (cgt2.tv_sec-cgt1.tv_sec) + (double) ((cgt2.tv_nsec-cgt1.tv_nsec) / (1.e+9));
  printf("R[0]=%d R[39999]=%d\n", R[0], R[TAM_R-1]);
  printf("Tiempo: %11.9f\n", ncgt);
```

Figura 1. Código C++ que suma dos vectores. My N deben ser parámetros de entrada al programa, usar valores mayores que 1000 en la evaluación.

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

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

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

(b) Modificar el código C (solo el trozo a evaluar) para reducir el tiempo de ejecución. Justificar los tiempos obtenidos (usando siempre -O2) a partir de la modificación realizada. En las ejecuciones de evaluación usar valores de N y M mayores que 1000. Incorporar los códigos modificados en el cuader

MODIFICACIONES REALIZADAS (al menos dos modificaciones):

Modificación A) –explicación-: reducción de bucles y operador ternario.

Modificación B) –**explicación-:** desenrrollado de bucle y operador ternario.

•••

CÓDIGOS FUENTE MODIFICACIONES

A) Captura figura1-modificado_A.c

```
printf("R[0]=%d R[39999]=%d\n", R[0], R[TAM R-1]);
```

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

```
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer2$ gcc figural-modificado-a.c -o figural-modificado-a elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer2$ ./figural-modificado-a

Vector R:
R[0]=24995000 R[39999]=-125010000
Tiempo: 0.674179991
```

B) Captura figura1-modificado_B.c

```
#include <stdio.h>
#include <stdlib.h>
#include <time.h>
#define TAM_S 5000
#define TAM_R 40000
```

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

elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer2\$ gcc figural-modificado-b.c -o figural-modificado-b

elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer2\$./figural-modificado-b

Vector R:

R[0]=24995000 R[39999]=-125010000

Tiempo: 0.681195113

TIEMPOS:

Modificación	ificación Breve descripción de las modificaciones	
Sin modificar		1.197413569
Modificación A)	reducción de bucles y operador ternario.	0.674179991
Modificación B)	desenrrollado de bucle y operador ternario.	0.681195113

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

En este caso la mejor opción es la a: en ambas modificaciones juntamos todo dentro de un solo bucle for interno, pero funciona mejor llamar al struct el menor número de veces posible acumulando lso resultados que ahorrarnos bucles con múltiples llamadas dentro.

3. 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
$$(i=0; i< N; i++)$$
 $y[i]= a*x[i] + y[i];$

Generar 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. Incorporar los códigos al cuaderno de prácticas y destacar las diferencias entre ellos. Sólo se debe evaluar el tiempo del núcleo DAXPY. N deben ser parámetro de entrada al programa.

CAPTURA CÓDIGO FUENTE: daxpy.c

```
#include <stdio.h>
#include <stdlib.h>
#include <time.h>
int main (int argc, char **argv){
  struct timespec cgt1, cgt2; //Medicion de tiempo
 double ncgt;
  const int A = 150;  //Constante arbitraria
  if ( argc < 2 ) {
  int n = atoi(argv[1]);
  int x[n], y[n];
  for (int i = 0; i \le n; i++){
   x[i] = i;
  clock_gettime(CLOCK_REALTIME,&cgt1);
  for (int i = 1; i \le n; i++){
   y[i] = A * x[i] + y[i];
  clock_gettime(CLOCK_REALTIME,&cgt2);
  ncgt=(double) (cgt2.tv_sec-cgt1.tv_sec) + (double) ((cgt2.tv_nsec-cgt1.tv_nsec) / (1.e+9));
  printf("\ny[0]: %d, y[%d]: %d\n", y[0], n, y[n]);
  printf("Tiempo: %11.9f\n", ncgt);
  return 0;
```

Tiempos ejec.	-O0	-Os	-O2	-O3
Longitud	0.015512014	0.003124218	0.003213841	0.002473006
vectores=1000000				

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

```
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ gcc -00 daxpy.c -o daxpy0 & gcc -00 daxpy.c -5
[1] 9717
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ gcc -0s daxpy.c -o daxpy5 & gcc -0s daxpy.c -5
[2] 9737
[1] Hecho gcc -00 daxpy.c -o daxpy0
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ gcc -02 daxpy.c -o daxpy2 & gcc -02 daxpy.c -5
[3] 9746
[2] Hecho gcc -0s daxpy.c -o daxpy5
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ gcc -03 daxpy.c -o daxpy3 & gcc -03 daxpy.c -5
[4] 9771
[3] Hecho gcc -02 daxpy.c -o daxpy2
```

```
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ ./daxpy0 1000

y[0]: 0, y[1000]: 152000
Tiempo: 0.000010027
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ ./daxpyS 1000

y[0]: 0, y[1000]: 152000
Tiempo: 0.000001124
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ ./daxpy2 1000

y[0]: 0, y[1000]: 152000
Tiempo: 0.000002250
elena@elena97om:~/Escritorio/6AÑO/AC/practicas/BP4/ejer3$ ./daxpy3 1000

y[0]: 0, y[1000]: 152000
Tiempo: 0.000001684
```

COMENTARIOS QUE EXPLIQUEN LAS DIFERENCIAS EN ENSAMBLADOR:

O0: No tiene ninguna optimización.

Os: Es parecido a O2, pero reduce el tamaño del archivo frente a este.

O2: Se emplean instrucciones más eficientes que las usadas en O0 y se reduce el tamaño de archivo respecto a este.

O3: Se genera un código más complejo con más llamadas a subrutinas y aumenta el tamaño del archivo.

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
jmp	.L3:	.L4:	.L9:
.L5	cmpl	leal	leaq
.L6:	%r13d, %eax	(%rax,%rax), %edx	-80(%rbp), %rsi
movq	jg	movl	xorl
-128(%rbp), %rax	.L10	%eax, (%r14,%rax,4)	%edi, %edi
movl	leal	movl	call
-148(%rbp), %edx	(%rax,%rax), %edx	%edx, (%rbx,%rax,4)	clock_gettime@PLT
movslq	movl	addq	movq
%edx, %rdx	%eax, (%r15,%rax,4)	\$1, %rax	-72(%rbp), %rax
movl	movl	cmpq	pxor
(%rax,%rdx,4), %eax	%edx, (%r14,%rax,4)	%rcx, %rax	%xmm0, %xmm0
imull	incq	jne	subq
-152(%rbp), %eax	%rax	. L4	-88(%rbp), %rax
movl	jmp		pxor
%eax, %ecx	.L3		%xmm1, %xmm1
movq			movl
-112(%rbp), %rax			0(,%rbx,4), %edx

movl			movl
-148(%rbp), %edx		0(%r13,%r12,4	
movslq			leaq
%edx, %rdx movl		.LC4(%rip), %	Krsi movl
(%rax,%rdx,4), %eax addl		%r14d, %ecx	movl
%eax, %ecx		\$1, %edi	
movq		0,	cvtsi2sdq
-112(%rbp), %rax movl		%rax, %xmm0	movq
-148(%rbp), %edx		-80(%rbp), %r	•
movslq			subq
%edx, %rdx movl		-96(%rbp), %r	ax cvtsi2sdq
%ecx, (%rax,%rdx,4)		%rax, %xmm1	
addl			xorl
\$1, -148(%rbp)		%eax, %eax	divod
.L5: movl		.LC3(%rip), %	divsd 6xmm0
-148(%rbp), %eax		(- -//	addsd
cmpl		%xmm1, %xmm0	
-140(%rbp), %eax			movsd
jle .L6		%xmm0, -104(%	Grbp) call
		printf_chk@)PLT
		104(%rbp) %	movsd
		-104(%rbp), %	leaq
		.LC5(%rip), %	
			movl
		\$1, %edi	ma 1
		\$1, %eax	movl
			call
		printf_chk@	PLT xorl
		%eax, %eax	X01 1
			movq
		-56(%rbp), %r	xorq
		%fs:40, %rbx	
			jne
		.L32	loog
		-40(%rbp), %r	leaq
		.5(.0.56), 701	popq
		%rbx	
		07.40	popq
		%r12	popq
		%r13	r = r = 1
		04:-4.4	popq
		%r14	popq
		%r15	4.44
			popq
		%rbp	.cfi_remem
		ber_state	
		fa 7, 8	.cfi_def_c
		, -	

			ret
		.L18:	
			.cfi_resto
		re_state	-
			movl
	-	\$1, -104(%rbp)
			jmp
		.L5	
		.L22:	ma1
			movl
		\$2, %esi	imn
		111	jmp
		.L11 .L3:	
		.LJ.	leaq
		-96(%rbp), %r	
			xorl
		%edi, %edi	X01 I
		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	call
		clock_gettime	
		_3	jmp
		.L9	5 P
		.L19:	
			movl
		\$2, -104(%rbp	
			jmp
		.L5	

- 4. **(a)** Paralizar con OpenMP en la CPU el código de la multiplicación resultante en el Ejercicio 1.(b). NOTA: usar para generar los valores aleatorios, por ejemplo, drand48_r().
 - **(b)** Calcular la ganancia en prestaciones que se obtiene en atcgrid4 para el máximo número de procesadores físicos con respecto al código inicial no optimizado del Ejercicio 1.(a) para dos tamaños de la matriz.
 - (a) MULTIPLICACIÓN DE MATRICES PARALELO:

CAPTURA CÓDIGO FUENTE: pmm-paralelo.c

(b) RESPUESTA