Politechnika Poznańska

Wydział Informatyki

Przetwarzanie Równoległe

Równoległe sumowanie komórek pamięci za pomocą wielu wątków przetwarzania

Autorzy: Adam Szczepański Mateusz Czajka

Prowadzący: dr Rafał Walkowiak



Spis treści

| 1 | Info | ormacje o projekcie | 3 | | | | | | | |
|----|------------------------|---|-----------|--|--|--|--|--|--|--|
| | 1.1 | Dane autorów | 3 | | | | | | | |
| | 1.2 | Historia projektu | 3 | | | | | | | |
| 2 | Wstęp | | | | | | | | | |
| | 2.1 | Opis problemu | 4 | | | | | | | |
| | 2.2 | Punkt odniesienia (algorytm sekwencyjny - kolejność ij) | 4 | | | | | | | |
| | $\frac{2.2}{2.3}$ | Badane algorytmy | 5 | | | | | | | |
| | 2.3 | | | | | | | | | |
| | | 2.3.1 Algorytm sekwencyjny - kolejność ji | 5 | | | | | | | |
| | | 2.3.2 Algorytm zrównoleglony - kolejność ij | 5 | | | | | | | |
| | | 2.3.3 Algorytm zrównoleglony - kolejność jj | 5 | | | | | | | |
| | | 2.3.4 Algorytm na sekcyjność pamięci | 6 | | | | | | | |
| | | 2.3.5 Algorytm na pobranie z wyprzedzeniem | 6 | | | | | | | |
| | 2.4 | Maska powinowactwa | 6 | | | | | | | |
| | | 2.4.1 Kod | 7 | | | | | | | |
| | | 2.4.2 Rezultat | 7 | | | | | | | |
| | | 2.4.2 Rezultat | ' | | | | | | | |
| 3 | Pon | niary efektywności | 8 | | | | | | | |
| | 3.1 | Przyśpieszenie obliczeń równoległych | 8 | | | | | | | |
| | | 3.1.1 Pomiary | 8 | | | | | | | |
| | | 3.1.2 Podsumowanie | 9 | | | | | | | |
| | 3.2 | Ilość wykonanych instrukcji na jeden cykl procesora | 9 | | | | | | | |
| | 3.3 | Współczynniki braku trafień do pamięci | 10 | | | | | | | |
| | ა.ა | | | | | | | | | |
| | | 3.3.1 Znaczenie danych | 10 | | | | | | | |
| | | 3.3.2 Pomiary | 10 | | | | | | | |
| | | 3.3.3 Omówienie | 11 | | | | | | | |
| 4 | Wp | ływ rozmiaru danych | 11 | | | | | | | |
| | 4.1 | Wstęp | 11 | | | | | | | |
| | 4.2 | Wyniki | 12 | | | | | | | |
| | 4.3 | · | 12 | | | | | | | |
| | 4.3 | Podsumowanie | 12 | | | | | | | |
| 5 | $\mathbf{W}\mathbf{p}$ | ływ sekcyjności pamięci | 12 | | | | | | | |
| | 5.1 | Wstęp | 12 | | | | | | | |
| | 5.2 | Wyniki | 13 | | | | | | | |
| c | 1 37 | przedzające pobronie donych | 13 | | | | | | | |
| 6 | - | przedzające pobranie danych | | | | | | | | |
| | 6.1 | Klasyczne podejście | 13 | | | | | | | |
| | 6.2 | Zastosowanie kwalifikatora $volatile$ | 14 | | | | | | | |
| 7 | Załączniki | | | | | | | | | |
| 8 | Koc | ł źródłowy | 14 | | | | | | | |
| Sp | ois ry | zsunków | 14 | | | | | | | |

Spis tablic 15

1 Informacje o projekcie

1.1 Dane autorów

Mateusz Czajka 106596 Adam Szczepański 106593

1.2 Historia projektu

1. Jest to pierwsza wersja projektu. Dokumentacja elektroniczna została przesłana w dniu 20 stycznia 2013.

2 Wstęp

2.1 Opis problemu

Głównym założeniem projektu było zapoznanie się biblioteką OpenMP na podstawie równoległego sumowania komórek tablicy. W ramach projektu zrealizowaliśmy 4 algorytmy sekwencyjne oraz 2 algorytmy zrównleglone. Celem zastosowania czterech różnych algorytmów sekwencyjnych było zbadanie wpływu sekcyjności pamięci podręcznej, wyprzdzającego pobrania danych do pamięcy podręcznej oraz kolejności uszeregowania pętli na czas realizacji zadania. W przypadku algorytmów zrównoleglonych badaliśmy wpływ kolejności uszeregowania pętli na końcowy rezultat.

Nasze badania podzieliliśmy na 3 spójne części. Kolejno badaliśmy

- rozmiar danych
- sekcyjność pamięci
- wyprzedzające pobranie

na czas realizacji problemu.

Sam problem sprowadzał się do zsumowania wartości komórek w tabeli. Dla zachowania czytelności w kolejności pętli zastosowaliśmy tablicę dwuwymiarową. Ponieważ sam problem jest prosty obliczeniowo zmuszeni byliśmy do stosowania maksymalnego rozmiaru tablicy tj $[2^{28} - wierszy]$ na $[2^4 - kolumn]$.

2.2 Punkt odniesienia (algorytm sekwencyjny - kolejność ij)

Punktem odniesienia dla wszystkich algorytmów był podstawowy algorytm sekwencyjny w którym sumowaliśmy elementy tablicy wierszami. Nazywany dalej sum_ij .

```
__declspec(noinline) int sum_ij() {
  int sum = 0;
  for (int i=0; i<ROWS; i++) {
    for (int j=0; j<COLS; j++) {
      sum += tab[i][j];
    }
  }
  return sum;
}</pre>
```

2.3 Badane algorytmy

2.3.1 Algorytm sekwencyjny - kolejność ji

Algorytm sekwencyjny ze zmienioną kolejnością pętli (sumujemy kolumnami). Nazywany dalej sum_-ji .

```
__declspec(noinline) int sum_ji() {
  int sum = 0;
  for (int j=0; j<COLS; j++) {
    for (int i=0; i<ROWS; i++) {
       sum += tab[i][j];
    }
  }
  return sum;
}</pre>
```

2.3.2 Algorytm zrównoleglony - kolejność ij

Algorytm sumujący wierszami realizowany równolegle. Nazywany dalej sum_par_ij.

```
__declspec(noinline) int sum_par_ij() {
  int sum = 0;
  int i;

#pragma omp parallel for default(none) shared(tab) private(i) reduction(+:sum)
  for (i=0; i<ROWS; i++) {
    for (int j=0; j<COLS; j++) {
        sum += tab[i][j];
    }
  }
  return sum;
}</pre>
```

2.3.3 Algorytm zrównoleglony - kolejność jj

Algorytm realizowany równolegle ze zmienioną kolejnością pętli (sumujemy kolumnami). Nazwywany dalej $sum_p ar_- ji$.

```
__declspec(noinline) int sum_par_ji() {
  int sum = 0;
  int j;
#pragma omp parallel for default(none) shared(tab) private(j) reduction(+:sum)
  for (j=0; j<COLS; j++) {
    for (int i=0; i<ROWS; i++) {
        sum += tab[i][j];
    }
}
return sum;</pre>
```

}

2.3.4 Algorytm na sekcyjność pamięci

Algorytm sekwencyjny w którym badaliśmy wpływ sekcyjności pamięci na czas wykonania zadania. Nazywany dalej sum_sec .

```
__declspec(noinline) int sum_sec() {
  int sum = 0;
  for (int j=0; j<COLS; j++) {
    for (int k=0; k<CACHE_LINES_ON_PAGE; k++) {
      for (int i=k; i<ROWS; i+=CACHE_LINES_ON_PAGE) {
        sum += tab[i][j];
      }
    }
  }
  return sum;
}</pre>
```

2.3.5 Algorytm na pobranie z wyprzedzeniem

Algorytm sekwencyjny w którym badaliśmy wpływ wyprzedzającego pobrania danych do pamięci podręcznej na czas realizacji zadania. Nazywany dalej sum_pf .

```
int tmp;
__declspec(noinline) int sum_pf() {
  int sum = 0;
  int i;
  for (i=0; i<ROWS-1; i++) {
    for (int j=0; j<COLS; j++) {
       sum += tab[i][j];
       tmp = tab[i+1][j];
    }
}

for (int j=0; j<COLS; j++) {
    sum += tab[i][j];
}
  return sum;
}</pre>
```

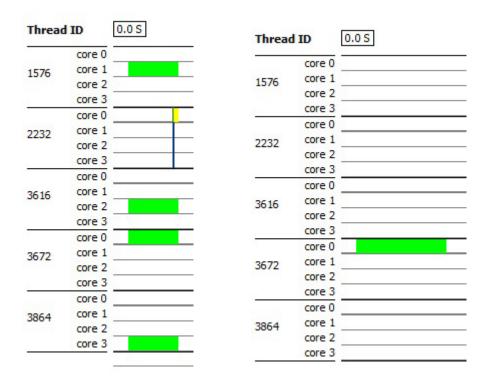
2.4 Maska powinowactwa

Aby wyeliminować przełączanie wykonywania wątku pomiędzy rdzeniami procesora zastosowaliśmy maski powinowactwa.

2.4.1 Kod

2.4.2 Rezultat

W rezultacie dla przetwarzania równoległego każdy wątek odpowiedzialny za sumowanie realizowany jest na oddzielnym procesorze. Co widoczne jest na poniższych wykresach:



Rysunek 1: Wykres wątków dla
 Rysunek 2: Wykres wątków dla sum_ij sum_par_ij

3 Pomiary efektywności

3.1 Przyśpieszenie obliczeń równoległych

3.1.1 Pomiary

Tablica 1: Porównanie szybkości wybranych algorytmów wobec sum_ij

| Algorytm | Czas wykonania w ms | Przyśpieszenie względem sum_ij | |
|------------------------|-----------------------|--------------------------------|--|
| $sum_{-}ij$ 253 | | 1.000 | |
| sum_ji | 2362 | 0.107 | |
| sum_par_ij 87 | | 2.908 | |
| <i>sum_par_ji</i> 1189 | | 0.213 | |

Warto zwrócić uwagę na fakt, że oprócz braku lub obecności zrównoleglenia znaczący wpływ na szybkość przetwarzania ma kolejność uszeregowania pętli. Zostało to szerzej omówione w sekcji 3.3. Przyśpieszenia dla poszczególnych uszeregowań pętli prezentują się następująco:

Tablica 2: Porównanie szybkości wybranych algorytmów (o kolejności pętli ij) wobec $sum_{\cdot}ij$

| Algorytm Czas wykonania | | Przyśpieszenie względem sum_ij | |
|-------------------------|------|-----------------------------------|--|
| sum_ij | 2362 | 1.000 | |
| sum_par_ij | 87 | 2.908 | |

Tablica 3: Porównanie szybkości wybranych algorytmów (o kolejności pętli ji) wobec sum_ji

| Algorytm Czas wykonania w ms | | Przyśpieszenie względem $sum_{-}ji$ | |
|--------------------------------|------|-------------------------------------|--|
| sum_ji | 2362 | 1.000 | |
| <i>sum_par_ji</i> 1189 | | 1.987 | |

3.1.2 Podsumowanie

Zrównoleglenie przetwarzania znacząco przyśpiesza jego czas przetwarzania. Znaczeny wpływ na wielkość wartości tego przyśpieszenia ma kolejność uszeregowania pętli. W przybliżeniu jest to:

- trzykrotne przyśpieszenie dla uszeregowania pętli ij
- \bullet dwukrotne przyśpieszenie dla uszeregowania pętli ji

3.2 Ilość wykonanych instrukcji na jeden cykl procesora

IPC (insructions per cycle) jest jednym z wyznaczników prędkości procesora. Oznacza on liczbę wykonywanych instrukcji przez procesor w jednym cyklu zegara. Wskaźnik IPC obliczaliśmy dla każdej funkcji na podstawie wzoru:

$$IPC = \frac{ret_instr}{CPU_clocks}. (1)$$

CPI (cycles per instruction) jest odwrotnością CPI:

$$CPI = \frac{CPU_clocks}{ret_instr}.$$
 (2)

Tablica 4: Wartości IPC i CPI dla poszczególnych algorytmów

| Algorytm | ret_instr | CPU_clocks | IPC | CPI |
|----------------|--------------|---------------|------|-------|
| sum_ij | 12640 | 14736 | 0.86 | 1.17 |
| sum_ji | 9780 | 136048 | 0.07 | 13.91 |
| sum_par_ij | 12684 | 20096 | 0.63 | 1.58 |
| sum_par_ji | 11524 | 259628 | 0.04 | 22.53 |
| sum_sec | 14420 | 1139472 | 0.01 | 79.02 |
| sum_pf | 14996 | 13140 | 1.14 | 0.88 |

Zgodnie z oczekiwaniami najwyższe wartości wzkaźnika IPC (a zarazem najniższe CPI) wysętpują w przypadku algorytmów, które spełniają zasadę lokalności przestrzennej, są to:

- *sum_ij*
- sum_par_ij
- *sum_pf*

IPC dla wyżej wymienionej grupy algorytmów jest kilkunasto a w niektórych przypadkach kilkudziesiącio krotonie większe niż dla algorytmów z pozostałej grupy.

Współczynniki braku trafień do pamięci 3.3

3.3.1 Znaczenie danych

Tablica 5: Badane dane

| Nazwa | Znaczenie |
|---------------|---|
| DC accesses | ilość odwołań do pamięci podręcznej |
| DC misses | ilość chybień do pamięci podręcznej |
| L2 requests | ilość odwołań do pamięci L2 |
| L2 misses | ilość chybień do pamięci L2 |
| L2 fill write | zapis danych w L2 powodowany usunięciem ich z pp L1 |
| DRAM accesses | ilość odwołań do pamięci dynamicznej |

3.3.2 Pomiary

Wartości DC_accesss oraz DC_missess odczytaliśmy bezpośrednio z programu Code Analyst. Natomiast DC_miss_ratio oraz DC_miss_rate obliczyliśmy w następujący sposób:

$$DC_miss_rate = \frac{DC_missess}{ret_instr}$$

$$DC_miss_ratio = \frac{DC_missess}{DC_accesss}$$

$$(3)$$

$$DC_miss_ratio = \frac{DC_missess}{DC_accesss}$$
 (4)

Tablica 6: Pomiary związane z odwołaniami do pamięci L1

| Algorytm DC accessess | | DC missess | DC miss ratio | DC miss rate |
|-----------------------|-------|------------|---------------|--------------|
| $sum_{-}ij$ | 5792 | 24 | 0% | 0% |
| $sum_{-}ji$ | 7620 | 4068 | 53% | 42% |
| sum_par_ij | 5188 | 20 | 0% | 0% |
| sum_par_ji | 7652 | 4456 | 58% | 38% |
| sum_sec | 18808 | 5256 | 28% | 36% |
| sum_pf | 9928 | 24 | 0% | 0% |

Podobnie jak w przypadku pamięci L1 wartości L2_requests, L2_misses oraz L2_fill_write odczytaliśmy bezpośrednio z programu CodeAnalyst. Natomiast L2_miss_ratio oraz L2_miss_rate obliczyliśmy w następujący sposób:

$$L2_miss_rate = \frac{L2_missess}{ret_instr}$$
 (5)

$$L2_miss_rate = \frac{L2_missess}{ret_instr}$$

$$L2_miss_ratio = \frac{L2_missess}{L2_requests + L2_fill_write}$$
(6)

Tablica 7: Pomiary związane z odwołaniami do pamięci L2

| | , | , | | - F 2 | |
|----------------|-------------|------------|---------------|--------------|---------------|
| Algorytm | L2 requests | L2 missess | L2 fill write | L2 miss rate | L2 miss ratio |
| sum_ij | 663 | 340 | 684 | 3% | 25% |
| $sum_{-}ji$ | 5452 | 7209 | 10908 | 56% | 30% |
| sum_par_ij | 1431 | 352 | 668 | 3% | 17% |
| sum_par_ji | 5472 | 23736 | 10844 | 47% | 16% |
| sum_sec | 11024 | 15945 | 21668 | 76% | 29% |
| sum_pf | 344 | 693 | 684 | 2% | 25% |

3.3.3 Omówienie

Znowu można wyróżnić 2 grupy algorytmów. Pierwsza z nich składająca się z:

- *sum_ij*
- sum_par_ij
- *sum_pf*

cechują się one lokalnością przestrzenną. Widać to bardzo wyraźnie przy współczynnikach DC_miss_ratio oraz DC_miss_rate. Dla wyżej wymienionej grupy wynoszą one w przybliżeniu 0%, a dla pozostałych algorytmów tj sum_ji , sum_par_ji oraz sum_sec po kilkadziesiąt procent. Związane to jest z faktem że grupa posiadająca cechę lokalności przestrzennej sumuje wierszami, zatem z raz wczytanej linii do pamięci podręcznej wszystkie elementy są wykorzystywane jeden po drugim.

Wpływ rozmiaru danych 4

4.1 Wstep

W celu sprawdzenia wpływu rozmiaru danych na czas realizacji zadania dokonaliśmy pomiarów czasów dla dwóch rozmiarów danych:

- Instancja A: $tab[2^{28}][2^4]$
- Instancja B: $tab[2^{24}][2^4]$

Spodziewaliśmy się, że czas obliczeń dla Instancji B będzie kilkukrotnie krótszy.

4.2 Wyniki

Wyniki pomiarów prezentują się następująco:

Tablica 8: Czas realizacji kodu dla poszczególnych funkcji w ms.

| Problem | sum_ij | sum_ji | sum_sec | sum_pf | sum_par_ij | sum_par_ji |
|--------------|-----------|-----------|------------|-----------|----------------|----------------|
| Instancja A | 253 | 2352 | 19552 | 225 | 87 | 1189 |
| Instancja B | 62 | 59 | 4890 | 57 | 23 | 304 |
| Stosunek A/B | 4.081 | 4.003 | 3.998 | 3.947 | 3.783 | 3.911 |

4.3 Podsumowanie

Wyniki eksperymentu są bardzo zadowalające. W przypadku każdego algorytmu rozmiar danych ma liniowy wpływ na czas przetwarzania. Jest to zgodne ze złożonością algorytmu która jest liniowa wobec n gdzie n to $liczba_wierszy*liczba_kolumn$.

W każdym przypadku czas realizacji dla Instancji B był około 4 razy krótszy niż dla Instancji A. Średnia tych wartości wynosi 3.954.

5 Wpływ sekcyjności pamięci

5.1 Wstęp

W trakcie naszych badań sprawdziliśmy wpływ sekcyjności pamięci na czas przetwarzania. Kluczowe dla tego podpunktu algorytmy to:

- $\bullet \ sum.ij$ podstawowy algorytm sekwencyjny, służący jako punkt odeniesienia
- \bullet sum_sec algorytm sekwencyjny w którym badaliśmy sekcyjność pamięci

W wykorzystywanym systemie obliczeniowym liczba lini w pojedynczym bloku pamięci L1 wynosi 64.

W celu sprawdzenia w jaki sposób sekcyjność wpływa na czas realizacji zmodyfikowaliśmy kod w następujący sposób:

```
for (int j=0; j<COLS; j++) {
  for (int k=0; k<CACHE_LINES_ON_PAGE; k++) {
   for (int i=k; i<ROWS; i+=CACHE_LINES_ON_PAGE) {
     sum += tab[i][j];
   }
}</pre>
```

Sumujemy kolumnami, wprowadzając dodatkową, środkową pętlę która sumuje co 64 element. Dzięki temu każdy sumowany element musi być pobrany z pamięci dynamicznej.

5.2 Wyniki

Zgodnie z założeniem, powinien to być najwolniejszy algorytm, ponieważ spodziewana jest bardzo wysoka liczba odwołań do pamięci RAM or wysoki stosunek braku trafień do pamięci podręcznej.

Tablica 9: Porównanie wybranych algorytmów pod kątem sekcyjności pamięci

| Algorytm | Czas [ms] | DC missess | L2 misses | DRAM access |
|------------|-----------|------------|-----------|-------------|
| sum_ij | 253 | 24 | 340 | 352 |
| sum_sec | 19552 | 5256 | 11024 | 5920 |

Wyniki przerosły nasze oczekiwania - sum_sec wykonał się ponad 77 razy wolniej. Pokazuje to jest wielkie znaczenie na szybkość realizacji przetwarzania ma zachowanie cechy lokalności przestrzennej. Wyraźnie widać, że w przypadku algorytmu użytego jako punkt odniesienia liczba braków trafień do pamięci podręcznej L1 była znacznie mniejsza. Sytuacja wygląda bardzo podobnie dla pamięci podręcznej L2. Konsekwencją braku trafień do pamięci podręcznej jest konieczność większej liczby pobrań danych z pamięci dynamicznej co odzwierciedla współczynnik $DRAM_accesses$.

6 Wyprzedzające pobranie danych

Wyprzedzające pobranie danych potraktowaliśmy jako swego rodzaju eksperyment. Nasze obawy związane z optymlizacją kody okazały się być uzasadnione. W celu realizacji problemu dodaliśmy zmienną tmp, na której nie były realizowane żadne operacje. Eksperyment podzieliliśmy na dwie części.

6.1 Klasyczne podejście

Początkowo spodziewaliśmy się, że kompilator usunie nic nie robiące przepisanie. Po przejrzeniu instrukcji Assemblera okazało się, że jest inaczej.

```
        4
        0xb41234
        94
        sum += tab[i];

        0xb41237
        add eax,[ecx+0sh]
        0341FC

        0xb41237
        add edx,[ecx]
        0311

        0xb41239
        add edi,[ecx+0sh]
        037104

        0xb41239
        add edi,[ecx+0sh]
        037908

        4
        0xb4125f
        95
        tmp = tab[i+1][j];

        0xb41247
        oxb41248
        add ecx,l6h
        85948

        0xb41246
        dec dword [ebp-0sh]
        FF49FC

        0xb41248
        dec dword [ebp-0sh]
        FF49FC

        0xb41249
        ing S-1ah (8xh41244)
        75F4

        0xb41249
        ing S-1ah (8xh41244)
        75F4
```

Rysunek 3: Instrukcja przypisania do zmiennej tmp

Tablica 10: Czasy algorytmów kluczowych dla wyprzedzającego pobrania

| Algorytm | Czas [ms] |
|-----------|-----------|
| sum_ij | 253 |
| sum_pf | 225 |

Przyśpieszenie związane z wyprzedzającym pobraniem danych okazało się jednak być bardzo nieznaczne.

6.2 Zastosowanie kwalifikatora volatile

volatile w C++ jest kwalifikatorem typu informującym kompilator, ze wartość zmiennej może się zmienić bez jego wiedzy i kontroli i że w związku z tym kompilator powinien zrezygnować z agresywnej optymalizacji i przy każdym odwołaniu do tej zmiennej wczytać nową wartość z komórki pamięci.

Kod odpowiedzialny za inicjalizację zmiennej tmp prezentował się następująco:

```
volatile int tmp;
```

Zastosowanie wyżej wymienionego kwalifikatora spowodowało zgodnie z oczekiwaniami pogorszenie czasu przetwarzania.

7 Załączniki

1. sum.cpp – plik z kodem źródłowym

8 Kod źródłowy

```
#include <cstdio>
#include <cstdlib>
#include <ctime>
#include <Windows.h>
#include <omp.h>

#define CACHE_LINE 64
#define CACHE_ALIGN __declspec(align(CACHE_LINE))
#define CACHE_LINES_ON_PAGE 512

// ROWS * COLS = 1<<28 = 256 MB (2^8 * 2^10 * 2^10)
#define ROWS ((size_t) 1<<24)
#define COLS ((size_t) 1<<4)

CACHE_ALIGN int tab[ROWS][COLS];
double start;</pre>
```

```
HANDLE thread_handle = GetCurrentThread();
inline void start_clock() {
 start = (double) clock() / CLK_TCK;
inline void print_elapsed_time(const char* operation_name) {
 double elapsed;
 double resolution;
 elapsed = (double) clock() / CLK_TCK;
 resolution = 1.0 / CLK_TCK;
 printf("%s: %8.4f sec\n", operation_name, elapsed-start);
}
void init_tab() {
 start_clock();
 int i;
#pragma omp parallel for shared(tab) private(i)
 for (i=0; i<ROWS; i++) {</pre>
   for (int j=0; j<COLS; j++) {</pre>
     tab[i][j] = rand();
 }
 print_elapsed_time("init");
typedef int (*function)();
inline void run_function(function f, const char* operation_name) {
 printf("----\n");
 start_clock();
 int sum = f();
 printf("sum: %d\n", sum);
 print_elapsed_time(operation_name);
__declspec(noinline) int sum_ij() {
 int sum = 0;
 for (int i=0; i<ROWS; i++) {</pre>
   for (int j=0; j<COLS; j++) {</pre>
     sum += tab[i][j];
   }
 }
 return sum;
```

```
}
__declspec(noinline) int sum_ji() {
  int sum = 0;
  for (int j=0; j<COLS; j++) \{
    for (int i=0; i<ROWS; i++) {</pre>
     sum += tab[i][j];
 return sum;
__declspec(noinline) int sum_sec() {
  int sum = 0;
 for (int j=0; j<COLS; j++) {</pre>
   for (int k=0; k<CACHE_LINES_ON_PAGE; k++) {</pre>
      for (int i=k; i<ROWS; i+=CACHE_LINES_ON_PAGE) {</pre>
        sum += tab[i][j];
     }
   }
  }
  return sum;
}
int tmp;
__declspec(noinline) int sum_pf() {
  int sum = 0;
  int i;
  for (i=0; i<ROWS-1; i++) {</pre>
   for (int j=0; j<COLS; j++) {</pre>
     sum += tab[i][j];
     tmp = tab[i+1][j];
   }
  }
  for (int j=0; j<COLS; j++) {</pre>
    sum += tab[i][j];
  return sum;
__declspec(noinline) int sum_par_ij() {
  int sum = 0;
  int i;
#pragma omp parallel for default(none) shared(tab) private(i) reduction(+:sum)
```

```
for (i=0; i<ROWS; i++) {</pre>
   for (int j=0; j<COLS; j++) {</pre>
     sum += tab[i][j];
   }
 }
 return sum;
}
__declspec(noinline) int sum_par_ji() {
 int sum = 0;
 int j;
#pragma omp parallel for default(none) shared(tab) private(j) reduction(+:sum)
 for (j=0; j<COLS; j++) {</pre>
   for (int i=0; i<ROWS; i++) {</pre>
     sum += tab[i][j];
 }
 return sum;
int main(int argc, char* argv[]) {
 omp_set_num_threads(4);
#pragma omp parallel
 {
   #pragma omp single
     printf("num threads: %d\n", omp_get_num_threads());
   }
   const int processor_count = 4;
   int th_id=omp_get_thread_num();
   DWORD_PTR mask = (1 << (th_id % processor_count));</pre>
   DWORD_PTR result = SetThreadAffinityMask(thread_handle, mask);
   if (result==0) {
     printf("error SetThreadAffnityMask\n");
   else {
     printf("previous mask for thread %d: %d\n",th_id,result);\\
     printf("new mask for thread %d : %d\n",
       th_id,SetThreadAffinityMask(thread_handle, mask));
   }
 }
```

```
init_tab();
 run_function(sum_ij, "sum_ij");
 run_function(sum_ji, "sum_ji");
 run_function(sum_sec, "sum_sec");
 run_function(sum_pf, "sum_pf");
 run_function(sum_par_ij, "sum_par_ij");
 run_function(sum_par_ji, "sum_par_ji");
 system("pause");
Spis rysunków
  1
     2
     3
     Spis tablic
     Porównanie szybkości wybranych algorytmów wobec sum_{-ij} . . .
  1
                                               8
  2
     Porównanie szybkości wybranych algorytmów (o kolejności pętli
     9
  3
     Porównanie szybkości wybranych algorytmów (o kolejności pętli
                                               9
     Wartości IPC i CPI dla poszczególnych algorytmów . . . . . . .
                                               9
  4
  5
     10
  6
     7
     Pomiary związane z odwołaniami do pamięci L2 . . . . . . . . . . . .
                                              11
  8
     Czas realizacji kodu dla poszczególnych funkcji w ms. . . . . . . .
  9
     Porównanie wybranych algorytmów pod kątem sekcyjności pamięci 13
  10
     Czasy algorytmów kluczowych dla wyprzedzającego pobrania . . 14
```