Эффективное программирование современных микропроцессоров и мультипроцессоров

Практическое задание 3

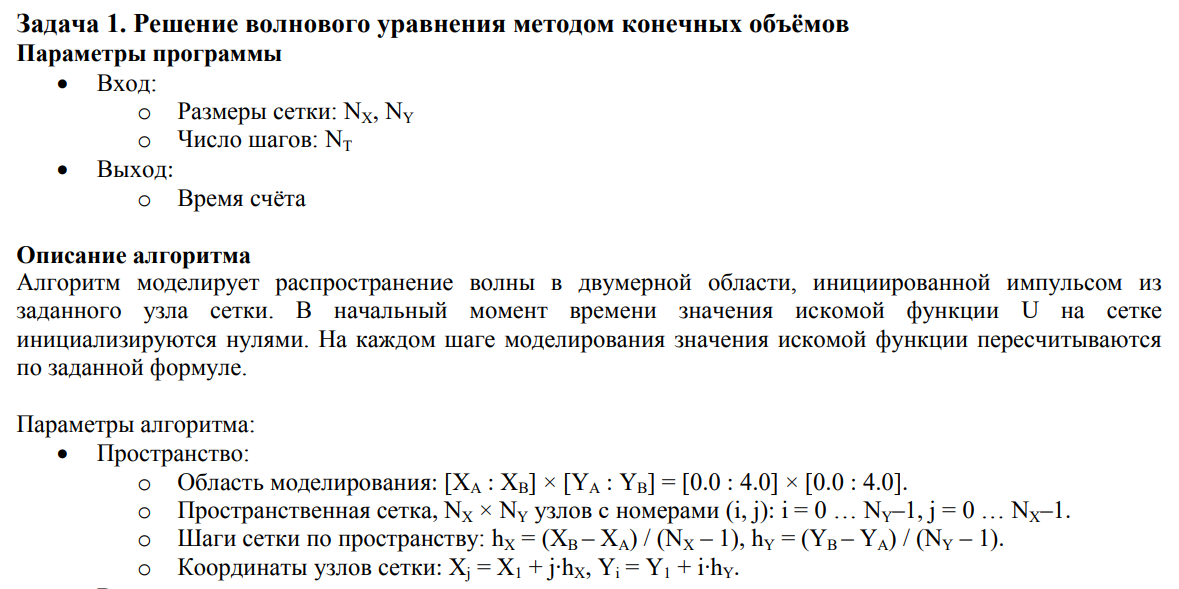
Выполнила:

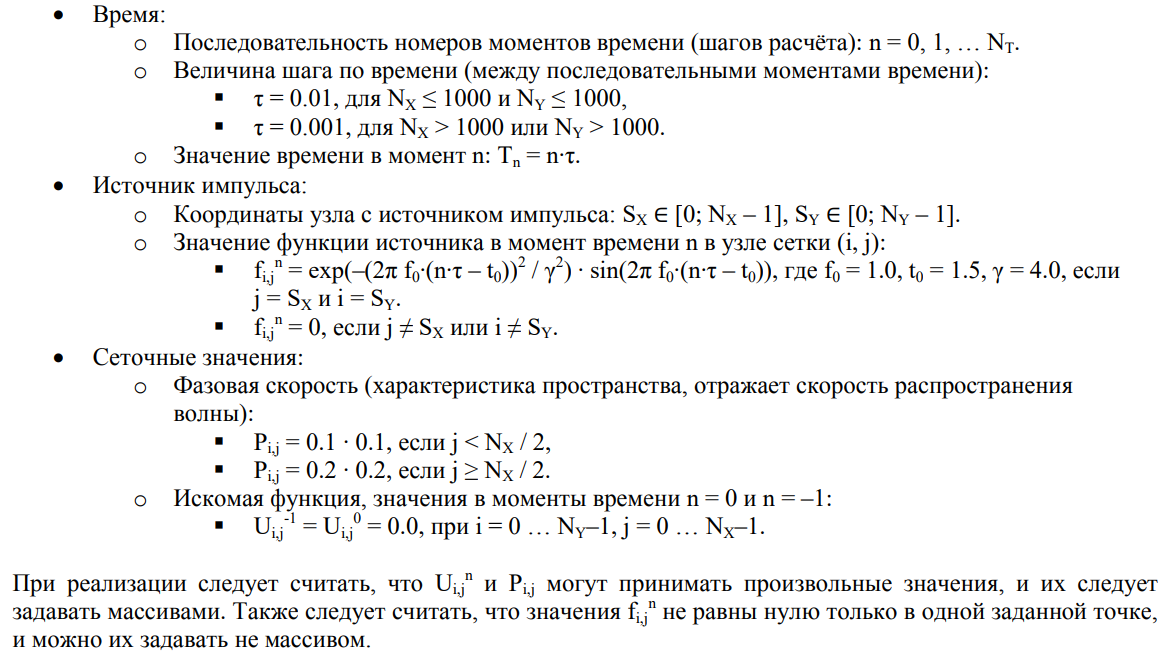
…

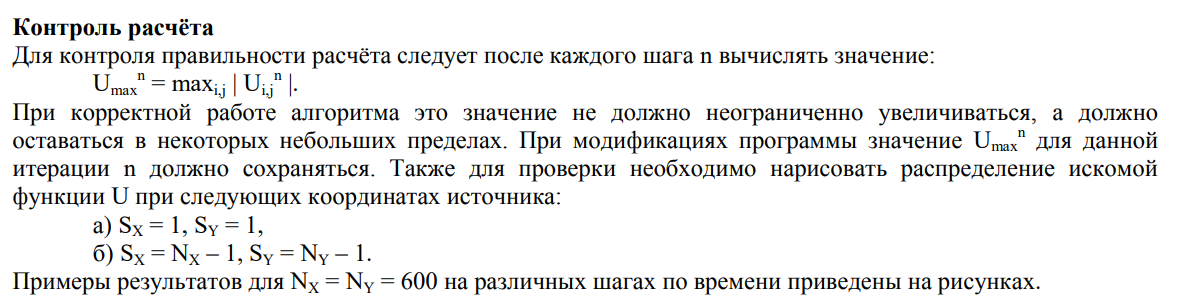
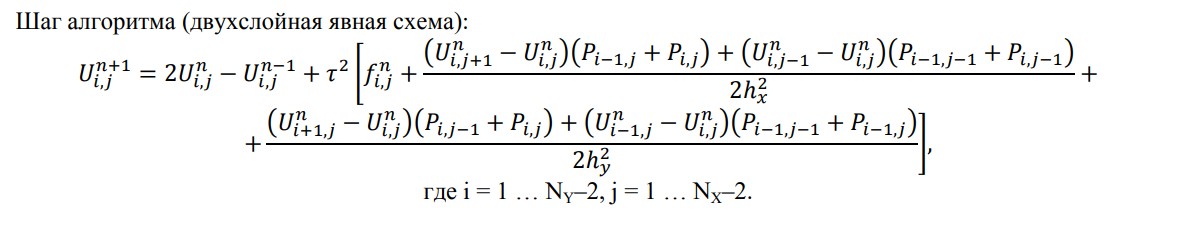
Группа …

Преподаватель:

…







**Результаты тестирования различных версий программы**

Nx = Ny = 8000, Nt = 100;

1. Изначальный вариант программы

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 204.2 | 47.2 | 37.5 | 22.9 |

1. Исключение повторяющихся вычислений

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 203.0 | 42.6 | 34.6 | 22.3 |

1. Замена всех операций деления на умножение

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 195.2 | 42.0 | 33.4 | 21.9 |

1. Заблаговременное обращение к массиву

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 139.7 | 38.7 | 32.9 | 22.4 |

1. Заблаговременное вычисление индекса элемента массива

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 102.8 | 31.5 | 24.5 | 23.7 |

1. Векторизация (n+=1)

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 101.2 | 18.2 | 15.9 | 15.4 |

1. n+=2

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 112.5 | 17.6 | 12.7 | 12.8 |

1. n+=3

|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 110.1 | 16.0 | 10.5 | 10.7 |

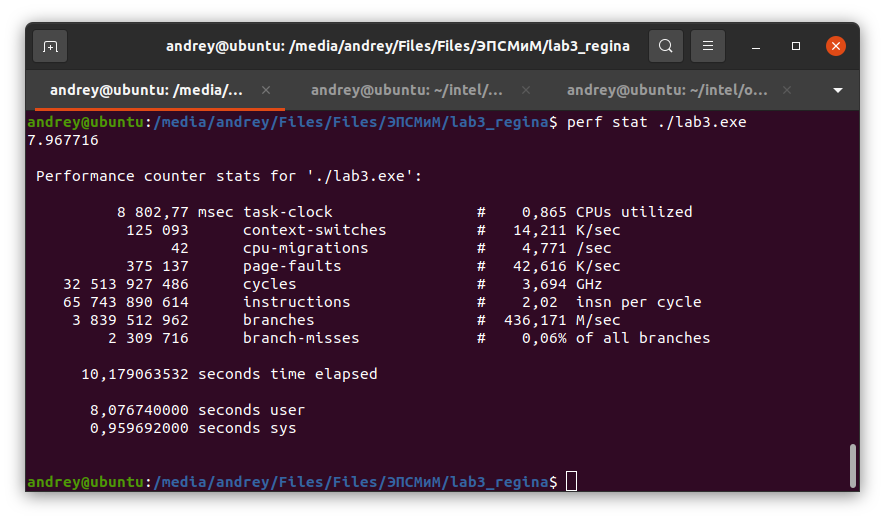
1. n+=4

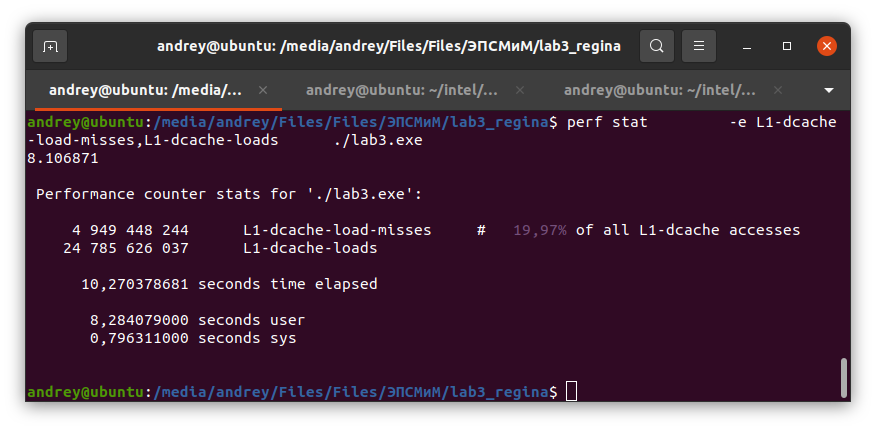
|  |  |  |  |
| --- | --- | --- | --- |
| Без оптимизации | -О1 | -О3 | -Ofast |
| 109.2 | 15.6 | 10.4 | 10.3 |

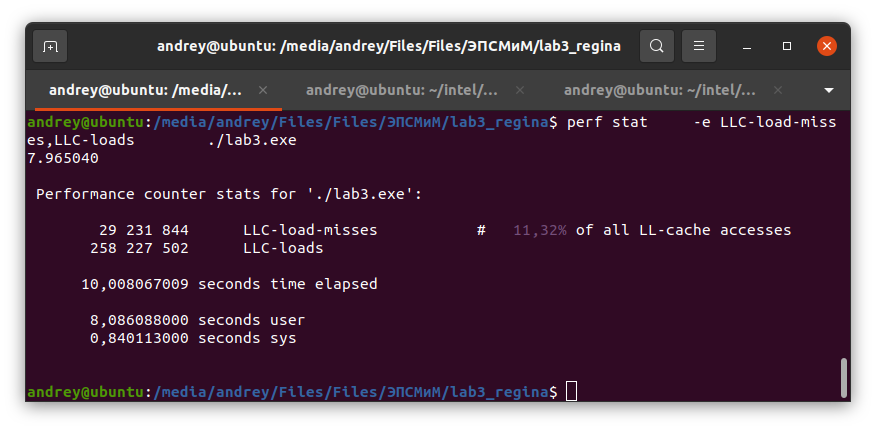
**Вывод**

Оптимизация по памяти ускорила программу в 1.5 раза. Лучшее время достигается при выполнении четырёх итераций за проход по сетке. Теперь результаты вычислений используются сразу, а не при следующем обходе массива, когда они могли бы уже не быть в кэш-памяти. Это привело к существенному уменьшению количества кэш-промахов и обращений к оперативной памяти. Профилирование также показывает, что нагрузка на память сократилась.

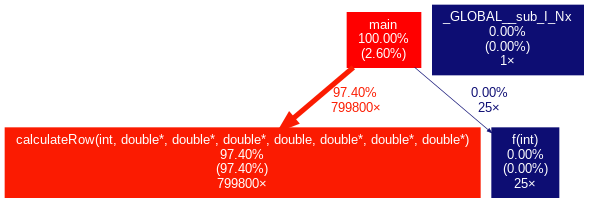
**Профилирование perf**



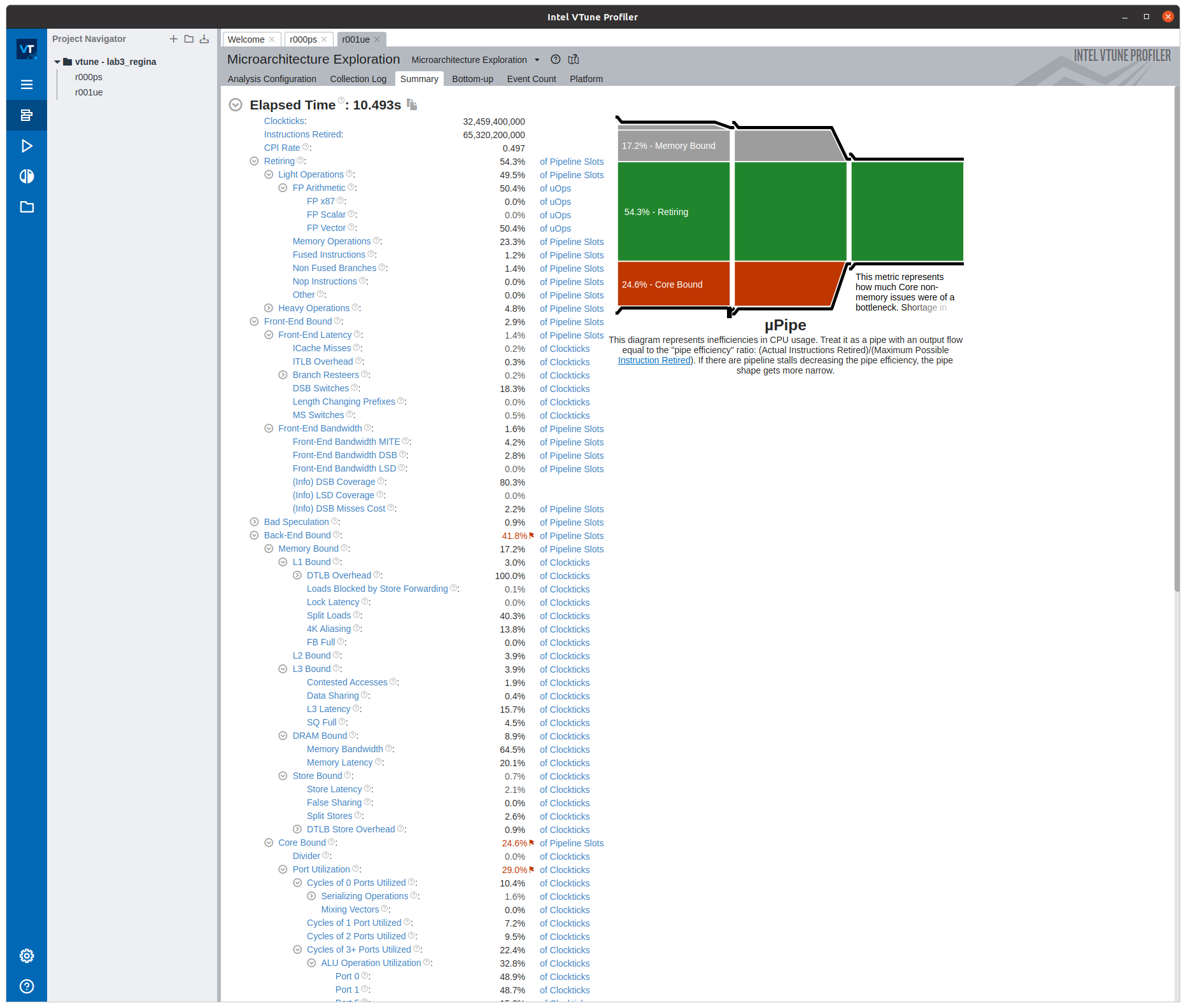




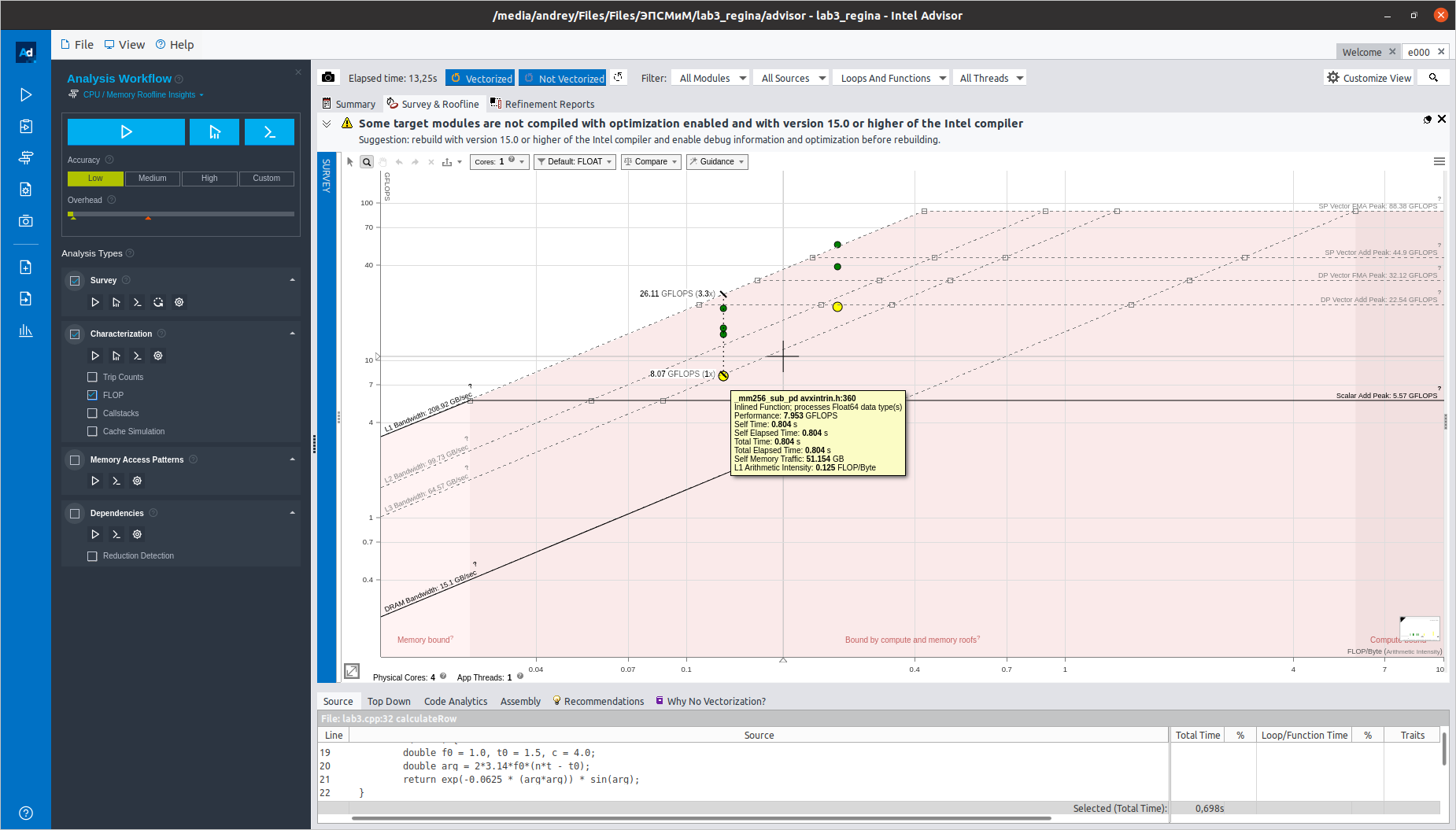
**Граф вызовов**



**Профилирование Vtune**



**Roofline-модель**



#include <iostream>

#include <math.h>

#include <cstdlib>

#include <ctime>

#include <stdio.h>

#include <cstring>

#include <memory>

#include <xmmintrin.h>

#include <immintrin.h>

using namespace std;

int Nx = 8000, Ny = 8000, Nt = 100;

int Sx = 1, Sy = 1;

double t;

const int ALIGN\_INDENT = 64;

const int VECTOR\_SIZE = 4;

double f (int n) {

double f0 = 1.0, t0 = 1.5, c = 4.0;

double arg = 2\*3.14\*f0\*(n\*t - t0);

return exp(-0.0625 \* (arg\*arg)) \* sin(arg);

}

void calculateRow(int i, double \*U\_prv, double \*U\_new, double \*P,

double f\_value, double \*d\_x\_arr, double \*d\_y\_arr, double \*t2\_arr) {

double two\_arr[VECTOR\_SIZE] = {2.0, 2.0, 2.0, 2.0};

double f\_arr[VECTOR\_SIZE];

int row = i \* Nx, row\_ = (i - 1) \* Nx, row1 = (i + 1) \* Nx;

for (int j = 1; j < Ny - 3; j += VECTOR\_SIZE) {

int rowcol = row + j, rowcol\_ = row + j - 1, rowcol1 = row + j + 1,

row\_col = row\_ + j, row1col = row1 + j, row\_col\_ = row\_col - 1;

if (i == Sx && (Sy - j) < VECTOR\_SIZE && (Sy - j) >=0)

f\_arr[(Sy - j)] = f\_value;

else

memset(f\_arr, 0, 4 \* sizeof(double));

\_\_m256d first = \_mm256\_loadu\_pd(U\_prv+rowcol1);

\_\_m256d second = \_mm256\_loadu\_pd(U\_prv+rowcol);

\_\_m256d res1 = \_mm256\_sub\_pd(first, second);

first = \_mm256\_loadu\_pd(P+row\_col);

second = \_mm256\_loadu\_pd(P+rowcol);

\_\_m256d res2 = \_mm256\_add\_pd(first, second);

\_\_m256d firstPath = \_mm256\_mul\_pd(res1, res2);

first = \_mm256\_loadu\_pd(U\_prv+rowcol\_);

second = \_mm256\_loadu\_pd(U\_prv+rowcol);

res1 = \_mm256\_sub\_pd(first, second);

first = \_mm256\_loadu\_pd(P+row\_col\_);

second = \_mm256\_loadu\_pd(P+rowcol\_);

res2 = \_mm256\_add\_pd(first, second);

\_\_m256d secondPath = \_mm256\_mul\_pd(res1, res2);

\_\_m256d lineDividend = \_mm256\_add\_pd(firstPath, secondPath);

first = \_mm256\_loadu\_pd(d\_x\_arr);

\_\_m256d firstLineResult = \_mm256\_mul\_pd(lineDividend, first);

first = \_mm256\_loadu\_pd(U\_prv+row1col);

second = \_mm256\_loadu\_pd(U\_prv+rowcol);

res1 = \_mm256\_sub\_pd(first, second);

first = \_mm256\_loadu\_pd(P+rowcol\_);

second = \_mm256\_loadu\_pd(P+rowcol);

res2 = \_mm256\_add\_pd(first, second);

firstPath = \_mm256\_mul\_pd(res1, res2);

first = \_mm256\_loadu\_pd(U\_prv+row\_col);

second = \_mm256\_loadu\_pd(U\_prv+rowcol);

res1 = \_mm256\_sub\_pd(first, second);

first = \_mm256\_loadu\_pd(P+row\_col\_);

second = \_mm256\_loadu\_pd(P+row\_col);

res2 = \_mm256\_add\_pd(first, second);

secondPath = \_mm256\_mul\_pd(res1, res2);

lineDividend = \_mm256\_add\_pd(firstPath, secondPath);

first = \_mm256\_loadu\_pd(d\_y\_arr);

\_\_m256d secondLineResult = \_mm256\_mul\_pd(lineDividend, first);

\_\_m256d sumOfTwoLines = \_mm256\_add\_pd(firstLineResult, secondLineResult);

\_\_m256d impulseVector = \_mm256\_loadu\_pd(f\_arr);

\_\_m256d bracketResult = \_mm256\_add\_pd(sumOfTwoLines, impulseVector);

first = \_mm256\_loadu\_pd(two\_arr);

second = \_mm256\_loadu\_pd(U\_prv+rowcol);

\_\_m256d totalResult = \_mm256\_mul\_pd(first, second);

second = \_mm256\_loadu\_pd(U\_new+rowcol);

totalResult = \_mm256\_sub\_pd(totalResult, second);

\_\_m256d tauSquareVector = \_mm256\_loadu\_pd(t2\_arr);

\_\_m256d mulResult = \_mm256\_mul\_pd(tauSquareVector, bracketResult);

totalResult = \_mm256\_add\_pd(totalResult, mulResult);

\_mm256\_storeu\_pd(U\_new+rowcol, totalResult);

}

}

int main(int argc, char\*\* argv) {

double Xa = 0.0, Ya = 0.0, Xb = 4.0, Yb = 4.0;

double hx = (Xb - Xa) / (Nx - 1), hy = (Yb - Ya) / (Ny - 1);

if (Nx <= 1000 && Ny <= 1000) t = 0.01;

else t = 0.001;

double\* P = (double\*)\_mm\_malloc(Nx \* Ny \* sizeof(double), ALIGN\_INDENT);

for (int i = 0; i < Ny; i++)

for (int j = 0; j < Nx; j++) {

if (j < Nx/2) P[i\*Nx + j] = 0.1\*0.1;

else P[i\*Nx + j] = 0.2\*0.2;

}

double\*\* U = (double\*\*)malloc(2 \* sizeof(double\*));

U[0] = (double\*)\_mm\_malloc(Nx \* Ny \* sizeof(double), ALIGN\_INDENT);

U[1] = (double\*)\_mm\_malloc(Nx \* Ny \* sizeof(double), ALIGN\_INDENT);

for (int i = 0; i < Ny; i++)

for (int j = 0; j < Nx; j++) {

U[0][i\*Nx + j] = 0.0;

U[1][i\*Nx + j] = 0.0;

}

double d\_x = 0.5 / (hx \* hx), d\_y = 0.5 / (hy \* hy), t2 = t \* t;

double t2\_arr[VECTOR\_SIZE] = {t2, t2, t2, t2};

double d\_x\_arr[VECTOR\_SIZE] = {d\_x, d\_x, d\_x, d\_x};

double d\_y\_arr[VECTOR\_SIZE] = {d\_y, d\_y, d\_y, d\_y};

struct timespec start, end;

clock\_gettime(CLOCK\_MONOTONIC\_RAW, &start);

for (int n = 1; n < Nt; n += 2) {

double\* U\_new = U[n%2];

double\* U\_prv = U[(n+1)%2];

double f\_value = f(n);

calculateRow(1, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

for (int i = 2; i < Ny - 1; i++) {

calculateRow(i, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

calculateRow(i-1, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

}

calculateRow(Ny-2, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

}

/\*for (int n = 1; n < Nt; n += 3) {

double\* U\_new = U[n%2];

double\* U\_prv = U[(n+1)%2];

double f\_value = f(n);

calculateRow(1, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=1 i=1

calculateRow(2, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=1 i=2

calculateRow(1, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=2 i=1

for (int i = 3; i < Ny - 1; i++) {

calculateRow(i, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

calculateRow(i-1, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

calculateRow(i-2, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

}

calculateRow(Ny-2, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=2 i = Ny-2

calculateRow(Ny-3, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=3 i = Ny-3

calculateRow(Ny-2, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=3 i = Ny-2

}\*/

/\*for (int n = 1; n < Nt; n += 4) {

double\* U\_new = U[n%2];

double\* U\_prv = U[(n+1)%2];

double f\_value = f(n);

calculateRow(1, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=1 i=1

calculateRow(2, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=1 i=2

calculateRow(1, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=2 i=1

calculateRow(3, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=1 i=3

calculateRow(2, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=2 i=2

calculateRow(1, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=3 i=1

for (int i = 4; i < Ny - 1; i++) {

calculateRow(i, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

calculateRow(i-1, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

calculateRow(i-2, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

calculateRow(i-3, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr);

}

calculateRow(Ny-2, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=2 i = Ny-2

calculateRow(Ny-3, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=3 i = Ny-3

calculateRow(Ny-2, U\_prv, U\_new, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=3 i = Ny-2

calculateRow(Ny-4, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=4 i = Ny-4

calculateRow(Ny-3, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=4 i = Ny-3

calculateRow(Ny-2, U\_new, U\_prv, P, f\_value, d\_x\_arr, d\_y\_arr, t2\_arr); //n=4 i = Ny-2

}\*/

clock\_gettime(CLOCK\_MONOTONIC\_RAW, &end);

double time = end.tv\_sec-start.tv\_sec + 0.000000001\*(end.tv\_nsec-start.tv\_nsec);

printf("%lf\n", time);

FILE\* file = fopen("file.dat", "wb");

fwrite(U[1], sizeof(double), Nx \* Ny, file);

fclose(file);

return 0;

}