МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет информационных технологий Кафедра параллельных вычислений

ОТЧЕТ

О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ

«Параллельная реализация решения системы линейных алгебраических уравнений с помощью OpenMP»

студента Бородина Артёма Максимовича 2 курса, 19205 группы Направление 09.03.01 – «Информатика и вычислительная техника»

> Преподаватель: к.т.н, доцент А.Ю. Власенко

СОДЕРЖАНИЕ

<u>ЦЕЛЬ</u>	
ЗАДАНИЕ	
ОПИСАНИЕ РАБОТЫ	
ЗАКЛЮЧЕНИЕ	
<u>Приложение 1.</u> Код последовательной программы	
<u>Приложение 2.</u> Код программы вар.1	9
Приложение 3. Код программы вар.2	12
Приложение 4. Исследования #pragma omp schedule()	16
Приложение 5. Замеры времени	17

ЦЕЛЬ

1. Решение СЛАУ, используя ОрепМР.

ЗАДАНИЕ

- 1. Последовательную программу из лабораторной работы 1, реализующую итерационный алгоритм решения системы линейных алгебраических уравнений вида Ax=b, распараллелить с помощью OpenMP. Реализовать два варианта программы:
 - Вариант 1: для каждого распараллеливаемого цикла создается отдельная параллельная секция **#pragma omp parallel for**,
 - Вариант 2: создается одна параллельная секция **#pragma omp parallel**, охватывающая весь итерационный алгоритм Уделить внимание тому, чтобы при запуске программы на различном числе OpenMP-потоков решалась одна и та же задача (исходные данные заполнялись одинаковым образом).
- 2. Замерить время работы двух вариантов программы при использовании различного числа процессорных ядер: от 1 до числа доступных в узле. Построить графики зависимости времени работы программы, ускорения и эффективности распараллеливания от числа используемых ядер. Исходные данные и параметры задачи подобрать таким образом, чтобы решение задачи на одном ядре занимало не менее 30 секунд.
- 3. Провести исследование на определение оптимальных параметров **#pragma omp for schedule(...)** при некотором фиксированном размере задачи и количестве потоков.
- 4. На основании полученных результатов сделать вывод о целесообразности использования первого или второго варианта программ

ОПИСАНИЕ РАБОТЫ

- 1. Были написаны варианты программы (<u>Приложение 1</u>, <u>Приложение 2</u>, <u>Приложение 3</u>), использующую векторные расширения и ОрепМР для решения СЛАУ методом сопряженных градиентов.
- 2. Были выбраны размер матрицы и точность, при которых время работы программы занимало примерно 30 секунд.
- 3. Было исследованы оптимальные параметры **#pragma omp schedule()** (Приложение 4). По полученным данным был сделан вывод что оба варианта работают наилучшим образом при выборе *guided* и размера куска равному размеру матрицы делённое на кол-во потоков. П размере куска меньше время работы увеличивается из-за того, что появляются дополнительные затраты на организацию работы параллельного кода. А при большем размере куска время увеличивается из-за того, что некоторые потоки бездействуют, ведь им не досталось куска.
- 4. Были проведены замеры времени работы программы на разном количестве потоков. По полученным данным был сделан график (Приложение 5). На графике мы можем заметить, что второй вариант программы работает примерно в 2 раза быстрее. Это происходит из-за того, что не тратится время на разделение на потоки и соединение обратно, так как есть только один параллельный участок, затрагивающий весь итерационный алгоритм, в отличии с первым вариантом, где происходит постоянное разделение на потоки перед каждым циклом.

ЗАКЛЮЧЕНИЕ

В ходе лабораторной работы были получены знания о разделении программы на потоки с использованием векторных расширений для более эффективного использования доступных ресурсов. Проанализировав обе лабораторные работы можно сделать вывод, что на системах с общей памятью OpenMP эффективнее, чем MPI.

Приложение 1. Код последовательной программы

```
#include <cmath>
oid matVecMul(const double *mat, const double *vec, int N, double *newVec) {
    <u>_m128d</u> vA = _mm_setzero_pd();
    vA += _mm_loadu_pd(&mat[i * N + j]) * _mm_loadu_pd(&vec[j]);
  newVec[i] = _mm_hadd_pd(vA, vA)[0];
oid mulByConst(const double *vec, double c, int size, double *newVec) {
  _mm_storeu_pd(&newVec[i], _mm_loadu_pd(&vec[i]) * _mm_set1_pd(c));
void subVec(const double *vec1, const double *vec2, int size, double *newVec) {
   void sumVec(const double *vec1, const double *vec2, int size, double *newVec) {
   _mm_storeu_pd(&newVec[i], _mm_loadu_pd(&vec1[i]) + _mm_loadu_pd(&vec2[i]));
double dotProduct(const double *vec1, const double *vec2, int size) {
  <u>_m128d</u> vA = _mm_setzero_pd();
  vA += _mm_loadu_pd(&vec1[i]) * _mm_loadu_pd(&vec2[i]);
 return _mm_hadd_pd(vA, vA)[0];
void printMat(double *mat, int rows, int columns, std::ostream &stream) {
  stream << std::endl;
double *solveSLAE(const double *A, double *b, int N) {
 auto *solution = new double[N]; // xn+1
 auto *prevSolution = new double[N]; // xn
 std::fill(prevSolution, prevSolution + N, 0);
```

```
auto *rNext = new double[N];
auto *alphaZ = new double[N];
auto *betaZ = new double[N];
double alpha;
double beta;
const double EPSILON = 1e-007;
double normb = sqrt(dotProduct(b, b, N));
bool diverge = false;
int divergeCount = 0;
int rightAnswerRepeat = 0;
while (res > EPSILON || rightAnswerRepeat < 5) {</pre>
 if (res < EPSILON) {
    ++rightAnswerRepeat;
   rightAnswerRepeat = 0;
  matVecMul(A, prevSolution, N, Atmp);
  subVec(b, Atmp, N, r);
  matVecMul(A, z, N, Atmp);
  alpha = dotRR / dotProduct(Atmp, z, N);
  mulByConst(z, alpha, N, alphaZ);
  matVecMul(A, alphaZ, N, Atmp);
  subVec(r, Atmp, N, rNext);
  beta = dotProduct(rNext, rNext, N) / dotRR;
  mulByConst(z, beta, N, betaZ);
  sumVec(rNext, betaZ, N, zNext);
  res = sqrt(dotRR) / normb;
  if (prevRes < res || res == INFINITY || res == NAN) {</pre>
    ++divergeCount;
   if (divergeCount > 10 || res == INFINITY || res == NAN) {
      diverge = true;
    divergeCount = 0;
  prevRes = res;
```

```
_mm_storeu_pd(&prevSolution[i], _mm_loadu_pd(&solution[i]));
    _mm_storeu_pd(&z[i], _mm_loadu_pd(&zNext[i]));
delete[](prevSolution);
delete[](Atmp);
delete[](rNext);
delete[](zNext);
delete[](alphaZ);
delete[](betaZ);
if (diverge) {
nt main(int argc, char *argv[]) {
const std::string &fileName = argv[2];
std::ofstream fileStream(fileName);
if (!fileStream) {
fileStream << "Matrix size: " << N << std::endl;
auto *b = new double[N];
auto *u = new double[N];
  u[i] = sin(2 * M_PI * i / double(N));
matVecMul(A, u, N, b);
double *solution = solveSLAE(A, b, N);
auto duration = duration_cast<nanoseconds>(endTime - startTime);
```

```
fileStream << "Answer:" << std::endl;
printMat(u, 1, N, fileStream);
fileStream << "SLAE solution:" << std::endl;
printMat(solution, 1, N, fileStream);
fileStream << "Time: " << duration.count() / double(1000000000) << "sec" << std::endl;
std::cout << "Time: " << duration.count() / double(1000000000) << "sec" << std::endl;
} else {
fileStream << "Does not converge" << std::endl;
}

delete[](solution);
delete[](b);
delete[](h);
return 0;
}</pre>
```

Приложение 2. Код программы вар.1

```
#include <immintrin.h>
pragma omp declare reduction(sseSum: __m128d: omp_out += omp_in) initializer (omp_priv = _mm_setzero_pd()#
#define TYPE guided
#define CHUNK chunkSize
#define BASE_CLAUSE default(none) num_threads(threadCount) schedule(TYPE, CHUNK)
int matrixSize = 1;
int threadCount = 1;
using namespace std::chrono;
void matVecMul(const double *mat, const double *vec, int N, double *newVec) {
#pragma omp parallel for shared(mat, vec, N, newVec, chunkSize)    BASE_CLAUSE
     m128d vA = _mm_setzero_pd();
    vA += _mm_loadu_pd(&mat[i * N + j]) * _mm_loadu_pd(&vec[j]);
   newVec[i] = _mm_hadd_pd(vA, vA)[0];
void mulByConst(const double *vec, double c, int size, double *newVec) {
pragma omp parallel for shared(c, vec, size, newVec, chunkSize) BASE_CLAUSE
   _mm_storeu_pd(&newVec[i], _mm_loadu_pd(&vec[i]) * _mm_set1_pd(c));
oid subVec(const double *vec1, const double *vec2, int size, double *newVec) {
pragma omp parallel for shared(vec1, vec2, size, newVec, chunkSize) BASE_CLAUSE
   void sumVec(const double *vec1, const double *vec2, int size, double *newVec) {
pragma omp parallel for shared(vec1, vec2, size, newVec, chunkSize) BASE_CLAUSE
   _mm_storeu_pd(&newVec[i], _mm_loadu_pd(&vec1[i]) + _mm_loadu_pd(&vec2[i]));
double dotProduct(const double *vec1, const double *vec2, int size) {
  __m128d vA = _mm_setzero_pd();
pragma parallel omp for shared(vec1, vec2, size, sum, chunkSize) reduction(sseSum:sum) BASE_CLAUSE#
   vA += _mm_loadu_pd(&vec1[i]) * _mm_loadu_pd(&vec2[i]);
 return _mm_hadd_pd(vA, vA)[0];
void printMat(double *mat, int rows, int columns, std::ostream &stream) {
```

```
double *solveSLAE(double *A, double *b, int N, std::ostream &stream) {
auto *solution = new double[N]; // xn+1
std::fill(solution, solution + N, 0);
 auto *prevSolution = new double[N]; // xn
 std::fill(prevSolution, prevSolution + N, 0);
 auto *Atmp = new double[N];
 auto *z = new double[N];
 auto *zNext = new double[N];
 auto *alphaZ = new double[N];
 auto *betaZ = new double[N];
 double beta;
 const double EPSILON = 1e-007;
 double normb = sqrt(dotProduct(b, b, N));
 double dotRR;
 bool diverge = false;
 int rightAnswerRepeat = 0;
 while (res > EPSILON || rightAnswerRepeat < 5) {</pre>
  if (res < EPSILON) {
    ++rightAnswerRepeat;
    rightAnswerRepeat = 0;
  matVecMul(A, prevSolution, N, Atmp);
  subVec(b, Atmp, N, r);
#pragma omp parallel for shared(z, r, N, chunkSize) BASE_CLAUSE
    _mm_storeu_pd(&z[i], _mm_loadu_pd(&r[i]));
  matVecMul(A, z, N, Atmp);
  dotRR = dotProduct(r, r, N);
  alpha = dotRR / dotProduct(Atmp, z, N);
  mulByConst(z, alpha, N, alphaZ);
  sumVec(prevSolution, alphaZ, N, solution);
  matVecMul(A, alphaZ, N, Atmp);
   subVec(r, Atmp, N, rNext);
   beta = dotProduct(rNext, rNext, N) / dotRR;
   mulByConst(z, beta, N, betaZ);
   sumVec(rNext, betaZ, N, zNext);
```

```
res = sqrt(dotRR) / normb;
    ++divergeCount;
    if (divergeCount > 10 || res == INFINITY || res == NAN) {
       diverge = true;
    divergeCount = 0;
  prevRes = res;
pragma omp parallel for shared(solution, prevSolution, r, rNext, z, zNext, N, chunkSize) #ASE_CLAUSE#
    _mm_storeu_pd(&prevSolution[i], _mm_loadu_pd(&solution[i]));
    _mm_storeu_pd(&r[i], _mm_loadu_pd(&rNext[i]));
    _mm_storeu_pd(&z[i], _mm_loadu_pd(&zNext[i]));
delete[](prevSolution);
 delete[](Atmp);
 delete[](alphaZ);
 delete[](betaZ);
nt main(int argc, char *argv[]) {
if (argc != 5) {
int N = atoi(argv[1]);
 threadCount = atoi(argv[2]);
 chunkSize = matrixSize / threadCount * atof(argv[4]);
 std::cout << "chunkSize: " << chunkSize << std::endl;
 const std::string &fileName = argv[3];
 if (!fileStream) {
  std::cout << "error with output file" << std::endl;
 auto *u = new double[N]
```

Приложение 3. Код программы вар.2

```
#include <immintrin.h>
pragma omp declare reduction(sseSum: __m128d: omp_out += omp_in) initializer (omp_priv = _mm_setzero_pd()#
int threadCount = 1;
double dotProduct(const double *vec1, const double *vec2, int size) {
 __m128d vA = _mm_setzero_pd();
   vA += _mm_loadu_pd(&vec1[i]) * _mm_loadu_pd(&vec2[i]);
 return _mm_hadd_pd(vA, vA)[0];
void printMat(double *mat, int rows, int columns, std::ostream &stream) {
   stream << std::endl;
double *solveSLAE(const double *A, double *b, int N) {
 std::fill(solution, solution + N, 0);
 auto *prevSolution = new double[N]; // xn
 std::fill(prevSolution, prevSolution + N, 0);
 auto *Atmp = new double[N];
 auto *r = new double[N];
 auto *z = new double[N];
 auto *rNext = new double[N];
 auto *alphaZ = new double[N];
 auto *betaZ = new double[N];
 double alpha;
 double beta;
 const double EPSILON = 1e-007;
 double normb = sqrt(dotProduct(b, b, N));
 double dotRR;
 _{m128d} vA = _{mm_setzero_pd()};
 __m128d vB = _mm_setzero_pd();
 bool diverge = false;
 int divergeCount = 0;
```

```
int rightAnswerRepeat = 0;
pragma omp parallel num_threads(threadCount) firstprivate(rightAnswerRepeat, divergeCount, diverge#
 while (res > EPSILON || rightAnswerRepeat < 5) {
#pragma omp single
    if (res < EPSILON) {
      ++rightAnswerRepeat;
      rightAnswerRepeat = 0;
#pragma omp for schedule(TYPE, CHUNK)
      <u>_m128d</u> sum = _mm_setzero_pd();
      sum += _mm_loadu_pd(&A[i * N + j]) * _mm_loadu_pd(&prevSolution[j]);
     Atmp[i] = _mm_hadd_pd(sum, sum)[0];
#pragma omp for schedule(TYPE, CHUNK)
   for (int i = 0; i < N; i += 2) {
     _mm_storeu_pd(&r[i], _mm_loadu_pd(&b[i]) - _mm_loadu_pd(&Atmp[i]));
     _mm_storeu_pd(&z[i], _mm_loadu_pd(&r[i]));
#pragma omp for schedule(TYPE, CHUNK)
      __m128d sum = _mm_setzero_pd();
      sum += _mm_loadu_pd(&A[i * N + j]) * _mm_loadu_pd(&z[j]);
     Atmp[i] = _mm_hadd_pd(sum, sum)[0];
#pragma omp single
     vA = _mm_setzero_pd();
#pragma omp for schedule(TYPE, CHUNK) reduction(sseSum: vA, vB)
   for (int i = 0; i < N; i += 2) {
    vA += _mm_loadu_pd(&r[i]) * _mm_loadu_pd(&r[i]);
     vB += _mm_loadu_pd(&Atmp[i]) * _mm_loadu_pd(&z[i]);
#pragma omp single
     dotRR = _mm_hadd_pd(vA, vA)[0];
     alpha = dotRR / _mm_hadd_pd(vB, vB)[0];
#pragma omp for schedule(TYPE, CHUNK)
     _mm_storeu_pd(&alphaZ[i], _mm_loadu_pd(&z[i]) * _mm_set1_pd(alpha));
     _mm_storeu_pd(&solution[i], _mm_loadu_pd(&prevSolution[i]) + _mm_loadu_pd(&alphaZ[i]));
#pragma omp for schedule(TYPE, CHUNK)
     __m128d sum = _mm_setzero_pd();
      sum += _mm_loadu_pd(&A[i * N + j]) * _mm_loadu_pd(&alphaZ[j]);
```

```
Atmp[i] = _mm_hadd_pd(sum, sum)[0];
#pragma omp single
     vA = _mm_setzero_pd();
#pragma omp for schedule(TYPE, CHUNK) reduction(sseSum: vA)
     \_mm\_storeu\_pd(\&rNext[i], \_mm\_loadu\_pd(\&r[i]) - \_mm\_loadu\_pd(\&Atmp[i]));
     vA += _mm_loadu_pd(&rNext[i]) * _mm_loadu_pd(&rNext[i]);
#pragma omp single
     beta = _mm_hadd_pd(vA, vA)[0] / dotRR;
#pragma omp for schedule(TYPE, CHUNK)
     _mm_storeu_pd(&betaZ[i], _mm_loadu_pd(&z[i]) * _mm_set1_pd(beta));
     _mm_storeu_pd(&zNext[i], _mm_loadu_pd(&rNext[i]) + _mm_loadu_pd(&betaZ[i]));
#pragma omp single
     res = sqrt(_mm_hadd_pd(vA, vA)[0]) / normb;
   if (prevRes < res || res == INFINITY || res == NAN) {
#pragma omp single
    if (divergeCount > 10 || res == INFINITY || res == NAN) {
      diverge = true;
#pragma omp single
       divergeCount = 0;
#pragma omp single
#pragma omp for schedule(TYPE, CHUNK)
   for (long i = 0; i < N; i += 2) {
     _mm_storeu_pd(&prevSolution[i], _mm_loadu_pd(&solution[i]));
     _mm_storeu_pd(&r[i], _mm_loadu_pd(&rNext[i]));
     _mm_storeu_pd(&z[i], _mm_loadu_pd(&zNext[i]));
 delete[](prevSolution);
 delete[](alphaZ);
 delete[](betaZ);
```

```
std::cout << "iterCount: " << iterCount << std::endl;
if (diverge) {
nt main(int argc, char *argv[]) {
if (argc != 5) {
int N = atoi(argv[1]);
threadCount = atoi(argv[2]);
chunkSize = matrixSize / threadCount * atof(argv[4]);
const std::string &fileName = argv[3];
if (!fileStream) {
fileStream << "Matrix size: " << N << " thread num: " << threadCount << std::endl;
auto *b = new double[N];
auto *u = new double[N];
  u[i] = sin(2 * M_PI * i / double(N));
    _{m128d} vA = _{mm_{setzero_{pd}}}
    vA += _mm_loadu_pd(&A[i * N + j]) * _mm_loadu_pd(&u[j]);
  b[i] = _mm_hadd_pd(vA, vA)[0];
auto startTime = system_clock::now();
double *solution = solveSLAE(A, b, N);
auto endTime = system_clock::now();
auto duration = duration_cast<nanoseconds>(endTime - startTime);
  printMat(u, 1, N, fileStream);
  printMat(solution, 1, N, fileStream);
  fileStream << "Time: " << duration.count() / double(1000000000) << "sec" << std::endl;
```

```
std::cout << "Time: " << duration.count() / double(1000000000) << "sec" << std::endl;
} else {
    fileStream << "Does not converge" << std::endl;
}

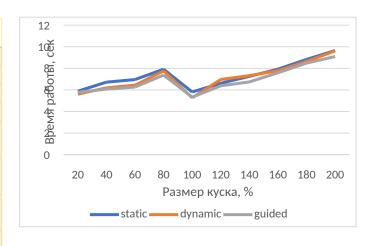
delete[](solution);
delete[](b);
delete[](A);
return 0;
}</pre>
```

Приложение 4. Исследования #pragma omp schedule()

За 100% размер куска берется N/кол-во потоков, где N - размер матрицы.

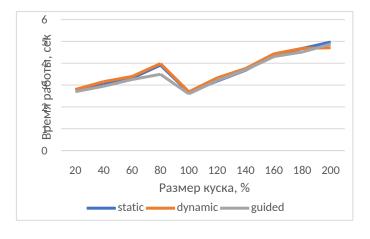
Bap.1:

ъар.	static	dynamic	guided
	Static	aynanne	Juliucu
20	5,88306	5,60548	5,77482
40	6,7197	6,19513	6,0746
60	6,96117	6,43523	6,26761
80	7,91898	7,76932	7,37753
10	5,80743	5,2682	5,33303
0			
12	6,61919	6,97014	6,38288
0			
14	7,25941	7,32339	6,74216
0			
16	7,92514	7,78371	7,58574
0			
18	8,82994	8,64281	8,48732
0			
20	9,65195	9,61938	9,09597
0			



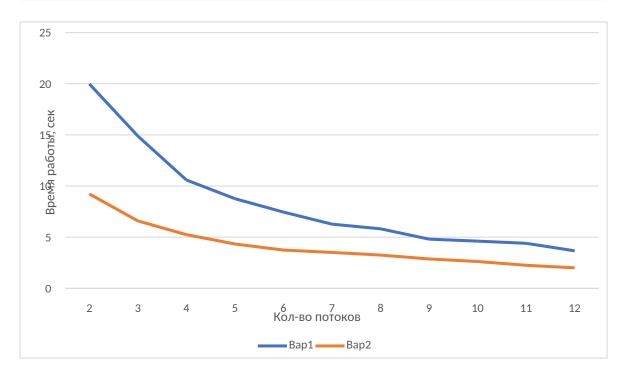
Bap.2:

	static	dynamic	guided
20	2,7910 3	2,79734	2,6994 2
40	3,0653 4	3,15653	2,9489 2
60	3,3191 7	3,39407	3,2567 3
80	3,9141 7	3,9814	3,4933 1
100	2,6642 5	2,69003	2,5973 4
120	3,1838 6	3,32867	3,2108 8
140	3,6776 1	3,75452	3,6957 9
160	4,3715 1	4,43063	4,3055 8
180	4,6738 5	4,68074	4,5142 6
200	4,9686 4	4,70993	4,8728 6



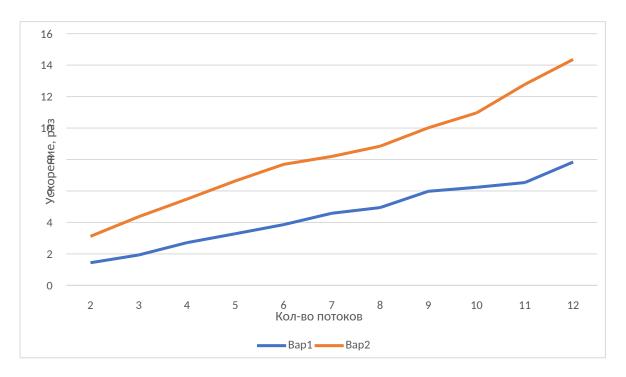
Приложение 5. Замеры времени

Время работы, сек											
	2	3	4	5	6	7	8	9	10	11	12
Вар											
1	19,98	14,88	10,59	8,76	7,45	6,27	5,82	4,81	4,61	4,40	3,67
Вар											
2	9,23	6,59	5,25	4,33	3,74	3,51	3,25	2,87	2,62	2,25	2,00



Время работы последовательной программы: 28,7534с

	Ускорение, раз											
	2	3	4	5	6	7	8	9	10	11	12	
Вар 1	1,44	1,93	2,72	3,28	3,86	4,58	4,94	5,98	6,23	6,53	7,84	
Вар 2	3,12	4,37	5,48	6,64	7,69	8,19	8,84	10,02	10,96	12,77	14,36	



Эффективность на поток, %											
	2	3	4	5	6	7	8	9	10	11	12
Вар											
1	0,72	0,64	0,68	0,66	0,64	0,65	0,62	0,66	0,62	0,59	0,65
Вар											
2	1,56	1,46	1,37	1,33	1,28	1,17	1,11	1,11	1,10	1,16	1,20

