1 Large array benchmark

1.1 Introduction

The size of data that must be analyzed keeps increasing year after year and the prize for DRAM are not getting cheaper. NVDIMM offer a lot of storage at a cheaper prize. This opens the opportunity to save money by offloading some of the data to the NVDIMM where the data will be analyzed the same way as the data on the DRAM. The downside to this strategy is that NVDIMM is slower than DRAM so the question is how much data can be offloaded to NVDIMM. If the user offload too much data to NVDIMM then the threads working on analyzing the data on DRAM will be idle while waiting for NVDIMM threads to complete.

   The goal of this benchmark is to find a formula that will make it possible to estimate how many threads should be delegated to work on the data located on NVDIMM. The methodology must also be able to calculate how much of the data on DRAM should be offloaded to NVDIMM

   There is two version of the code that is being used. They have been programmed slightly different in the way read data. This is to explore which of them is the fastest.

1.2 Formula

In order to use the formula one must measure the NVDIMM and DRAM speed using a benchmark made in a previous chapter where speed was measured when data was transferred from DRAM-DRAM and NVDIMM-NVDIMM simultaneously with different amount threads allocated to the different processes. By using using the speed from the benchmark in the formula below along with the size of the data that will be used in the calculation the user can easily calculate how much data can be transferred to NVDIMM without losing time. The formula only decide how much data can be allocated to NVDIMM with a certain amount of threads. This means that the user must probably use the formula several times where the number of NVDIMM threads varies from one to five in order to find the best combination of threads and data allocated to the NVDIMM process.

Formula
$$\frac{Total\_data - nvdimm\_data}{dram\_speed} = \frac{nvdimm\_data}{nvdimm\_speed}$$

$$nvdimm\_data = \frac{nvdimm\_speed * Total\_data}{nvdimm\_speed + dram\_speed}$$

Calculation

I have created a benchmark that will test this formula to see if it is accurate. This benchmark has an two dimensional array filled with data. The benchmark start at element (1,1) of the array where it sum ups all of its eight neighbors and then takes the average. The result is stored in the same position in another two dimensional array. The benchmark does this for every element between (1,1) and (m-2,n-2). The benchmark repeats this process ten times and after each time the benchmark will swap both the DRAM arrays and NVDIMM array. The time is measured at the beginning of the process and at the end, this time is called total_time in the code. Each thread will also measure the time they takes to complete their own tasks, in the code this is called individual_time.

In listing 1 is an example of the serial code of the benchmark. The array has not been divided into two. The thread will will repeat the while-loop on line 1 k_length amount of time. It will start the time measurement at line 2. From line 3-10 the thread will pass through the the 2D-array row by row starting from element (1,1) and end at element (m-2,n-2). At each element the thread will calculate the average of all the elements neighbors, eight neighbors in total. The thread will then stop the time measurement at line 11 and swap the arrays at line 12-14. It will also increase k by one and start over if k is still lower than k_length.

Listing 1: Serial code

```
1  while(k<K_length){
2    total_time[k] = mysecond();
3    for( i=1; i<m-2; i++){
4      for( j=1; j<n-2; j++){
5        temp = A[i-1][j-1] + A[i-1][j] + A[i-1][j+1]+
6               A[i][j-1]   +         A[i][j+1]+
7               A[i+1][j-1] + A[i+1][j] + A[i+1][j+1];
8        B[i][j] = temp/8;
9      }
10   }
11   total_time[k] = mysecond() - total_time[k];
12   temp_array = A;
```

```
13    A = B;
14    B = temp_array;
15    k++;
16  }
```

Distribution of m

1.3 First version

There are two groups of threads that works in parallel in this program.
The first group of threads works on the part of the data that is stored
on DRAM and the other works on the data stored on NVDIMM. One
thread in each group works on data that borders with the other group.
In the DRAM group that is the thread with the highest thread_id. Each
of the elements in the last row of data will have three neighbors that
exist on the NVDIMM side. This means that the thread must access
the NVDIMM in order to get the data. The NVDIMM thread with the
lowest thread_id also have elements in the first row of data that have
three neighbours that exist in DRAM that must be accessed by the
thread directly.

Explaination of code:
The code below only shows the calculation process, it does not the rest
of the code. Allocation of memory have been done by all the threads,
as a result the data have been spread across all the memory channels.
The data is a 2d array where the rows of data on DRAM will be divided
equally between the DRAM threads, the rows of data on NVDIMM will
also be divided equally between the NVDIMM threads. The variable
slice_start hold the index of the row where the tread must start at and
slice_end holds the index of the row the thread must stop at. Array A
and B are DRAM array and array C and D are NVDIMM arrays. The
average found by adding together eight neighbors in A will be placed
in same position in B. The same is true for C and D.

The process are repeated K_length amount of time, usually ten
times in my tests. One way of knowing if the test result are correct is to
run the test several time and look for consistency. The code measures
the time taken to complete one iteration of calculation, this is done in
the beginning of the code at line 5 and at the end at line 84 by a single
thread. All the threads then get divided into the DRAM and NVDIMM
at line 8. If the thread_id of a thread is less than the dram_threads is

3

it will do calculation on the data in DRAM, the rest will fail the if test and move on to the else bracket at line 42. Dram_threads is the total amount of threads that will be working on data in DRAM.

At line 11 the thread with the highest thread_id will pass the if test and the rest will move on to line 30. The thread with highest thread_id will then measure time at line 12 and end the measurement in line 29, this is the start and the end of the bracket. The thread will then enter a double for-loop at line 13-20 that will go through elements from position (slice_start,1) until (slice_end-1,n-1), this leaves out the last row assigned to the tread, that row will be dealt with later. At each element the for-loop it will add all of its eight neighbors together at line 15-17 and divide by eight at line 18. The thread will then enter a new for-loop at line 23, this for-loop will calculate average of the last row on DRAM. Elements of this row have three neighbors that exist in NVDIMM. The thread will access the NVDIMM directly when adding the eight neighbors at line 24-26. Data on NVDIMM are accessed by the thread at line 26, the thread is using a library developed for this purpose.

For all the other DRAM threads that jumped to line 30 will start by taking time measurement at the beginning and at the end of the bracket at line 31 and 40. The code from line 32-39 is identical to line 13-20 describe before.

The group of NVDIMM enters the else bracket at line 42 where the thread with the lowest thread_id will pass the if-sentence at line 44, the rest will move on to the else bracket at line 62. The thread will then measure time at line 45 and end the measurement in line 61. It will then enter a for-loop at line 47 and will begin calculating the average of the neighbors of the elements in the first row. The first row have three of its eight neighbors in the row above and they exist in the DRAM. Once done the thread will move on to a new for-loop at line 53. This for-loop will go through the rest of the portion of data the thread have been given and calculate the average of each elements neighbors.

The rest of the NVDIMM threads will move into the else bracket at line 62. The code here is very similar to the code at 31-40 that has been described at a previous paragraph. The only difference is that the code at line 66-68 where the code reads from NVDIMM instead DRAM.

All the threads will wait a barrier at line 75 until all threads are done. After that on thread will enter a single bracket where array A and B will swap places, array C and D will also swap places. The time it took for this one iteration will be registered at line 84. After this the code will move back to line 1.

Listing 2: First version

```
1  while(k<K_length){
2    #pragma omp barrier
3    #pragma omp single
4    {
5      total_time[k] = mysecond();
6    }
7    //Divides threads into DRAM threads and NVDIMM threads.
8    if( thread_id < dram_threads ){
9
10     //for the thread bordering on NVDIMM thread.
11     if( thread_id==(dram_threads-1) ){
12       individual_time[k][thread_id] = mysecond();
13       for( i=slice_start; i<slice_end-1; i++){
14         for( j=1; j<nMinusOne; j++){
15           temp = A[i-1][j-1] + A[i-1][j] + A[i-1][j+1]+
16                  A[i][j-1]    +        A[i][j+1]+
17                  A[i+1][j-1] + A[i+1][j] + A[i+1][j+1];
18           B[i][j] = temp*inverseEigth;
19         }
20       }
21
22       i = slice_end-1;
23       for( j=1; j<nMinusOne; j++){
24         temp = A[i-1][j-1] + A[i-1][j] + A[i-1][j+1]+
25                A[i][j-1]    +        A[i][j+1]+
26                D_RO(C)[i*n+j] + D_RO(C)[i*n+j] +
27                  D_RO(C)[i*n+j];
27         B[i][j] = temp*inverseEigth;
28         }
29       individual_time[k][thread_id] = mysecond() -
                individual_time[k][thread_id];
30     }else{
31       individual_time[k][thread_id] = mysecond();
32       for( i=slice_start; i<slice_end; i++){
33         for( j=1; j<nMinusOne; j++){
34           temp = A[i-1][j-1] + A[i-1][j] + A[i-1][j+1]+
35                  A[i][j-1]    +        A[i][j+1]+
36                  A[i+1][j-1] + A[i+1][j] + A[i+1][j+1];
37           B[i][j] = temp*inverseEigth;
38         }
39       }
40       individual_time[k][thread_id] = mysecond() -
```

```
                  individual_time[k][thread_id];
41       }
42    }else{
43        //for the thread bordering on DRAM thread.
44        if( thread_id==dram_threads ){
45          individual_time[k][thread_id] = mysecond();
46          i=0;
47          for( j=1; j<nMinusOne; j++){
48            temp =
                  A[dram_part-1][j-1]+A[dram_part-1][j]+A[dram_part-1][j+1]+
49              D_RO(C)[i*n+(j-1)]      +
                  D_RO(C)[i*n+(j+1)]+
50              D_RO(C)[(i+1)*n+(j-1)] + D_RO(C)[(i+1)*n+j]
                  + D_RO(C)[(i+1)*n+(j+1)];
51            D_RW(D)[i*n+j] = temp*inverseEigth;
52          }
53          for( i=slice_start+1; i<slice_end-1; i++){
54            for( j=1; j<nMinusOne; j++){
55              temp = D_RO(C)[(i-1)*n+(j-1)] +
                    D_RO(C)[(i-1)*n+j] + D_RO(C)[(i-1)*n+(j+1)]+
56                D_RO(C)[i*n+(j-1)]      +
                    D_RO(C)[i*n+(j+1)]+
57                D_RO(C)[(i+1)*n+(j+1)] +
                    D_RO(C)[(i+1)*n+j] +
                    D_RO(C)[(i+1)*n+(j+1)];
58              D_RW(D)[i*n+j] = temp*inverseEigth;
59            }
60          }
61          individual_time[k][thread_id] = mysecond() -
                individual_time[k][thread_id];
62        }else{
63          individual_time[k][thread_id] = mysecond();
64          for( i=slice_start; i<slice_end; i++){
65            for( j=1; j<nMinusOne; j++){
66              temp = D_RO(C)[(i-1)*n+(j-1)] +
                    D_RO(C)[(i-1)*n+j] + D_RO(C)[(i-1)*n+(j+1)]+
67                D_RO(C)[i*n+(j-1)]      +
                    D_RO(C)[i*n+(j+1)]+
68                D_RO(C)[(i+1)*n+(j-1)] +
                    D_RO(C)[(i+1)*n+j] +
                    D_RO(C)[(i+1)*n+(j+1)];
69              D_RW(D)[i*n+j] = temp*inverseEigth;
70            }
```

```
71          }
72          individual_time[k][thread_id] = mysecond() -
                individual_time[k][thread_id];
73        }
74      }
75      #pragma omp barrier
76      #pragma omp single
77      {
78        tempArray = B;
79          B=A;
80          A=tempArray;
81        temp_nvdimm = C;
82        C = D;
83        D = temp_nvdimm;
84        total_time[k] = mysecond() - total_time[k];
85        k++;
86      }
87      #pragma omp barrier
88  }//End of while
```

Table 1 shows the calculation of how much data must be allocated
to NVDIMM in order for the DRAM and NVDIMM to complete their
tasks simultaneously. M in row one in table shows how many rows
the 2D-array has. The n in row two shows how many elements each
row has. The total MB in row three is calculated in following manner,
m*n/(8*1000000). Row five is the beginning of five column. First col-
umn shows how many threads are used to calculate data on NVDIMM.
Second and third column is the DRAM and NVDIMM speed in MB per
second. The speeds comes from a benchmark described in a previous
chapter where data is transferred from DRAM-DRAM and NVDIMM-
NVDIMM simultaneously. Column four uses the formula described in
the beginning of the chapter, the numbers are in MB. The last column
converts the result in fourth column into number of rows of the 2D-
array that will be placed on NVDIMM.

Each row in table two is a result of one test. In table two the first
two column shows m and n. Third column shows how many rows are
assigned to NVDIMM. Fourth and fifth column shows the number of
DRAM and NVDIMM threads the test will have. Column six shows
the average speed of all the DRAM threads in the test. The test is re-
peated ten times so if eleven DRAM threads then there are 99 DRAM
threads that will be taken average of. The first test is excluded because

m	100000			
n	100000			
total MB	80,000			
speed				
Nvm-thr	dram	nvm	nvmpart	rows
1	63003.00	3036.04	3677.88	2299
2	60302.95	6032.18	7274.80	4547
3	58967.56	8404.24	9979.54	6237
4	54617.41	10551.97	12953.28	8096
5	55545.81	13330.65	15483.55	9677
6	52102.34	18159.78	20676.61	12923

Table 1: First version, distribution

the times are way higher then all the tests that comes after. The two next column is the dram minimum and dram maximum. Dram minimum is found by first finding the fastest DRAM thread in each of the nine tests and then take the average of them. Dram maximum is calculated the same way as dram minimum, The only difference is that this is for the slowest speed. Column nine, ten and eleven shows nvdimm average, nvdimm minimum and nvdimm maximum. These column are similar to dram average, dram minimum and dram maximum. The different is that the times is for the nvdimm threads. The last three column is the total average, total minimum and total maximum. The total time is only measured once for each test so total minimum and total maximum shows the fastest and slowest test. Total average is the average of all the tests except the first test.

Thread	Prediction DRAM Speed	First version DRAM Size MB	DRAM Pred. Time	DRAM Time	NVDIMM Speed	NVDIMM Part MB	NVDIMM Pred. Time	NVDIMM Time	Size of n in MB
1	61568.86	7814.69	0.2539	0.2717	1433.87	183.90	0.2565	0.3389	0.2530
2	58704.29	7634.84	0.2601	0.2742	5264.15	363.76	0.1382	0.3397	0.2530
3	55454.72	7499.76	0.2705	0.2780	7421.89	498.84	0.1344	0.3132	0.2530
4	52112.79	7351.02	0.2821	0.2849	10193.13	647.58	0.1271	0.3081	0.2530
5	48564.56	7224.54	0.2975	0.2930	13010.17	774.06	0.1190	0.2999	0.2530
6	45866.16	6965.00	0.3037	0.3013	16101.65	1033.59	0.1284	0.3261	0.2530
1	61568.86	15612.65	0.5072	0.6004	1433.87	386.38	0.5389	0.7711	0.3578
2	58704.29	15271.70	0.5203	0.5862	5264.15	727.33	0.2763	0.7454	0.3578
3	55454.72	15001.23	0.5410	0.6190	7421.89	997.79	0.2689	0.6954	0.3578
4	52112.79	14703.94	0.5643	0.6159	10193.13	1295.09	0.2541	0.6789	0.3578
5	48564.56	14450.64	0.5951	0.6128	13010.17	1548.39	0.2380	0.6653	0.3578
6	45866.16	13931.53	0.6075	0.6562	16101.65	2067.50	0.2568	0.7109	0.3578
1	61568.86	31268.78	1.0157	1.3334	1433.87	735.72	1.0262	1.5873	0.5060
2	58704.29	30549.24	1.0408	1.3584	5264.15	1455.26	0.5529	1.6423	0.5060
3	55454.72	30008.33	1.0823	1.4043	7421.89	1996.17	0.5379	1.6386	0.5060
4	52112.79	29413.27	1.1288	1.4622	10193.13	2591.23	0.5084	1.6775	0.5060
5	48564.56	28907.27	1.1905	1.5704	13010.17	3097.23	0.4761	1.6138	0.5060
6	45866.16	27868.46	1.2152	1.5622	16101.65	4136.04	0.5137	2.3127	0.5060
1	61568.86	62539.15	3.0473	2.9340	1433.87	1471.27	3.0783	4.1249	0.7156
2	58704.29	61100.07	3.1224	3.0671	5264.15	2910.35	1.6586	4.5790	0.7156
3	55454.72	60018.09	3.2469	3.0477	7421.89	3992.33	1.6137	3.6125	0.7156
4	52112.79	58828.04	3.3866	3.2977	10193.13	5182.38	1.5253	4.5699	0.7156
5	48564.56	57816.19	3.5715	3.2531	13010.17	6194.23	1.4283	4.8186	0.7156
6	45866.16	55738.08	3.6457	3.1883	16101.65	8272.34	1.5413	5.3689	0.7156
1	61568.86	78160.80	3.8085	3.9240	1433.87	1839.20	3.8481	4.9424	0.8000
2	58704.29	76362.40	3.9024	4.0606	5264.15	3637.60	2.0730	7.7780	0.8000
3	55454.72	75010.40	4.0579	4.0350	7421.89	4989.60	2.0168	5.5041	0.8000
4	52112.79	73523.20	4.2325	4.2847	10193.13	6476.80	1.9062	6.0299	0.8000
5	48564.56	72258.40	4.4636	4.2282	13010.17	7741.60	1.7851	5.3086	0.8000
6	45866.16	69661.60	4.5564	4.2876	16101.65	10338.40	1.9262	5.5472	0.8000

Table 2: First version, Prediction

| m	n	nvdimm length	dram threads	nvdimm threads	dram average	dram min	dram max	nvdimm average	nvdimm min	nvdimm max	total average	total min	total max
31620	31620	727	15	1	0.2658	0.2611	0.2717	0.3387	0.3386	0.3389	0.3388	0.3383	0.3392
31620	31620	1438	14	2	0.2679	0.2631	0.2742	0.3144	0.2894	0.3397	0.3395	0.3375	0.3422
31620	31620	1972	13	3	0.2729	0.2691	0.2780	0.2814	0.2653	0.3132	0.3132	0.3128	0.3134
31620	31620	2560	12	4	0.2785	0.2738	0.2849	0.2697	0.2550	0.3081	0.3079	0.3077	0.3081
31620	31620	3060	11	5	0.2864	0.2807	0.2930	0.2570	0.2445	0.2999	0.2997	0.2984	0.3010
31620	31620	4086	10	6	0.2955	0.2896	0.3013	0.2805	0.2707	0.3261	0.3261	0.3257	0.3267
44720	44720	1080	15	1	0.5573	0.5319	0.6004	0.7709	0.7057	0.7711	0.7718	0.7037	0.9598
44720	44720	2033	14	2	0.5549	0.5279	0.5862	0.6936	0.5974	0.7454	0.7454	0.6725	1.0010
44720	44720	2789	13	3	0.5665	0.5300	0.6190	0.6114	0.5620	0.6954	0.6954	0.6317	0.9309
44720	44720	3620	12	4	0.5755	0.5361	0.6159	0.5805	0.5313	0.6789	0.6758	0.6131	0.9672
44720	44720	4328	11	5	0.5834	0.5417	0.6128	0.5446	0.5008	0.6653	0.6686	0.5981	0.9613
44720	44720	5779	10	6	0.5995	0.5454	0.6562	0.5930	0.5584	0.7109	0.7206	0.6497	0.9279
63250	63250	1454	15	1	1.25251	1.22963	1.3334	1.5873	1.42011	1.5873	1.5873	1.4201	1.9038
63250	63250	2876	14	2	1.25697	1.22827	1.3584	1.5066	1.25703	1.6423	1.6427	1.3948	2.1539
63250	63250	3945	13	3	1.26384	1.23511	1.4043	1.4083	1.17484	1.6386	1.6001	1.291	2.3742
63250	63250	5121	12	4	1.27431	1.24472	1.4622	1.3695	1.14489	1.6775	1.6354	1.2794	2.2677
63250	63250	6121	11	5	1.31668	1.2662	1.5704	1.27	1.09073	1.6138	1.6805	1.3012	2.1018
63250	63250	8174	10	6	1.32123	1.26102	1.5622	1.4875	1.21667	2.3127	1.8423	1.327	2.7096
89450	89450	2056	15	1	2.8283	2.8024	2.9340	3.1952	3.1952	4.1249	3.1952	2.8860	4.1249
89450	89450	4067	14	2	2.9314	2.7818	3.0671	4.3454	3.2986	4.5790	4.5590	3.3266	5.2170
89450	89450	5579	13	3	2.9126	2.7635	3.0477	2.9296	2.3666	3.6125	3.6302	2.8349	5.5610
89450	89450	7242	12	4	3.0095	2.8432	3.2977	3.6364	2.8651	4.5699	4.3398	3.7687	4.8478
89450	89450	8656	11	5	3.0529	2.8716	3.2531	3.5520	2.6531	4.8186	4.5032	3.5581	5.1694
89450	89450	11560	10	6	3.0446	2.9456	3.1883	3.7067	2.8888	5.3689	4.6017	3.5022	5.7443
100000	100000	2299	15	1	3.7294	3.6904	3.9240	4.5748	4.2794	4.9424	4.5957	3.6680	5.9562
100000	100000	4547	14	2	3.8163	3.7377	4.0606	5.2638	4.5267	7.7780	6.0034	3.8320	7.7781
100000	100000	6237	13	3	3.8600	3.7936	4.0350	4.3152	3.3177	5.5041	5.1849	3.8963	6.5493
100000	100000	8096	12	4	3.8528	3.8004	4.2847	4.7541	3.4915	6.0299	6.0331	5.0625	6.9402
100000	100000	9677	11	5	3.9760	3.7601	4.2282	3.9186	3.0569	5.3086	4.9039	4.2381	5.5243
100000	100000	12923	10	6	3.9310	3.6650	4.2876	4.4501	3.4429	5.5472	5.4688	4.3413	6.0225

Table 3: First version, result
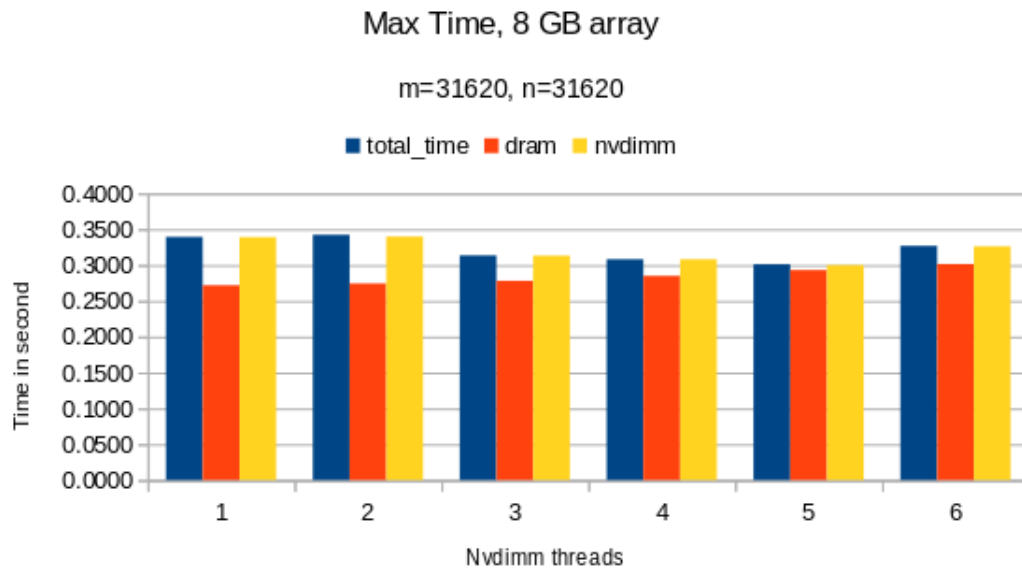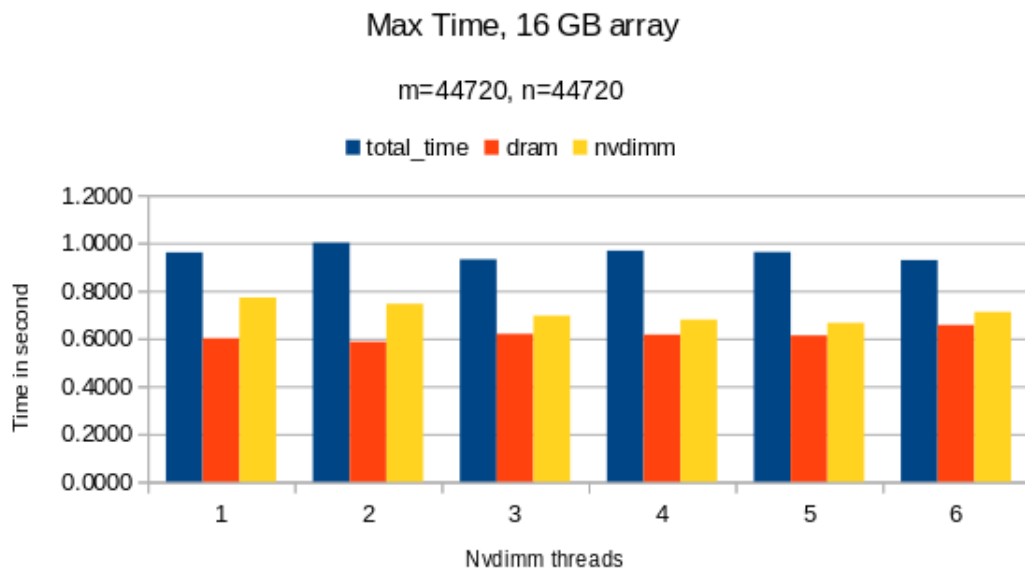
Figure 1: First version, max time
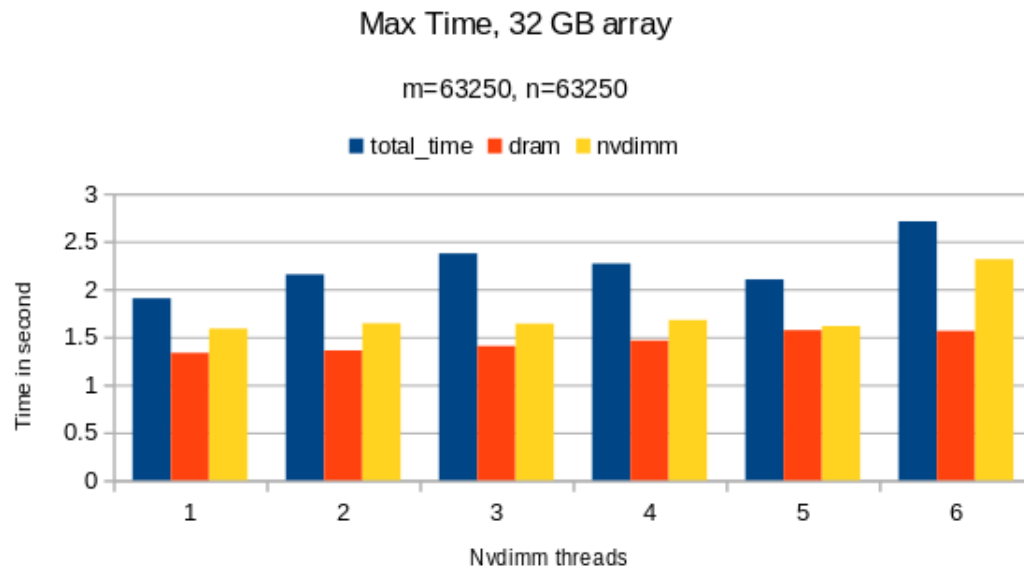

Figure 2: First version, max time
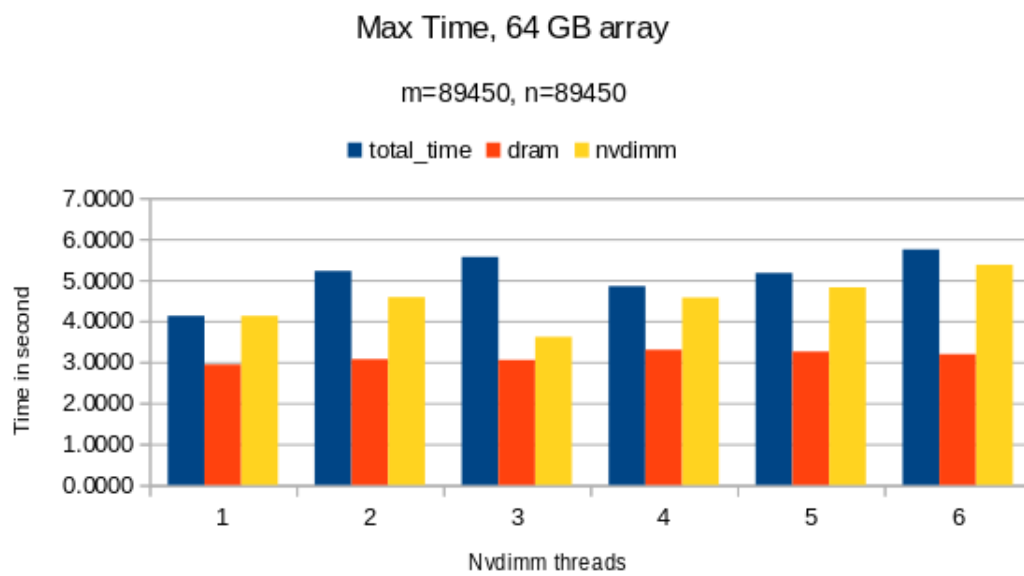
Figure 3: First version, max time

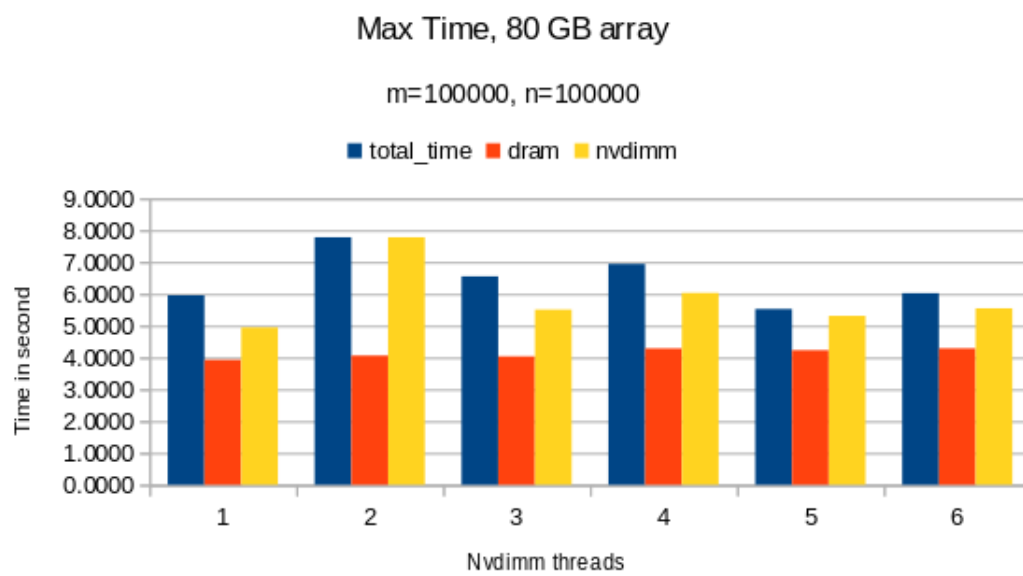Figure 4: First version, max time

Figure 5: First version, max time

The formula have been tested when the total array size are 8 GB, 16 GB, 32 GB, 64 GB and 80 GB. When the workload are 8 GB and 16 GB the workload between DRAM and NVDIMM is almost perfectly balanced. When the workload is increased to 32 GB the NVDIMM time starts to become slower than DRAM. The test with six NVDIMM threads have become a lot slower. When the workload is increased to 64 GB and 80 GB the NVDIMM time gets even slower compared to the DRAM time.

1.4 Second version

Same as the first version there are two groups of threads that works in parallel in this program. The first group of threads works on the part of the data that is stored on DRAM and the other works on the data stored on NVDIMM. In this version the two threads that has a row of elements with neighbours in the other type of memory will not directly access this data. Instead the two arrays will have their own ghost array on their memory that they will access instead of fetching data from the other side.

The while loop in line 1 will repeat itself for K_length amount of time usually ten times, this is decided by the user of the benchmark. All the threads will wait for all the other threads to arrive at line 2 before one of the threads will start the time measurement in line 5. When the threads arrives at line 8 the all the DRAM threads will pass the if-test and all the NVDIMM threads will move on to line 19. The DRAM threads will first start the time measurement in line 9 and will then start to calculate the average for every elements in the part of the array that have been allocated to the each of the threads, this will happen from line 10-17. The threads will then stop the time measurement in line 18.

The NVDIMM threads will enter the else bracket at line 19. The threads will then start time measurement at line 20. From line 21-28 the threads will calculate average on the elements that have been assigned to each thread. At line 29 the threads will stop the time measurements.

The DRAM and NVDIMM threads will then encounter a barrier at line 31 where they will wait for all the other threads to finish. All threads will then update the ghost arrays at line 34 and 35. At line 34 the threads are copying the second row in the NVDIMM array into the last row in the DRAM array. In line 35 they are copying the second last row in the DRAM array into the first row in the NVDIMM array.

Once all the threads are done copying their parts of the rows, one thread will enter the single at line 37. This thread will swap the DRAM arrays and then the NVDIMM array. The thread will then stop the time measurement that was at the beginning of the while-loop at line 47. At line 48 the k variable gets increased by one. The threads will then return to line 1.

Listing 3: Second version

```
1  while(k<K_length){
2    #pragma omp barrier
3    #pragma omp single
4    {
5      total_time[k] = mysecond();
6    }
7    //Divides threads into DRAM threads and NVDIMM threads.
8    if( thread_id < dram_threads ){
9      individual_time[k][thread_id] = mysecond();
10     for( i=slice_start; i<slice_end; i++){
11       for( j=1; j<nMinusOne; j++){
12         temp = A[i-1][j-1] + A[i-1][j] + A[i-1][j+1]+
13               A[i][j-1]    +      A[i][j+1]+
14               A[i+1][j-1] + A[i+1][j] + A[i+1][j+1];
15         B[i][j] = temp*inverseEigth;
16       }
17     }
18     individual_time[k][thread_id] = mysecond() -
            individual_time[k][thread_id];
19   }else{
20   individual_time[k][thread_id] = mysecond();
21     for( i=slice_start; i<slice_end; i++){
22       for( j=1; j<nMinusOne; j++){
23         temp = D_RO(C)[(i-1)*n+(j-1)] +
              D_RO(C)[(i-1)*n+j] + D_RO(C)[(i-1)*n+(j+1)]+
24             D_RO(C)[i*n+(j-1)]      +
                  D_RO(C)[i*n+(j+1)]+
25             D_RO(C)[(i+1)*n+(j-1)] + D_RO(C)[(i+1)*n+j]
                  + D_RO(C)[(i+1)*n+(j+1)];
26         D_RW(D)[i*n+j] = temp*inverseEigth;
27       }
28     }
29     individual_time[k][thread_id] = mysecond() -
            individual_time[k][thread_id];
30   }
31   #pragma omp barrier
32   #pragma omp for
33   for(a=0; a<n; a++){
34     A[dram_part_Minus_One][a] = D_RO(C)[n+a];
35     D_RW(C)[a] = A[dram_part_Minus_Two][a];
36   }
37   #pragma omp single
38   {
```

```
39      tempArray = B;
40      B=A;
41      A=tempArray;
42
43      temp_nvdimm = C;
44      C = D;
45      D = temp_nvdimm;
46
47      total_time[k] = mysecond() - total_time[k];
48      k++;
49    }
50    #pragma omp barrier
51 }//End of while
```
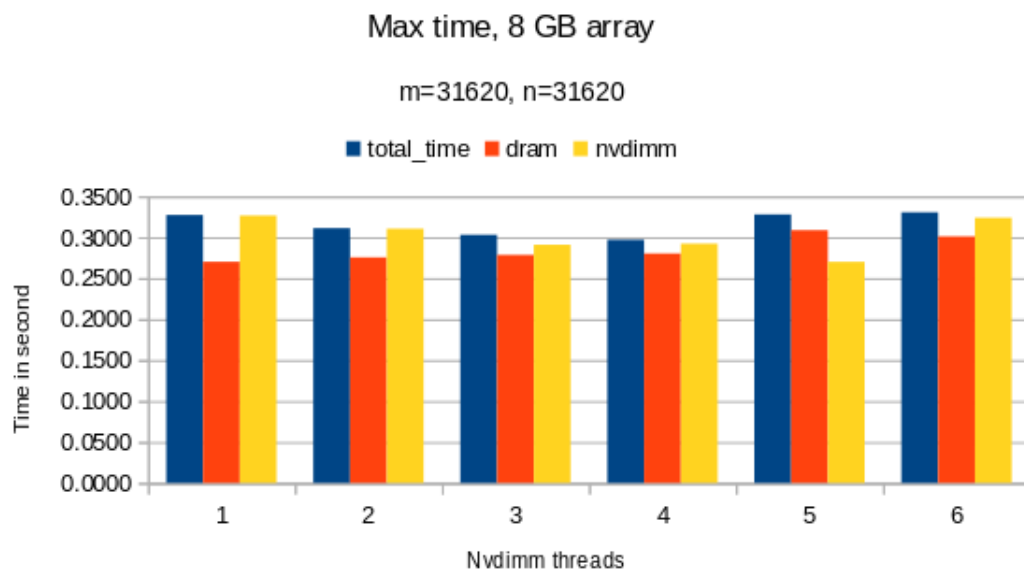

Figure 6: second version, max time

	Prediction	Second Version							
	DRAM	DRAM	DRAM	DRAM	NVDIMM	NVDIMM	NVDIMM	NVDIMM	Size of n
Thread	Speed	Size MB	Pred. Time	Time	Speed	Part MB	Pred. Time	Time	in MB
1	61568.86	7814.69	0.2539	0.2700	1433.87	183.90	0.2565	0.3265	0.2530
2	58704.29	7634.84	0.2601	0.2754	5264.15	363.76	0.1382	0.3104	0.2530
3	55454.72	7499.76	0.2705	0.2785	7421.89	498.84	0.1344	0.2908	0.2530
4	52112.79	7351.02	0.2821	0.2802	10193.13	647.58	0.1271	0.2922	0.2530
5	48564.56	7224.54	0.2975	0.3087	13010.17	774.06	0.1190	0.2700	0.2530
6	45866.16	6965.00	0.3037	0.3011	16101.65	1033.59	0.1284	0.3238	0.2530
1	61568.86	15612.65	0.5072	0.6105	1433.87	386.38	0.5389	0.7408	0.3578
2	58704.29	15271.70	0.5203	0.6316	5264.15	727.33	0.2763	0.7379	0.3578
3	55454.72	15001.23	0.5410	0.6431	7421.89	997.79	0.2689	0.6812	0.3578
4	52112.79	14703.94	0.5643	0.6560	10193.13	1295.09	0.2541	0.6750	0.3578
5	48564.56	14450.64	0.5951	0.6400	13010.17	1548.39	0.2380	0.6592	0.3578
6	45866.16	13931.53	0.6075	0.7165	16101.65	2067.50	0.2568	0.6861	0.3578
1	61568.86	31268.78	1.0157	1.3261	1433.87	735.72	1.0262	1.5071	0.5060
2	58704.29	30549.24	1.0408	1.3598	5264.15	1455.26	0.5529	1.6736	0.5060
3	55454.72	30008.33	1.0823	1.3968	7421.89	1996.17	0.5379	1.5666	0.5060
4	52112.79	29413.27	1.1288	1.4743	10193.13	2591.23	0.5084	1.5287	0.5060
5	48564.56	28907.27	1.1905	1.5775	13010.17	3097.23	0.4761	1.5859	0.5060
6	45866.16	27868.46	1.2152	1.6176	16101.65	4136.04	0.5137	1.8406	0.5060
1	61568.86	62539.15	3.0473	2.9056	1433.87	1471.27	3.0783	3.9167	0.7156
2	58704.29	61100.07	3.1224	3.1172	5264.15	2910.35	1.6586	3.9871	0.7156
3	55454.72	60018.09	3.2469	3.1205	7421.89	3992.33	1.6137	3.9524	0.7156
4	52112.79	58828.04	3.3866	3.2843	10193.13	5182.38	1.5253	4.6903	0.7156
5	48564.56	57816.19	3.5715	3.3156	13010.17	6194.23	1.4283	4.0358	0.7156
6	45866.16	55738.08	3.6457	3.3277	16101.65	8272.34	1.5413	4.9185	0.7156
1	61568.86	78160.80	3.8085	3.9705	1433.87	1839.20	3.8481	5.1265	0.8000
2	58704.29	76362.40	3.9024	3.8766	5264.15	3637.60	2.0730	6.1921	0.8000
3	55454.72	75010.40	4.0579	4.0847	7421.89	4989.60	2.0168	6.3739	0.8000
4	52112.79	73523.20	4.2325	4.0826	10193.13	6476.80	1.9062	6.4060	0.8000
5	48564.56	72258.40	4.4636	4.3185	13010.17	7741.60	1.7851	5.2292	0.8000
6	45866.16	69661.60	4.5564	4.1754	16101.65	10338.40	1.9262	5.4224	0.8000

Table 4: Second version, Prediction

m	n	nvdimm length	dram threads	nvdimm threads	dram average	dram min	dram max	nvdimm average	nvdimm min	nvdimm max	total average	total min	total max
31620	31620	727	15	1	0.2627	0.2574	0.2700	0.3163	0.3162	0.3265	0.3164	0.3052	0.3272
31620	31620	1438	14	2	0.2649	0.2580	0.2754	0.3078	0.3069	0.3104	0.3088	0.3066	0.3111
31620	31620	1972	13	3	0.2706	0.2660	0.2785	0.2852	0.2818	0.2908	0.2905	0.2826	0.3030
31620	31620	2560	12	4	0.2751	0.2714	0.2802	0.2803	0.2751	0.2922	0.2923	0.2871	0.2969
31620	31620	3060	11	5	0.2877	0.2812	0.3087	0.2654	0.2626	0.2700	0.3103	0.2956	0.3279
31620	31620	4086	10	6	0.2936	0.2894	0.3011	0.2988	0.2919	0.3238	0.3268	0.3212	0.3305
44720	44720	1080	15	1	0.5624	0.5283	0.6105	0.7408	0.6508	0.7408	0.7440	0.6510	1.0257
44720	44720	2033	14	2	0.5664	0.5266	0.6316	0.7197	0.6308	0.7379	0.7372	0.6241	1.1207
44720	44720	2789	13	3	0.5744	0.5328	0.6431	0.6388	0.5690	0.6812	0.7201	0.6105	1.1788
44720	44720	3620	12	4	0.5800	0.5381	0.6560	0.6237	0.5503	0.6750	0.7112	0.5944	0.9808
44720	44720	4328	11	5	0.6005	0.5472	0.6400	0.5923	0.5316	0.6592	0.6873	0.6041	0.8672
44720	44720	5779	10	6	0.6150	0.5443	0.7165	0.6318	0.5887	0.6861	0.7364	0.6627	1.0788
63250	63250	1454	15	1	1.2549	1.2403	1.3261	1.4427	1.3735	1.5071	1.4528	1.2606	1.6861
63250	63250	2876	14	2	1.2620	1.2427	1.3598	1.5605	1.3367	1.6736	1.6401	1.2717	2.0893
63250	63250	3945	13	3	1.2647	1.2416	1.3968	1.4216	1.1760	1.5666	1.6460	1.2694	2.6450
63250	63250	5121	12	4	1.2864	1.2436	1.4743	1.3458	1.1560	1.5287	1.6146	1.2952	2.2785
63250	63250	6121	11	5	1.3177	1.2667	1.5775	1.3194	1.1801	1.5859	1.6600	1.3247	2.1018
63250	63250	8174	10	6	1.3231	1.2665	1.6176	1.4227	1.3059	1.8406	1.7482	1.3105	2.2546
89450	89450	2056	15	1	2.8438	2.8170	2.9056	2.8137	2.8137	3.9167	3.0696	2.8250	3.9325
89450	89450	4067	14	2	2.9161	2.8384	3.1172	3.2239	2.9246	3.9871	3.6211	2.8220	4.1466
89450	89450	5579	13	3	2.9185	2.8582	3.1205	3.0320	2.6291	3.9524	3.5822	2.7917	4.5728
89450	89450	7242	12	4	3.0148	2.9170	3.2843	3.1972	2.7765	4.6903	3.9077	2.8639	4.6944
89450	89450	8656	11	5	3.0399	2.9426	3.3156	3.1701	2.4809	4.0358	3.9958	3.3680	4.7907
89450	89450	11560	10	6	3.0656	2.9548	3.3277	3.7235	2.9749	4.9185	4.5136	3.3499	5.1769
100000	100000	2299	15	1	3.7444	3.6900	3.9705	4.2077	4.1984	5.1265	4.3952	3.6912	5.1932
100000	100000	4547	14	2	3.7030	3.6505	3.8766	4.5366	4.4105	6.1921	4.9242	3.7079	6.9846
100000	100000	6237	13	3	3.8682	3.7749	4.0847	5.7169	5.1832	6.3739	6.3394	5.6729	6.9403
100000	100000	8096	12	4	3.8714	3.8035	4.0826	5.2261	3.9844	6.4060	6.4100	5.7252	6.9710
100000	100000	9677	11	5	3.9831	3.8362	4.3185	4.2058	3.2200	5.2292	5.3170	4.3486	6.4441
100000	100000	12923	10	6	3.9253	3.6374	4.1754	4.4446	3.5566	5.4224	5.4257	4.8637	6.7254

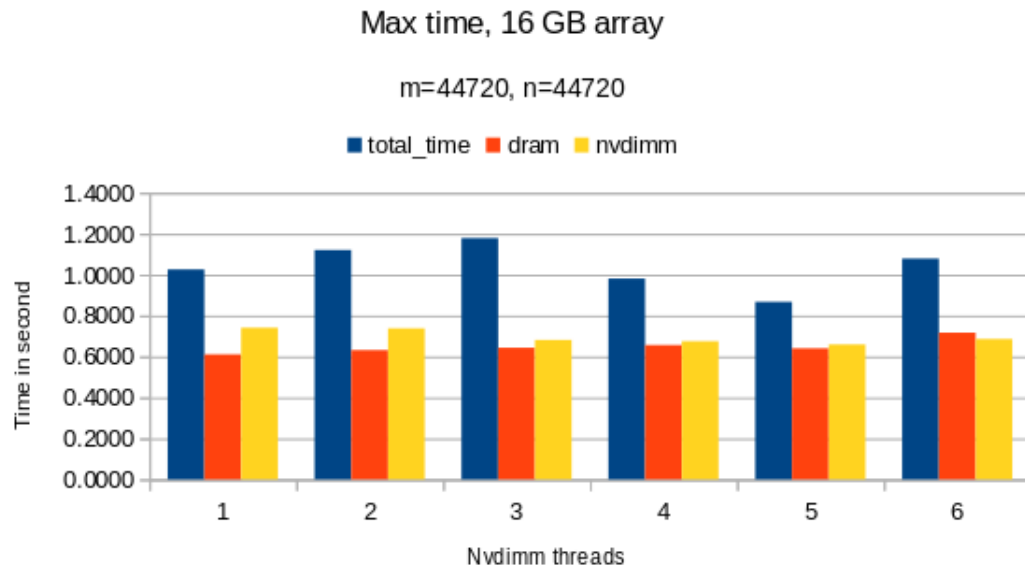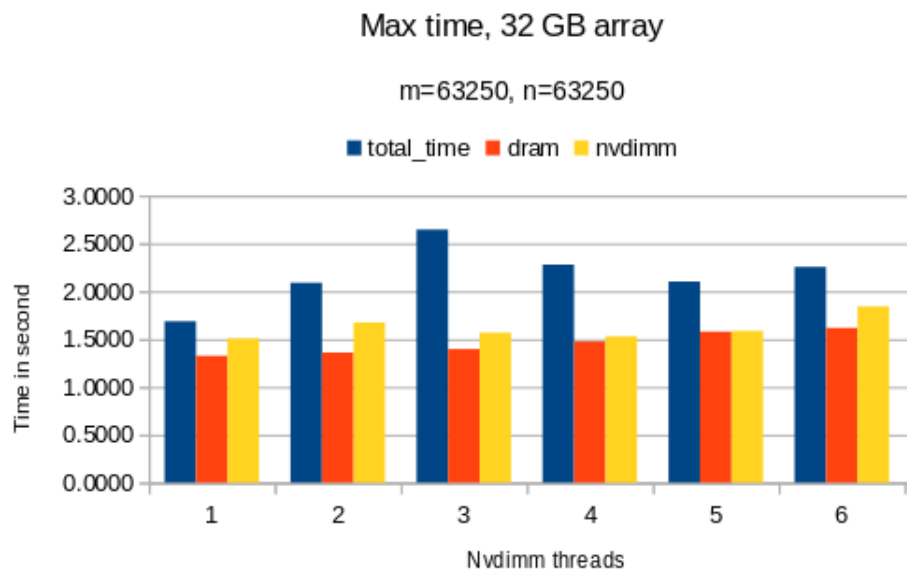Table 5: second version, result

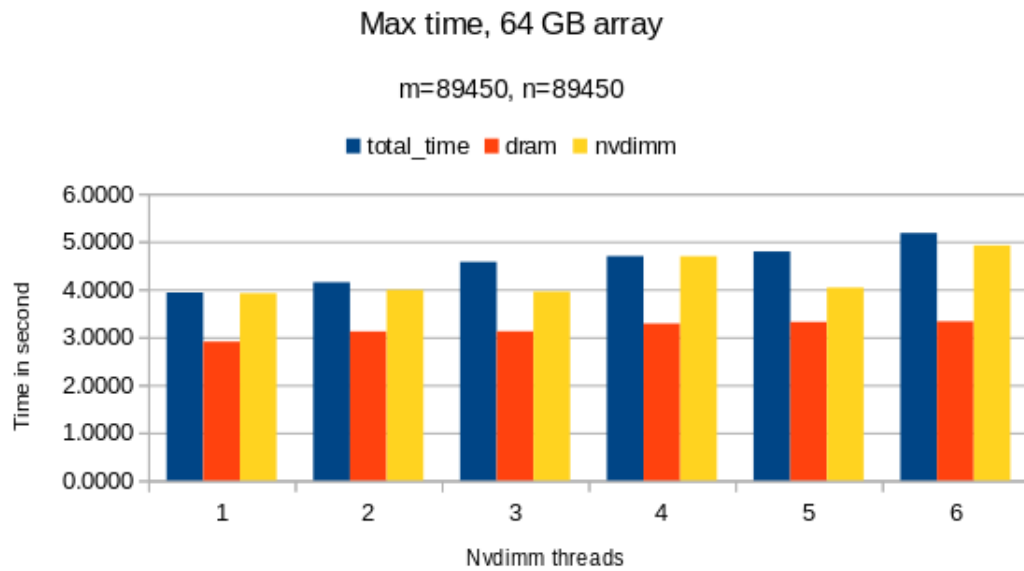Figure 7: second version, max time

Figure 8: second version, max time

Max time, 64 GB array

m=89450, n=89450

■ total_time ■ dram ■ nvdimm

Figure 9: second version, max time

Max time, 80 GB array

m=100000, n=100000
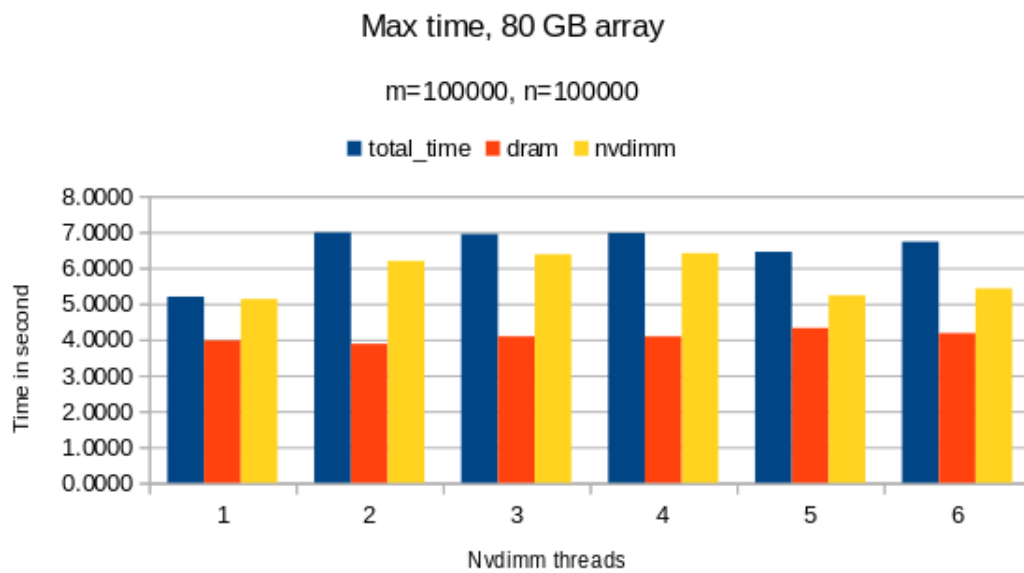
■ total_time ■ dram ■ nvdimm

Figure 10: second version, max time

The formula have been tested when the total array size are 8 GB, 16 GB, 32 GB, 64 GB and 80 GB. When the array size is 8 GB, 16 GB and 32 GB the workload between DRAM and NVDIMM is almost perfectly balanced. When the array size is increased to 64 GB and 80 GB the NVDIMM becomes slower and slower.

1.4.1 Comparisons

This