

CSCE 435 Group project

0. Group number:

6

1. Group members:

1. Yida Zou
2. Brian Chen
3. Sam Hollenbeck
4. Alex Pantazopol

2. Project topic

We chose to use the suggested topic idea:

Choose 3+ parallel sorting algorithms, implement in MPI and CUDA. Examine and compare performance in detail (computation time, communication time, how much data is sent) on a variety of inputs: sorted, random, reverse, sorted with 1% perturbed, etc. Strong scaling, weak scaling, GPU performance.

2. Project description

We will communicate via iMessage.

Algorithms we will implement:

- Sample Sort (MPI)
- Sample Sort (CUDA)
- Merge Sort (MPI)
- Merge Sort (CUDA)
- Bitonic Sort (MPI)
- Bitonic Sort (CUDA)

We will compare the performance of these three algorithms with the metrics stated in the project topic.

To vary our algorithms, we will apply the following communication and parallelization strategies:

- fork/join parallelism
- point-to-point communication

3. Pseudocode

```
function SampleSort(unsortedList, t): //t = thread count
    //Divide input into samples depending on number of threads
    sampleSize = calculateSampleSize(unsortedList, t)
    sample = selectSample(unsortedList, sampleSize)

    //Distribute the sample to all processors
    distribute(sample) //Using MPI

    //Each thread sorts the given sample locally
    sortedSample = sortLocally(sample) //sort using cudaMemcpy

    //Gather sorted samples
    sortedSamples = communicate(sample) //Using MPI

    //Merge and sort all samples together
    sortedLists = mergeSublists(sortedSamples)

    return sortedSublist

function void merge_sort(arr, numThreads):

    # Initialize MPI environment
    MPI_Init()

    # Get the MPI rank and size
    rank = MPI_Comm_rank(MPI_COMM_WORLD)
    size = MPI_Comm_size(MPI_COMM_WORLD)

    # Divide and distribute data across nodes
    local_data = distribute_data(rank, size)

    # Sort local data using CUDA
    local_data = cuda_merge_sort(local_data)
```

```

# Communicate data between nodes, and merge lists
sorted_data = gather_and_merge(local_data, rank, size)

# Finalize MPI environment
MPI_Finalize()

return sorted_data

function void bitonic_sort(arr, numThreads) {

    MPI_Init(...);
    rank = MPI_Comm_rank(...);
    size = MPI_Comm_size(...);

    MPI_Scatter(...);

    for (...) { // major step
        for (...) { // minor step
            sortOnGPU(...);
        }
    }

    MPI_Gather(...);
    MPI_Finalize();
    return sorted_data;
}

```

3. Evaluation Plan

We evaluated our sorting algorithm through the following:

Runtimes between parallel sorting algorithms on GPU and CPU (MPI vs CUDA)

Scaling the number of threads or processors

Scaling the problem size (length of array to sort)

Comparing different input types (Random, Sorted, Reverse Sorted, 1 % perturbed)

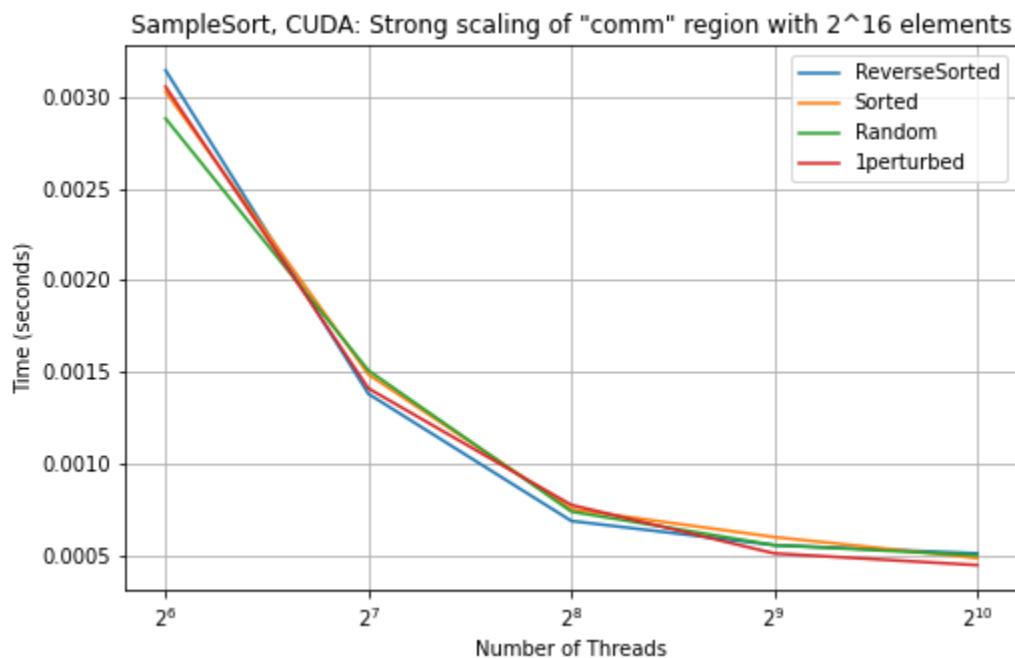
3. Project implementation

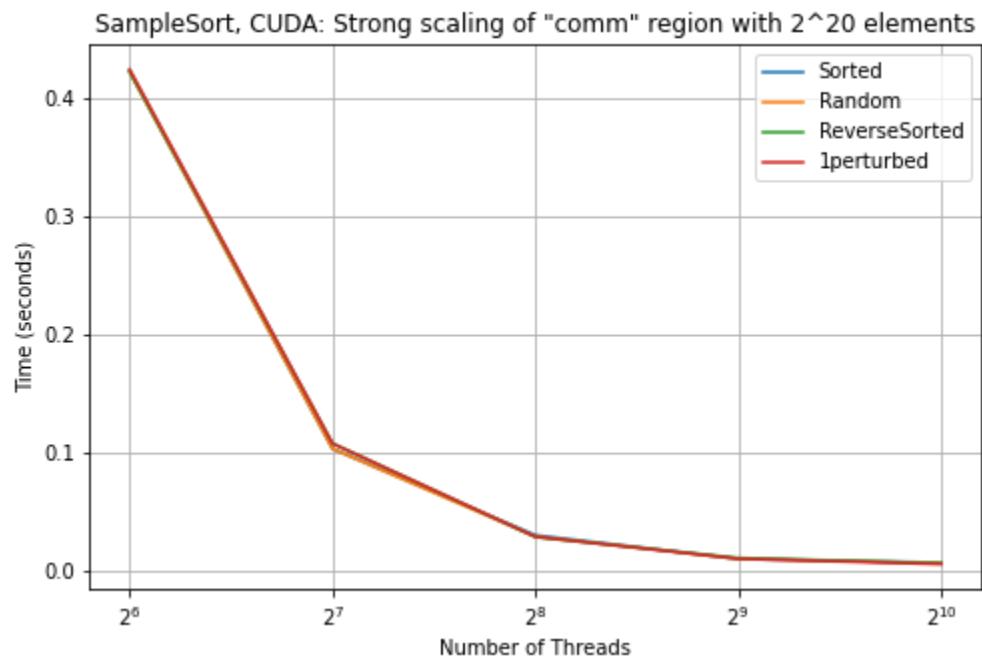
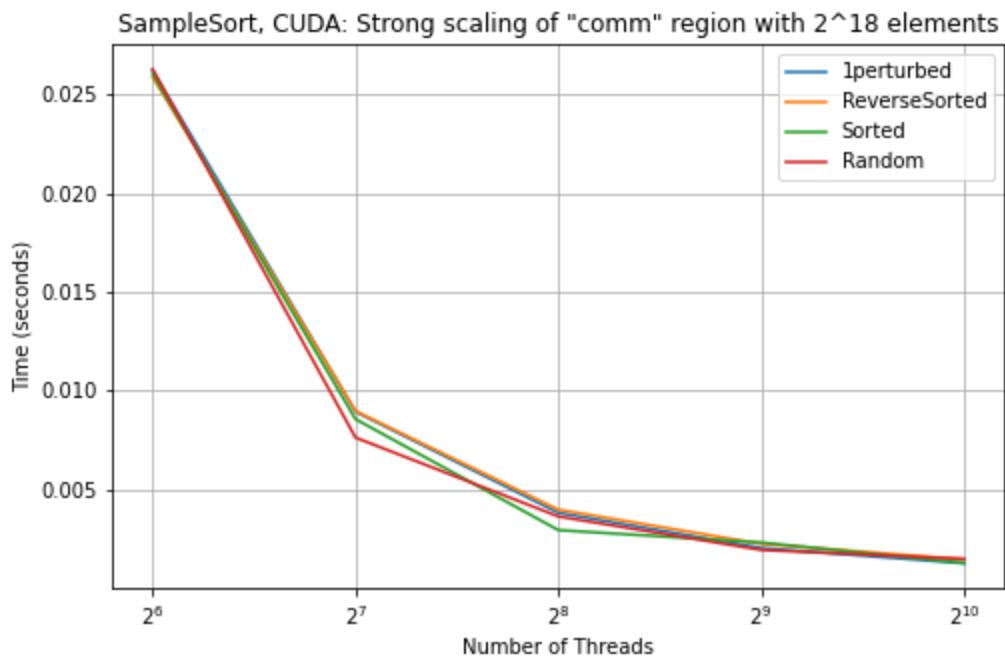
Our implementations can be found in their respective folders (BitonicSort, MergeSort, SampleSort) with each folder having subfolders for MPI and CUDA.

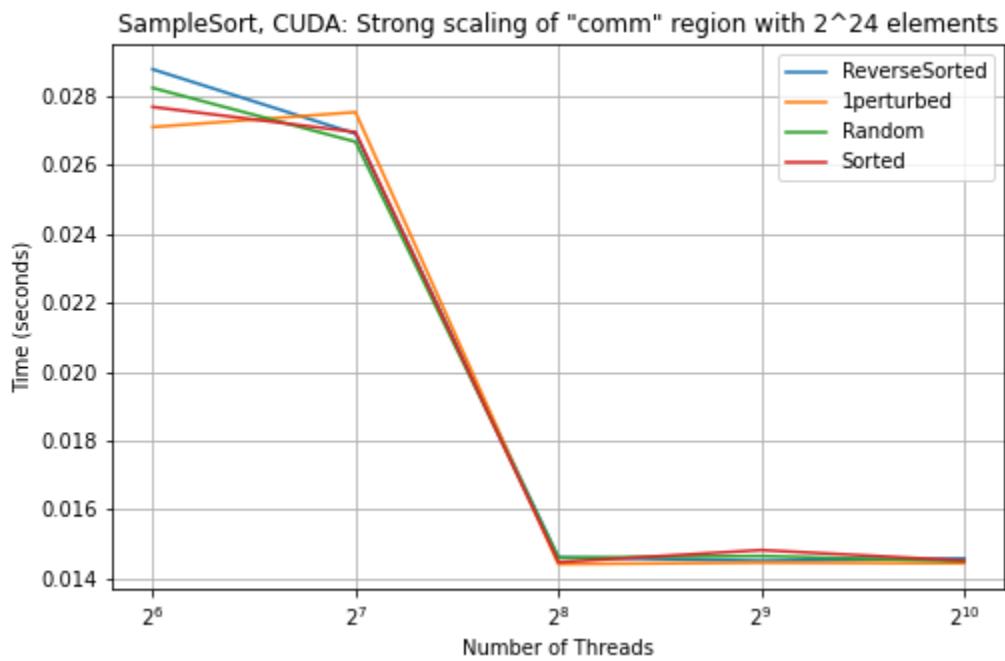
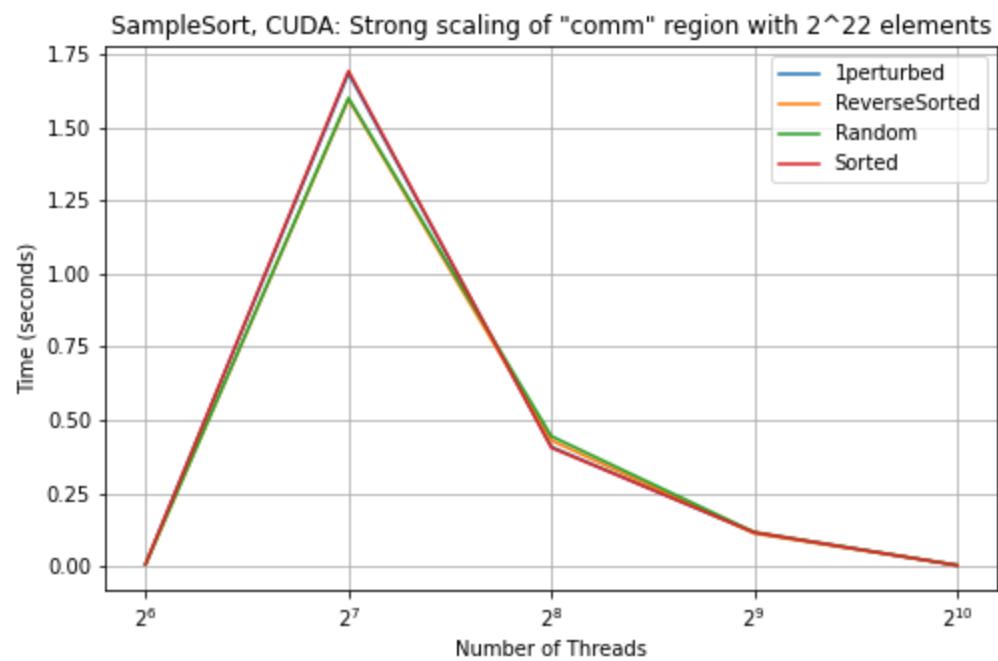
4. Performance evaluation

Sample Sort CUDA

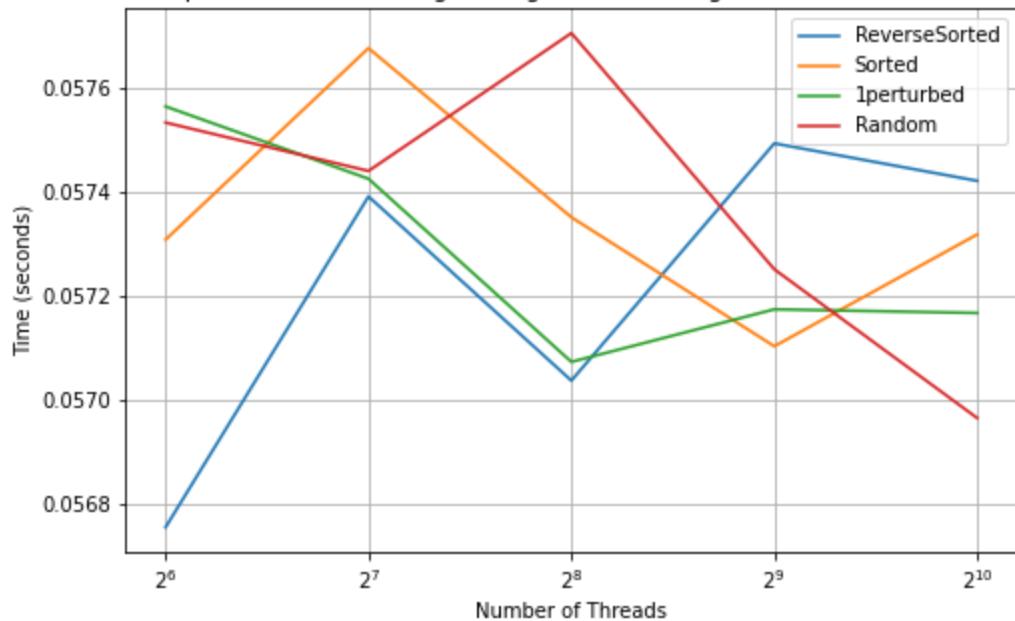
Graphs



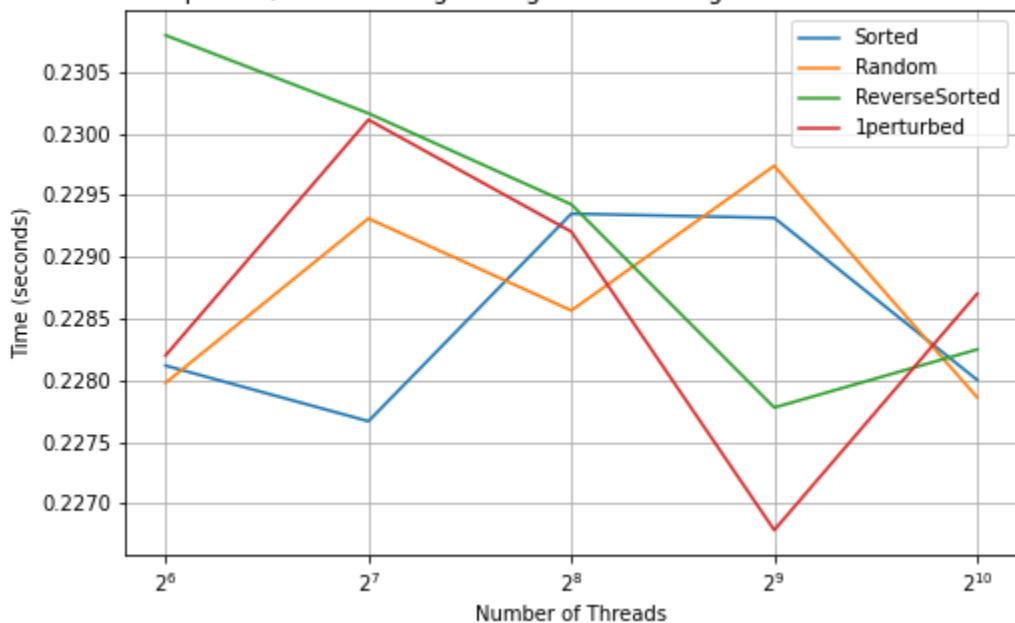


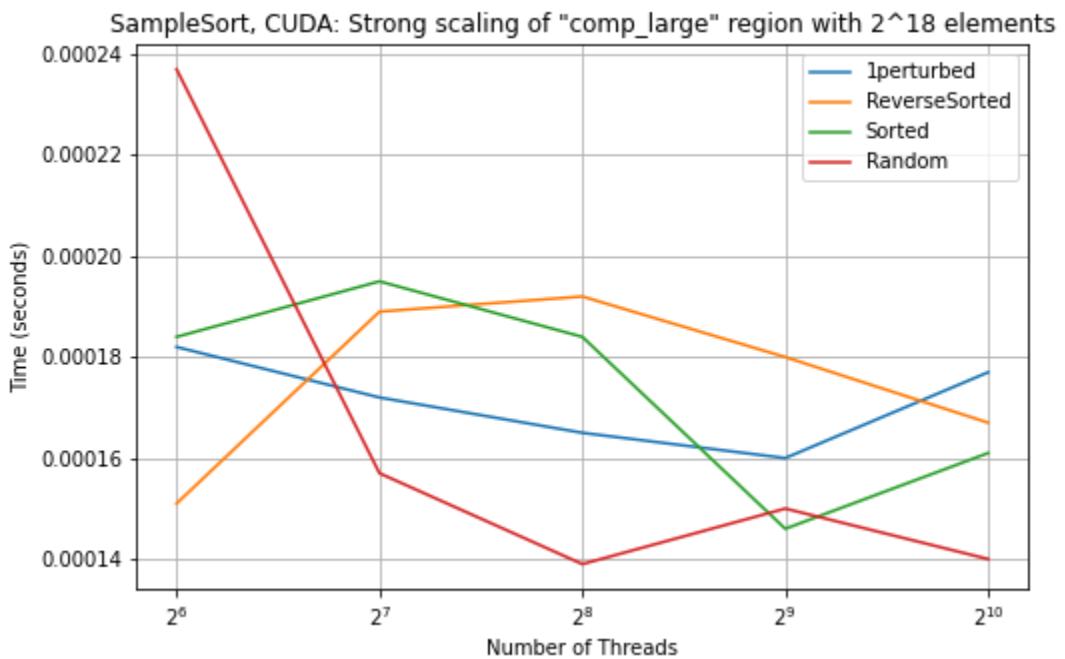
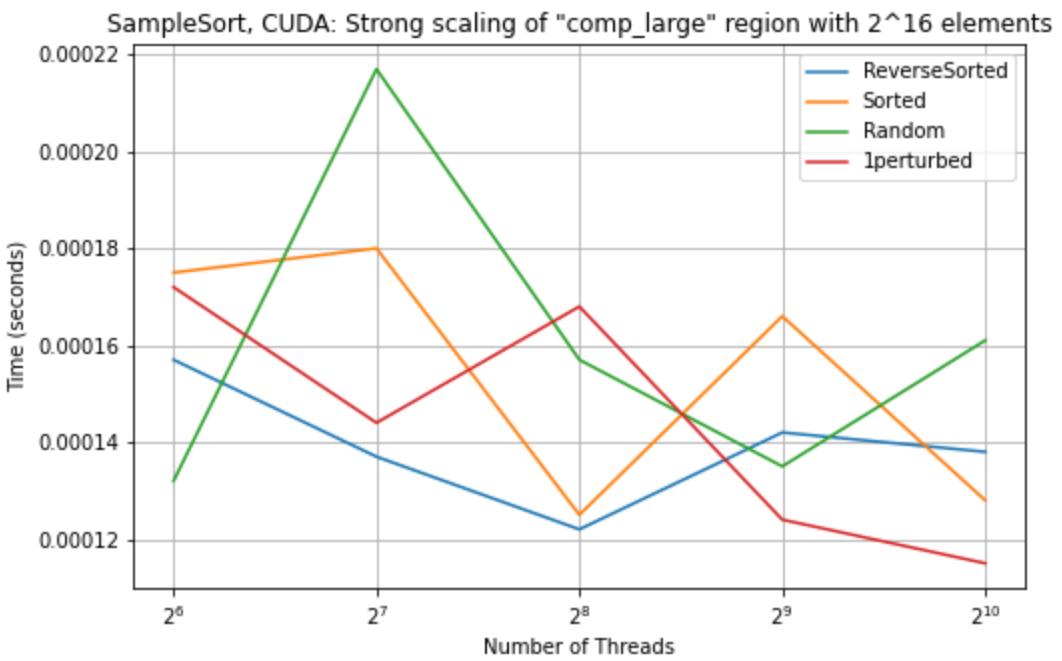


SampleSort, CUDA: Strong scaling of "comm" region with 2^{26} elements

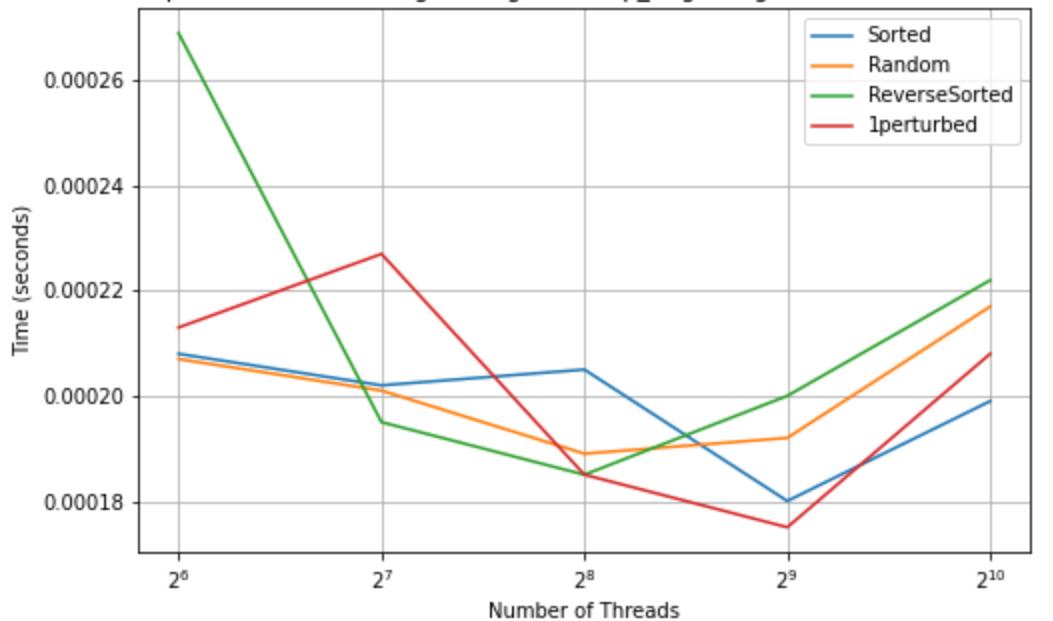


SampleSort, CUDA: Strong scaling of "comm" region with 2^{28} elements

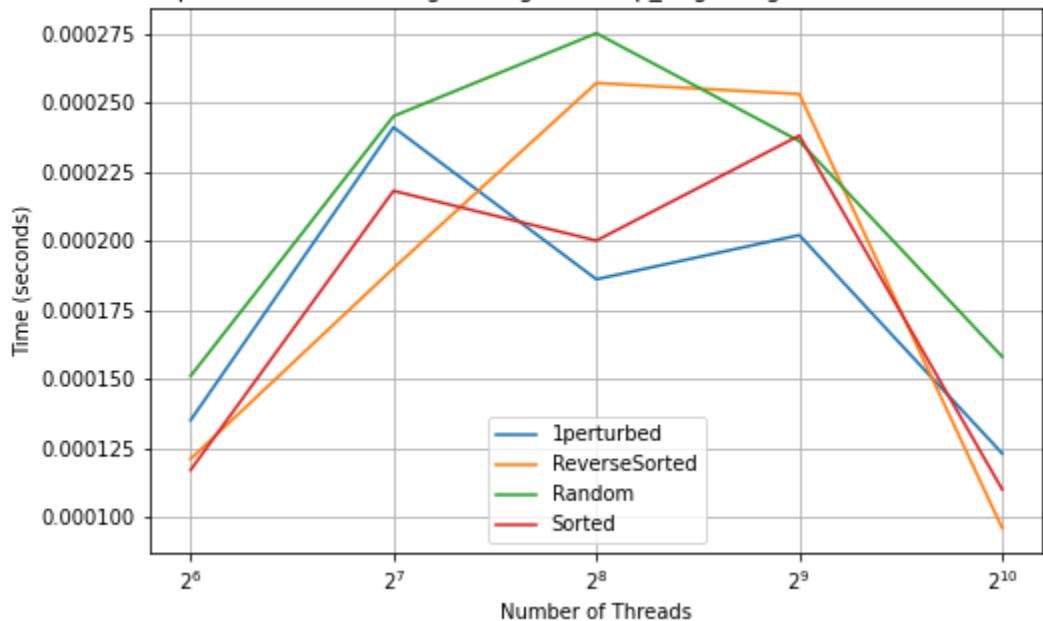




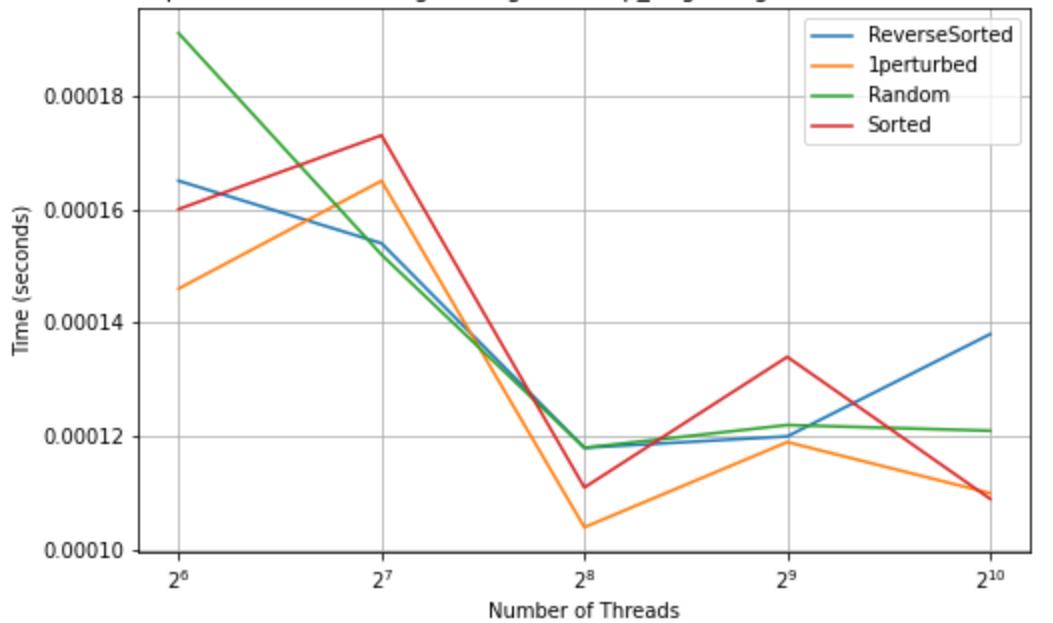
SampleSort, CUDA: Strong scaling of "comp_large" region with 2^{20} elements



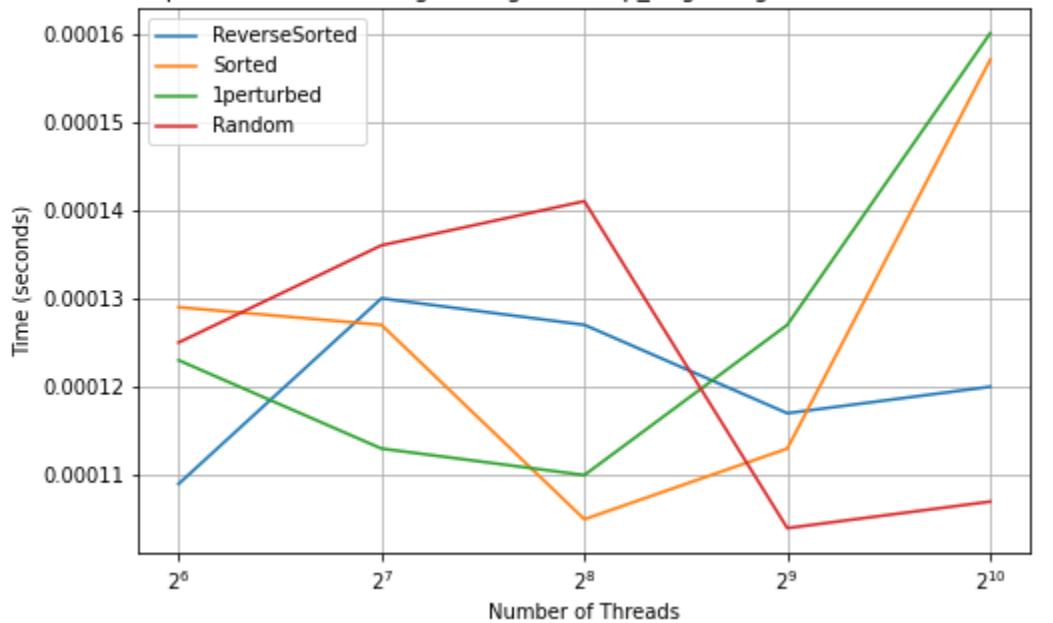
SampleSort, CUDA: Strong scaling of "comp_large" region with 2^{22} elements



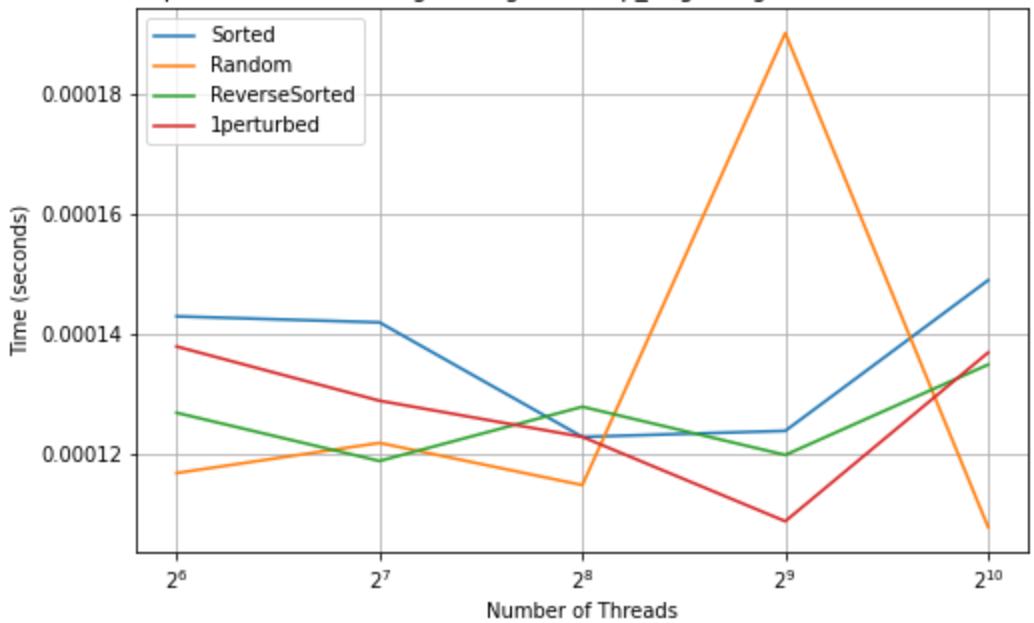
SampleSort, CUDA: Strong scaling of "comp_large" region with 2^{24} elements



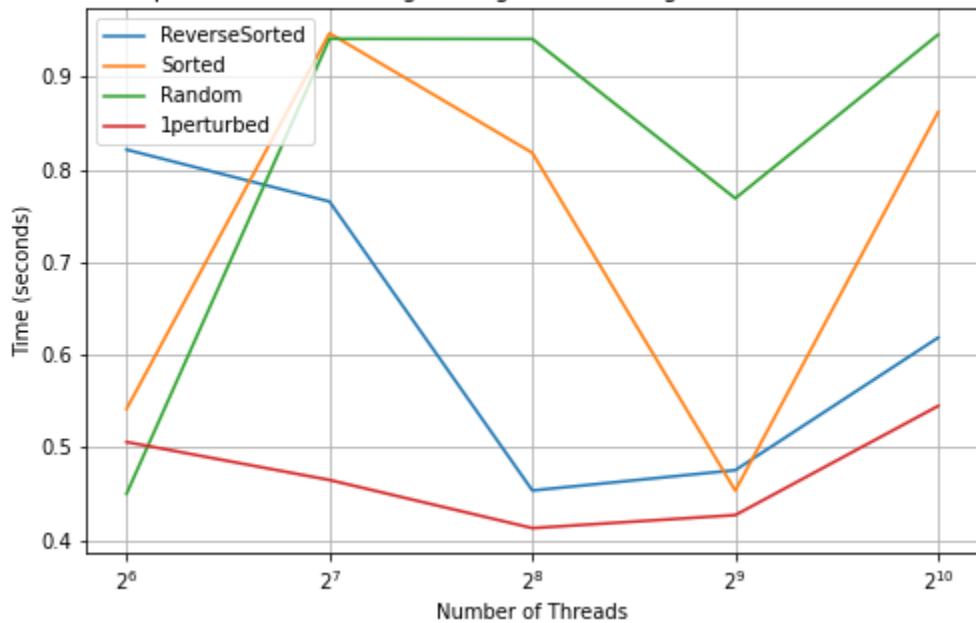
SampleSort, CUDA: Strong scaling of "comp_large" region with 2^{26} elements



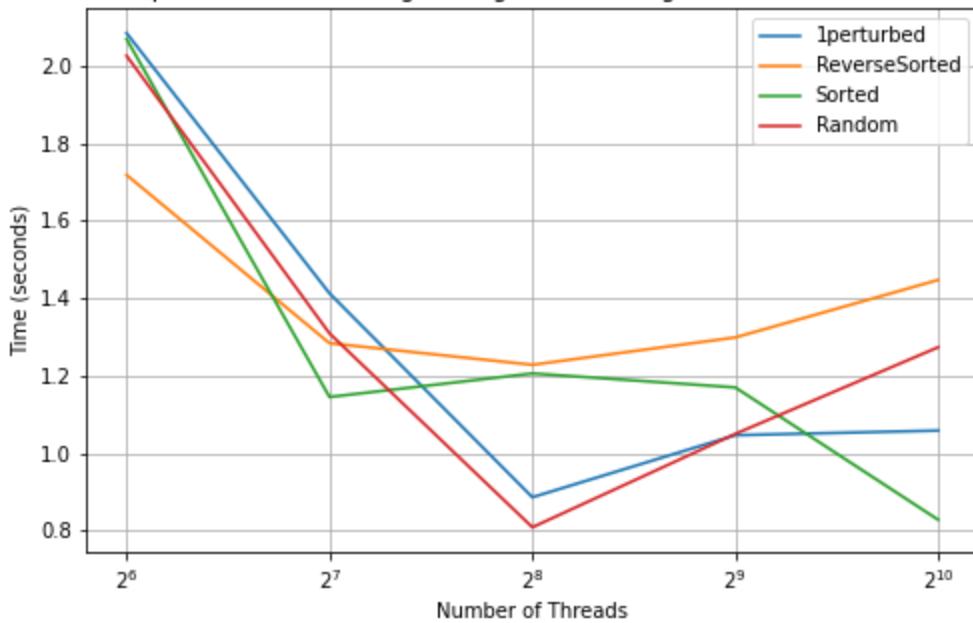
SampleSort, CUDA: Strong scaling of "comp_large" region with 2^{28} elements



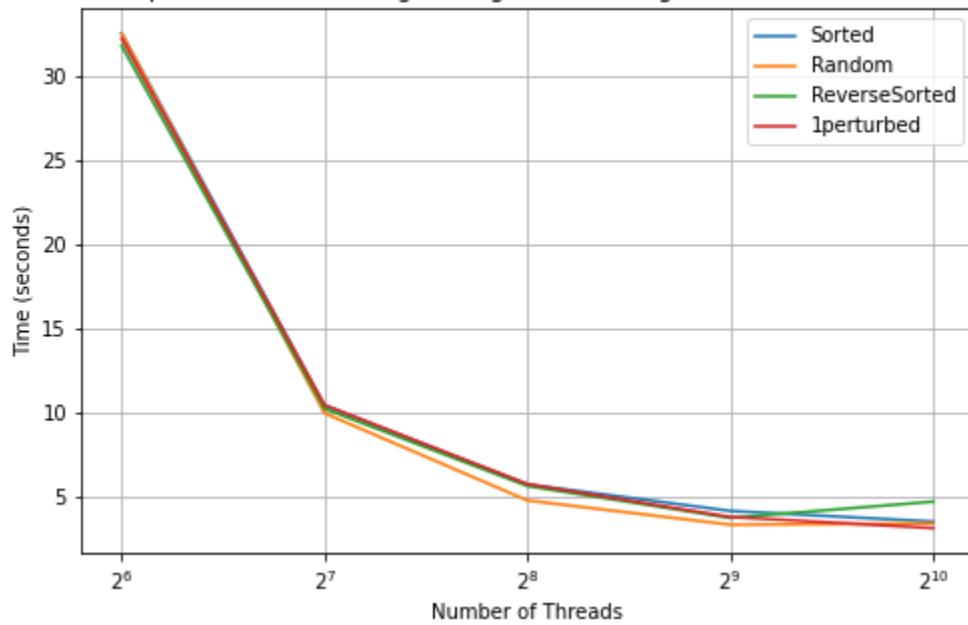
SampleSort, CUDA: Strong scaling of "main" region with 2^{16} elements



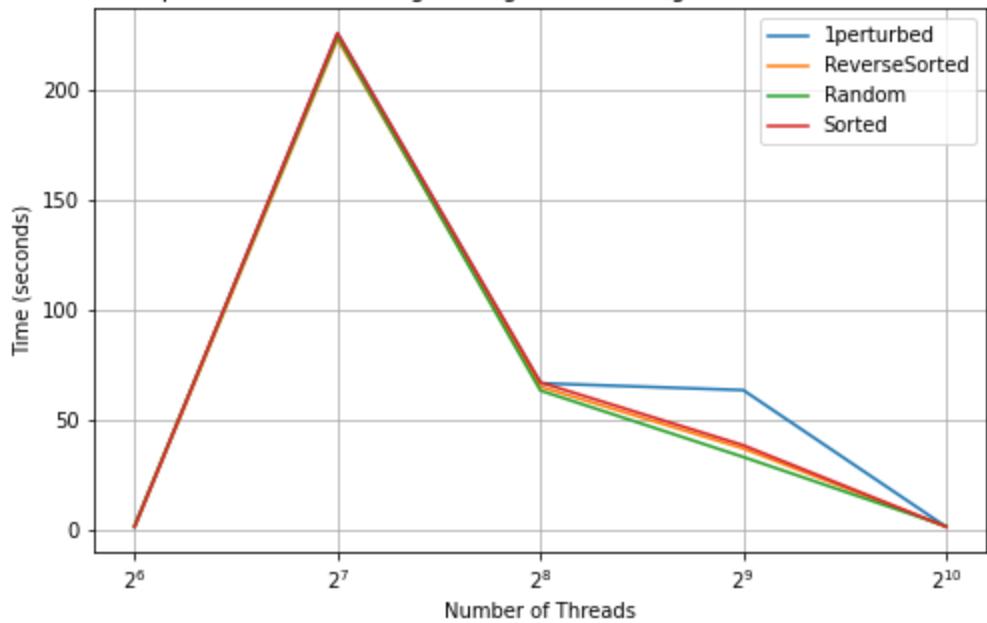
SampleSort, CUDA: Strong scaling of "main" region with 2^{18} elements



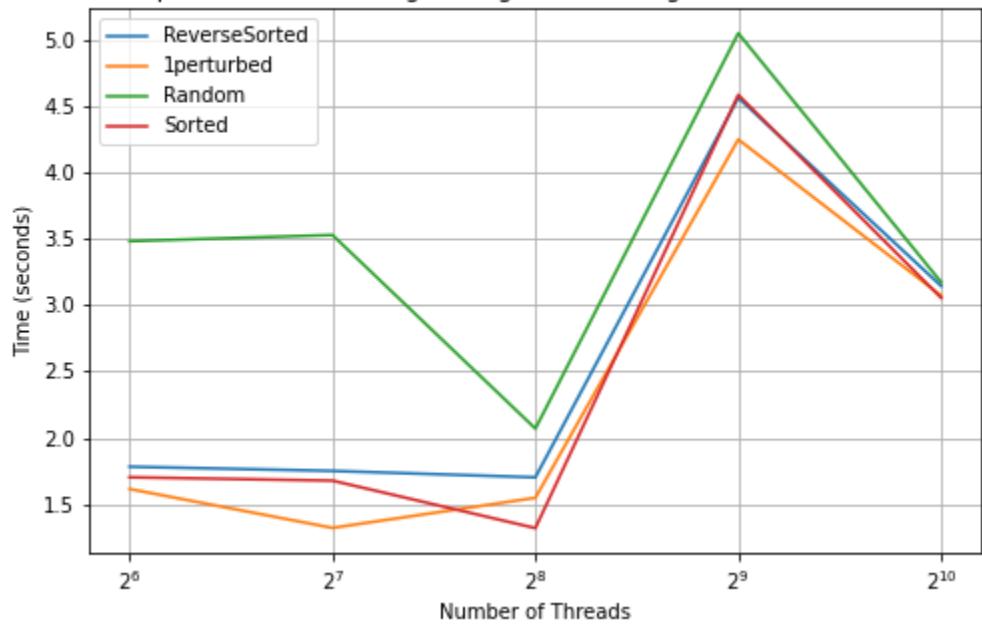
SampleSort, CUDA: Strong scaling of "main" region with 2^{20} elements

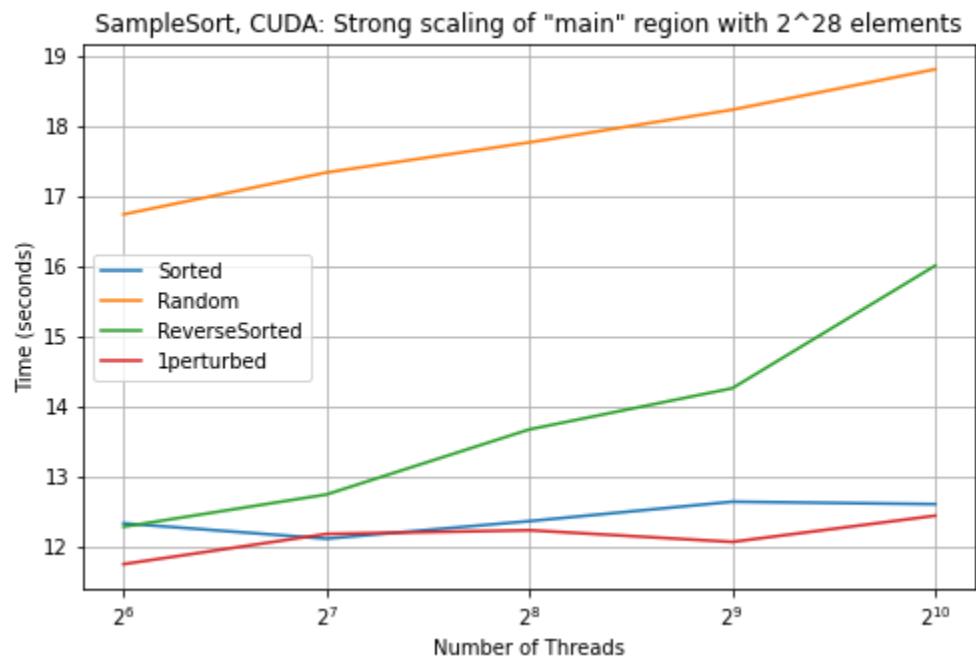
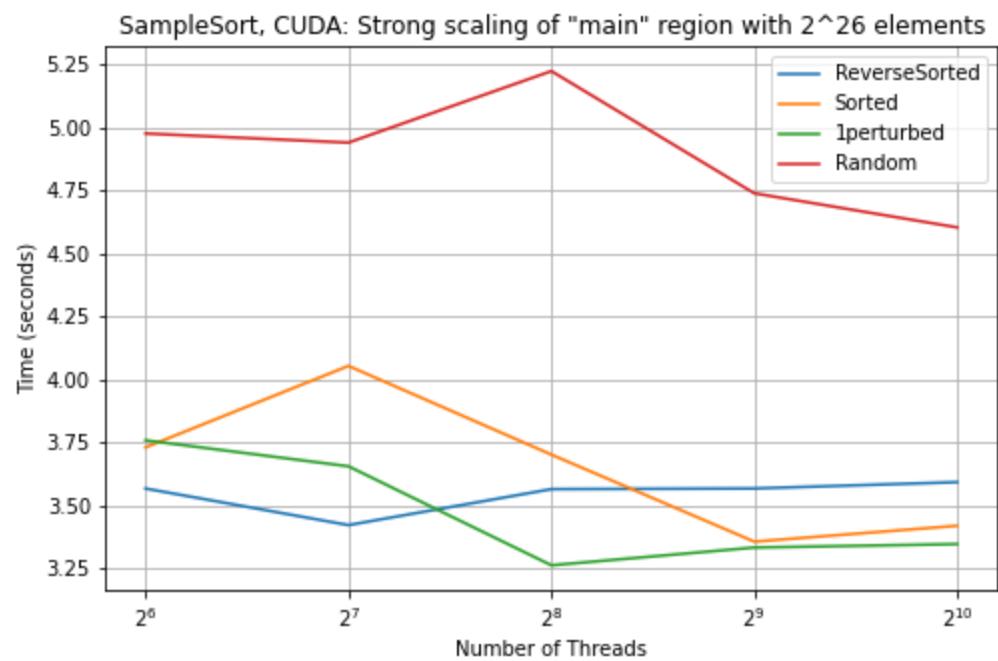


SampleSort, CUDA: Strong scaling of "main" region with 2^{22} elements

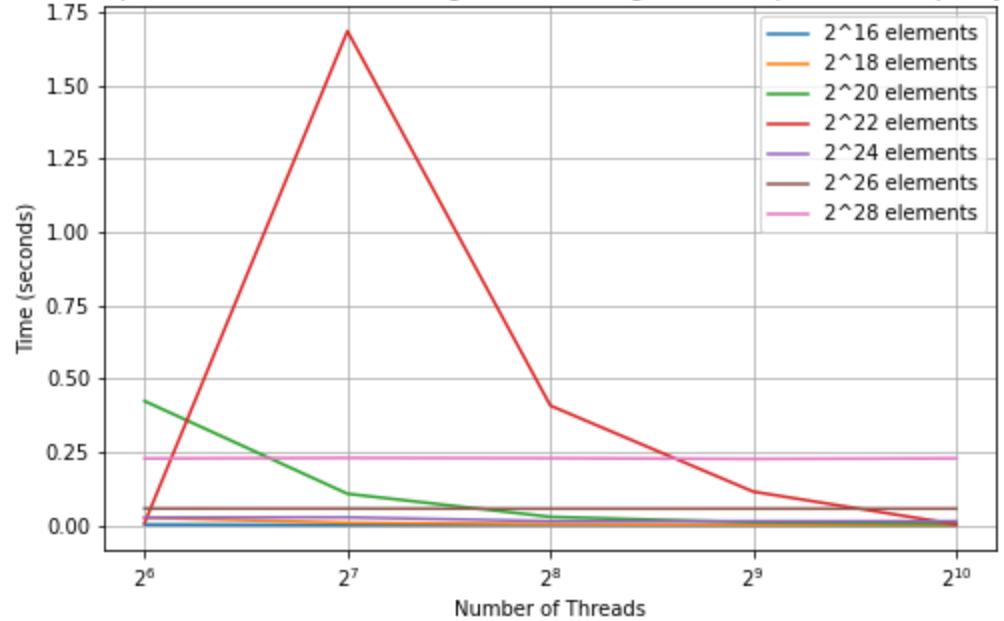


SampleSort, CUDA: Strong scaling of "main" region with 2^{24} elements

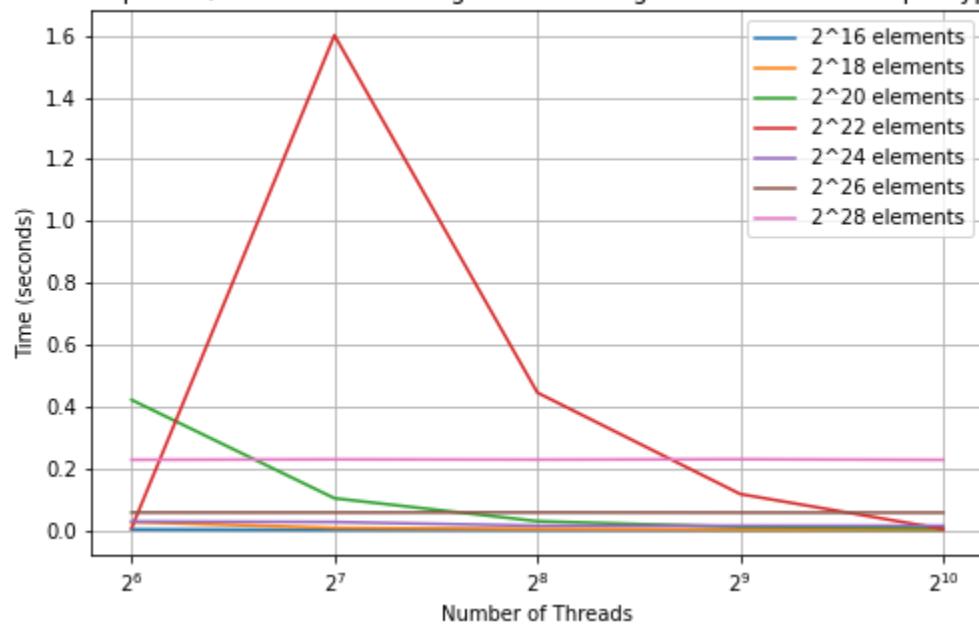




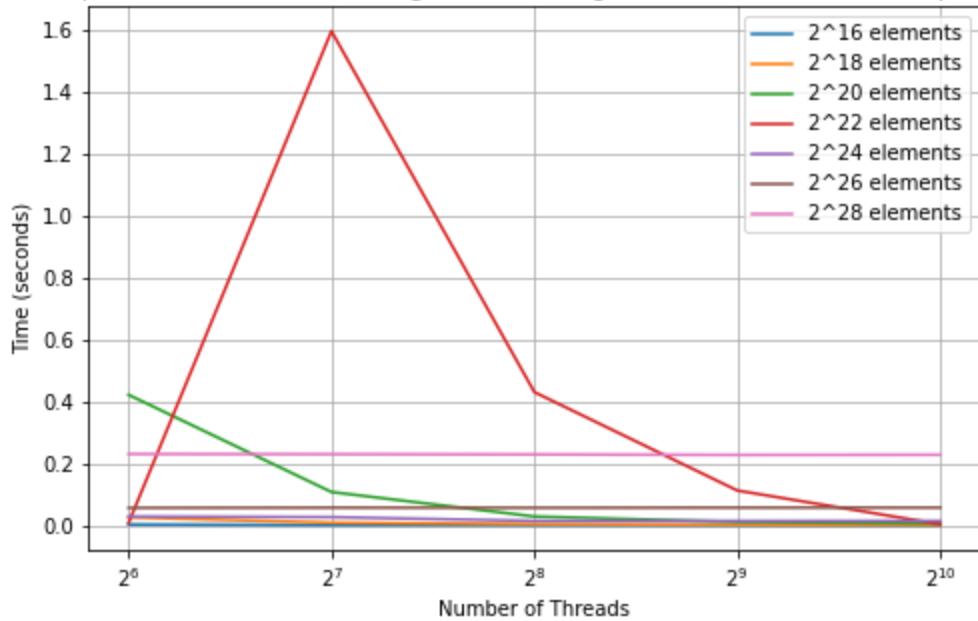
SampleSort, CUDA: Weak scaling of "comm" region with "1perturbed" input type



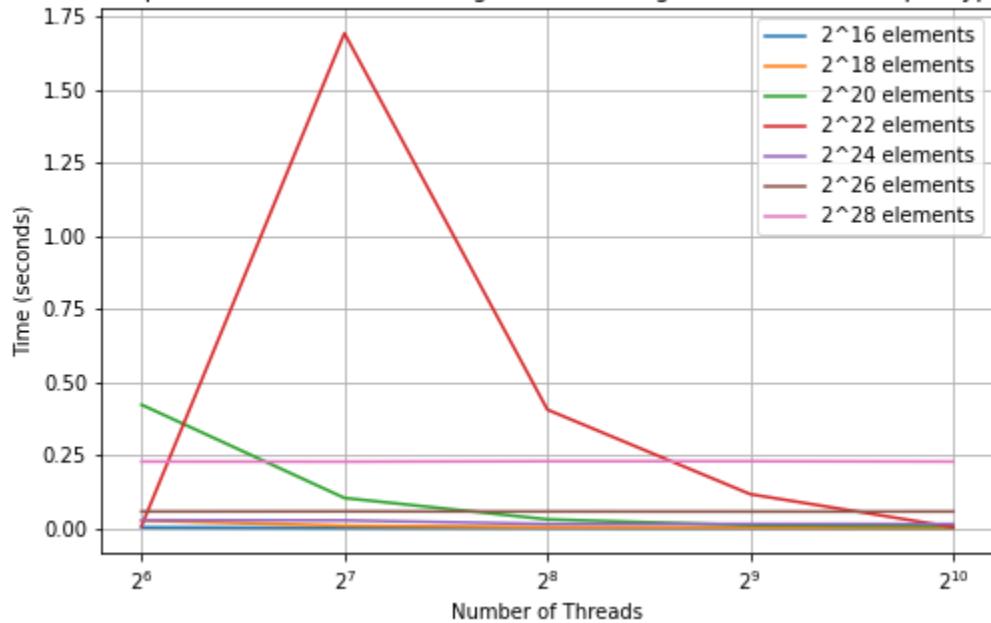
SampleSort, CUDA: Weak scaling of "comm" region with "Random" input type



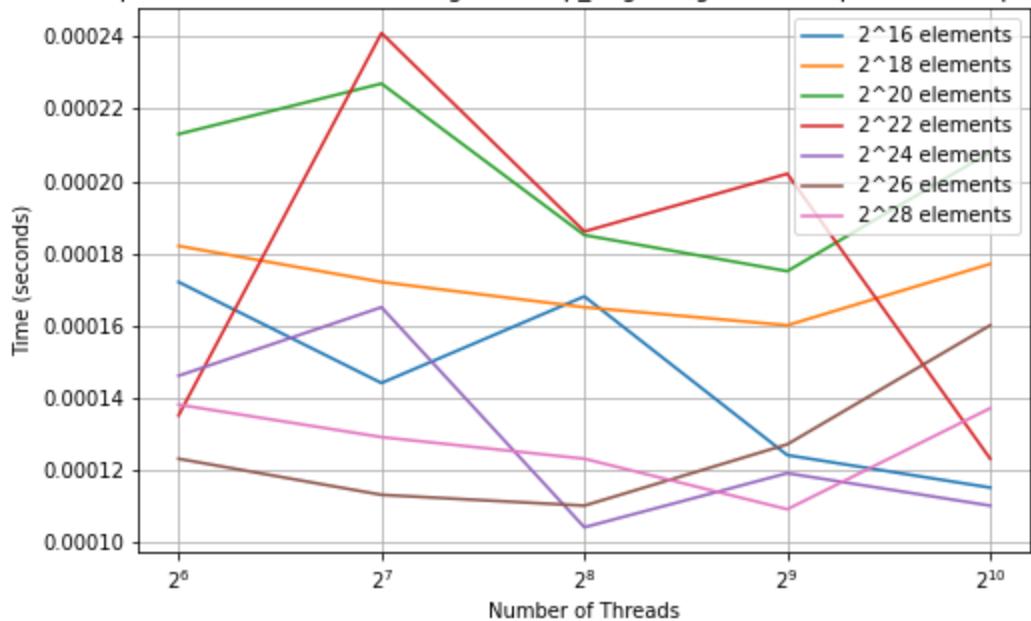
SampleSort, CUDA: Weak scaling of "comm" region with "ReverseSorted" input type



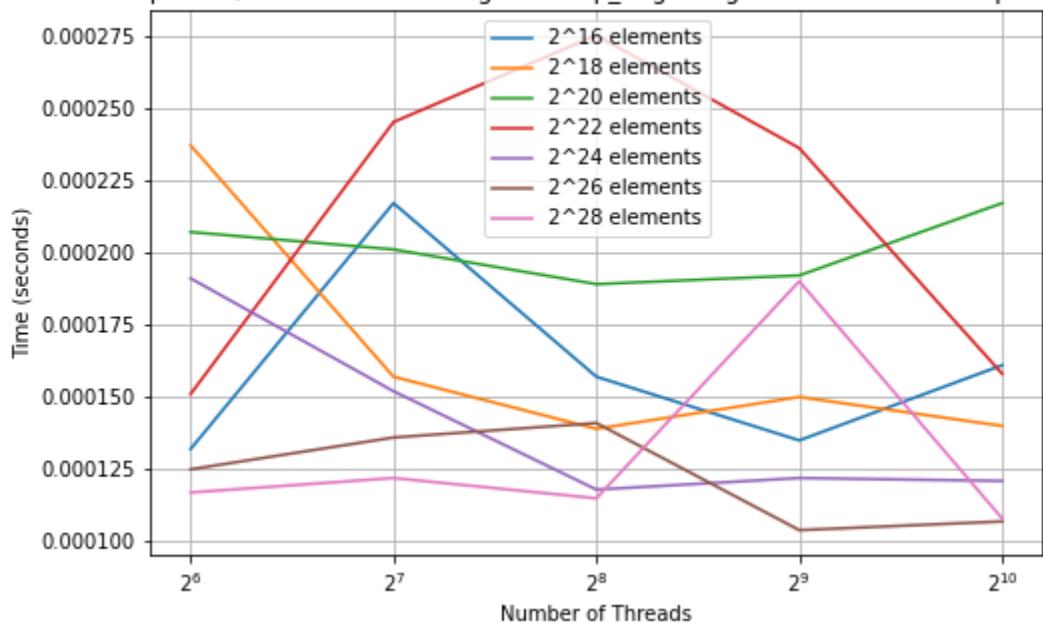
SampleSort, CUDA: Weak scaling of "comm" region with "Sorted" input type

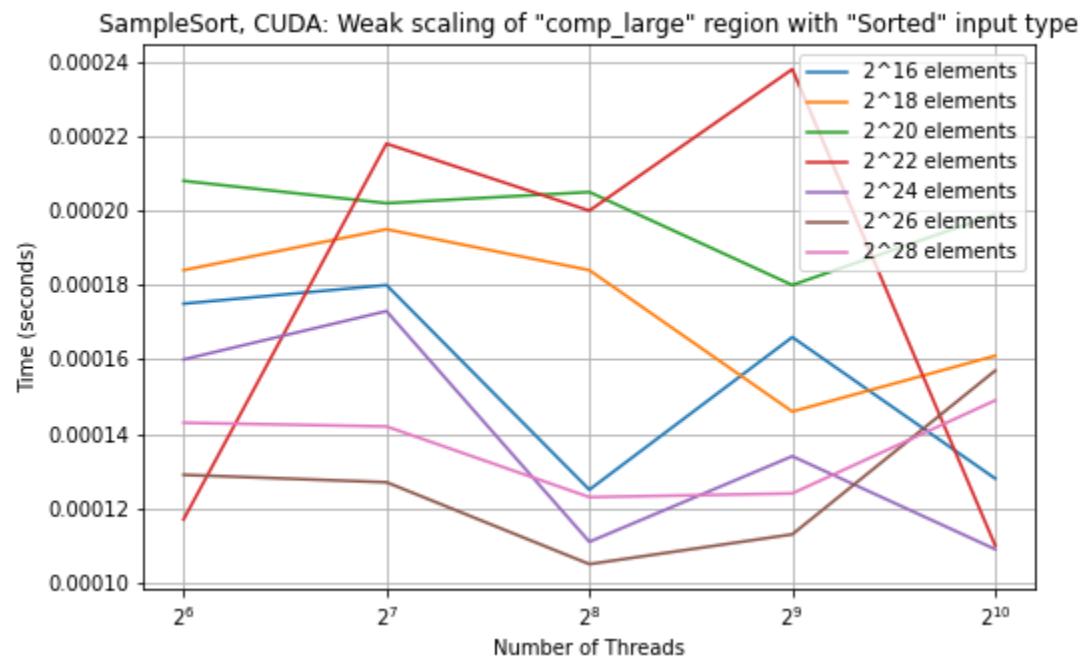
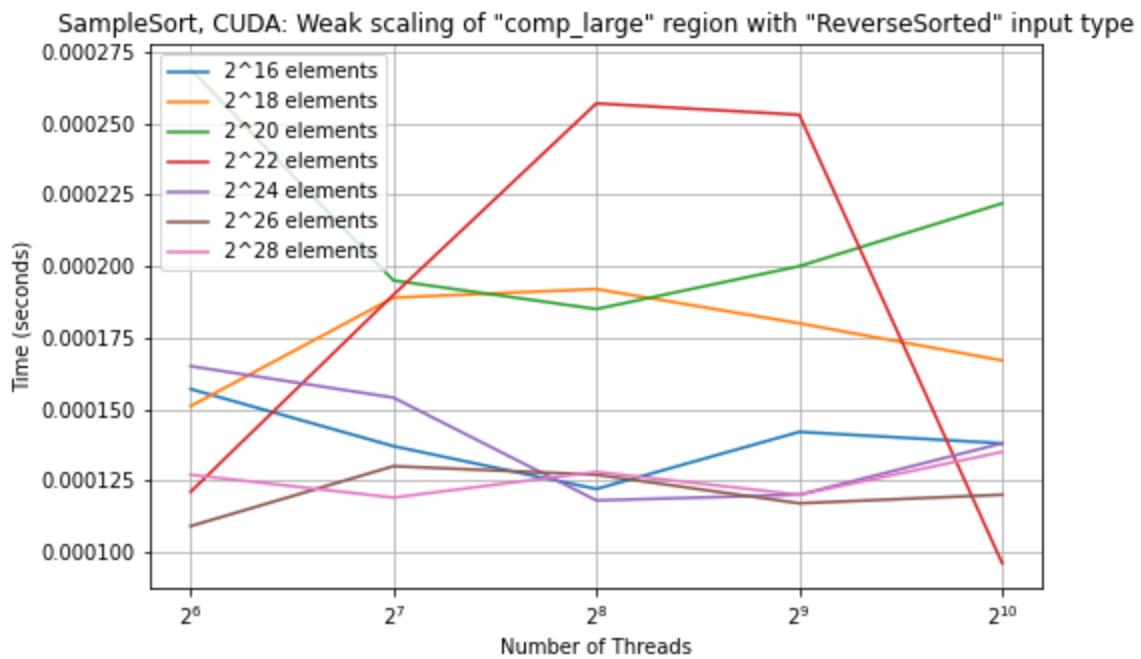


SampleSort, CUDA: Weak scaling of "comp_large" region with "1perturbed" input type

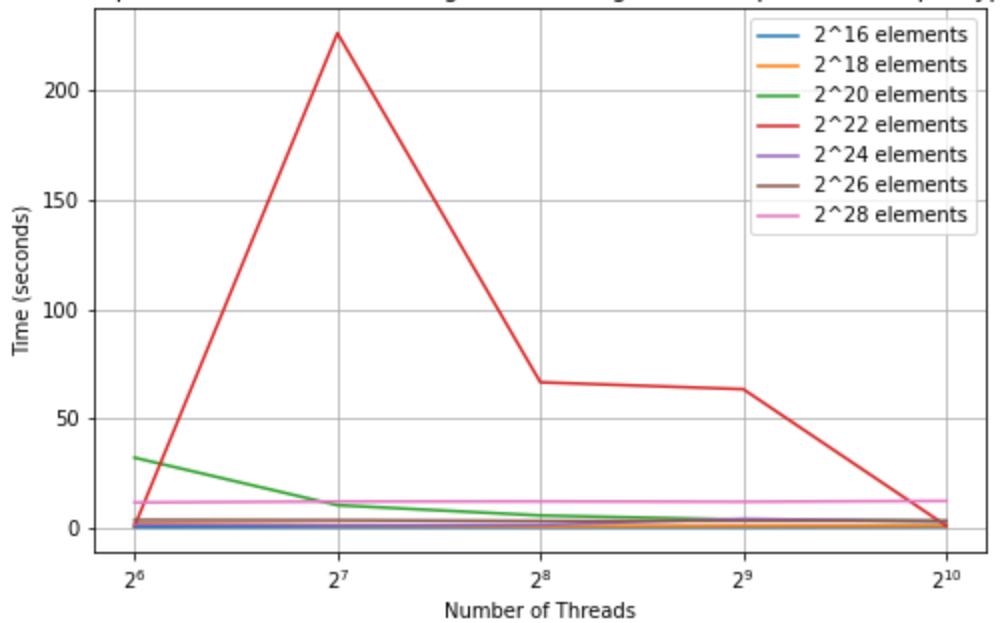


SampleSort, CUDA: Weak scaling of "comp_large" region with "Random" input type

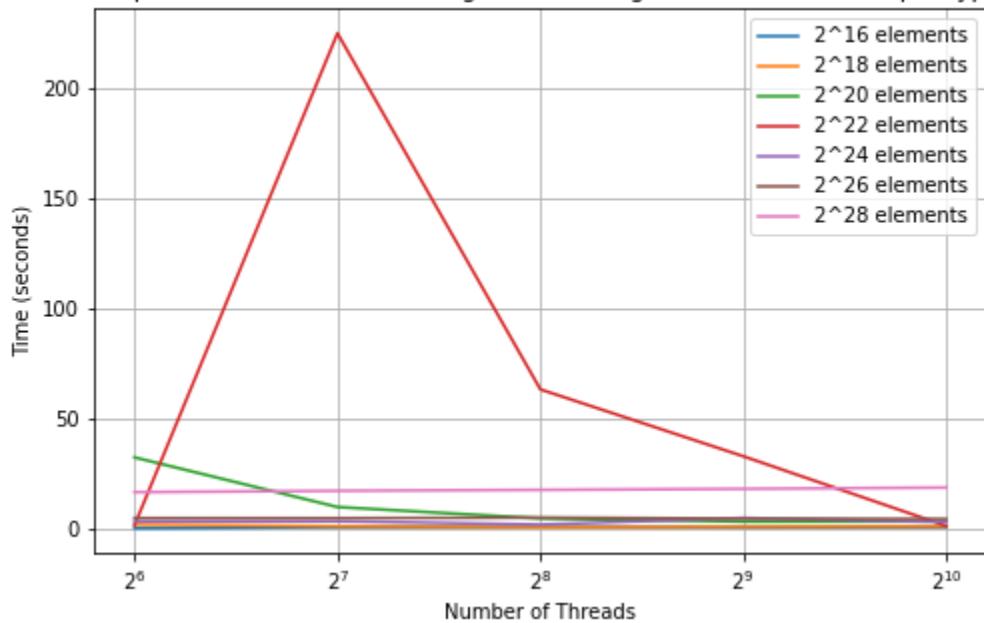




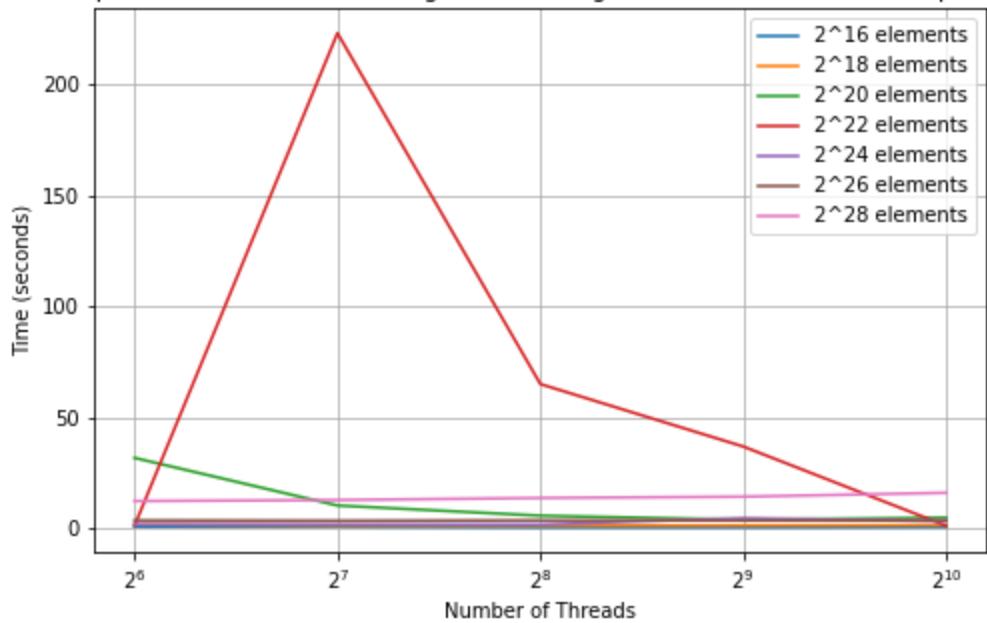
SampleSort, CUDA: Weak scaling of "main" region with "1perturbed" input type



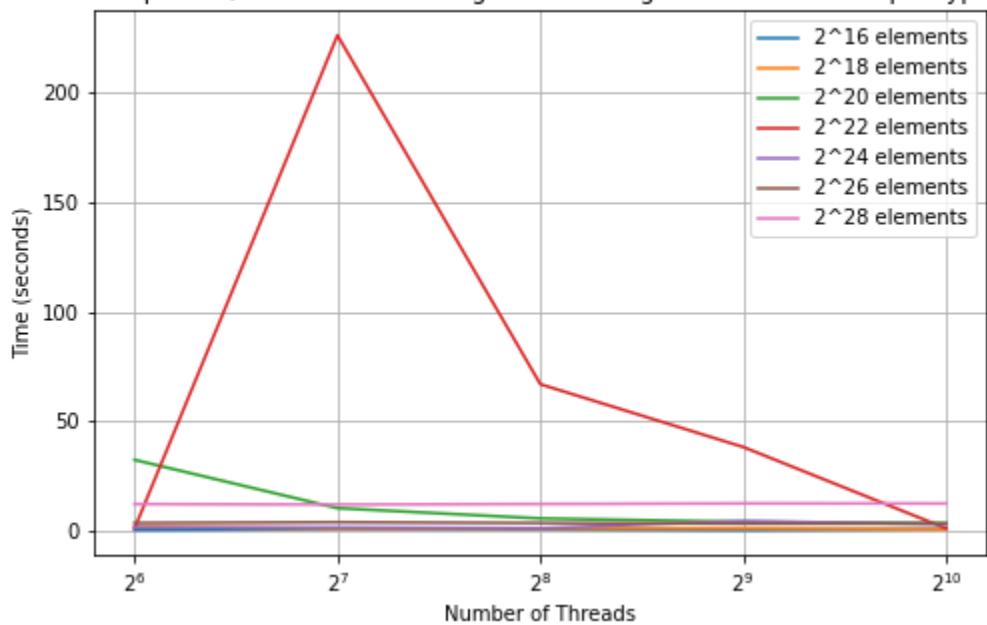
SampleSort, CUDA: Weak scaling of "main" region with "Random" input type



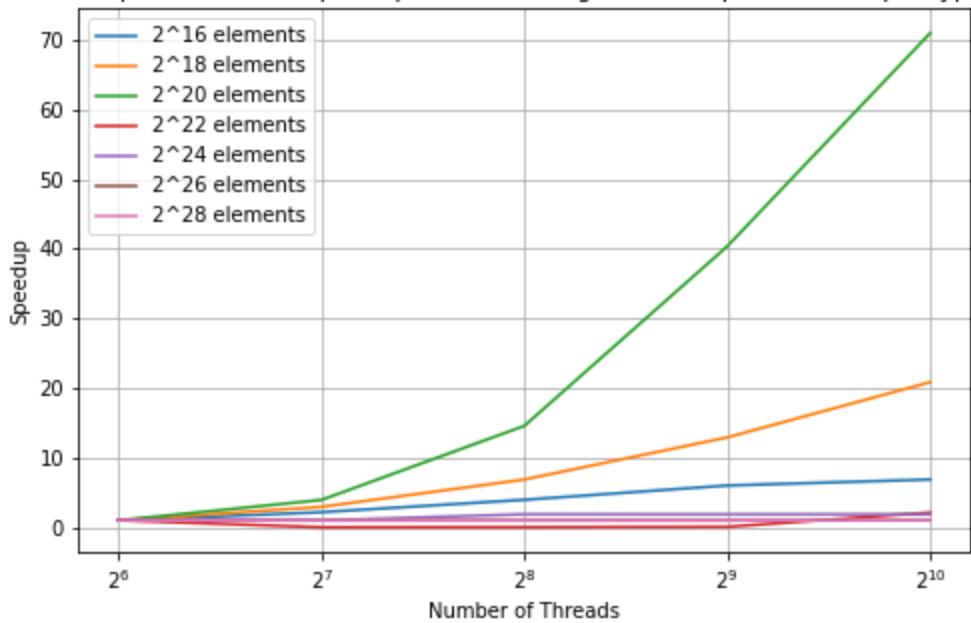
SampleSort, CUDA: Weak scaling of "main" region with "ReverseSorted" input type



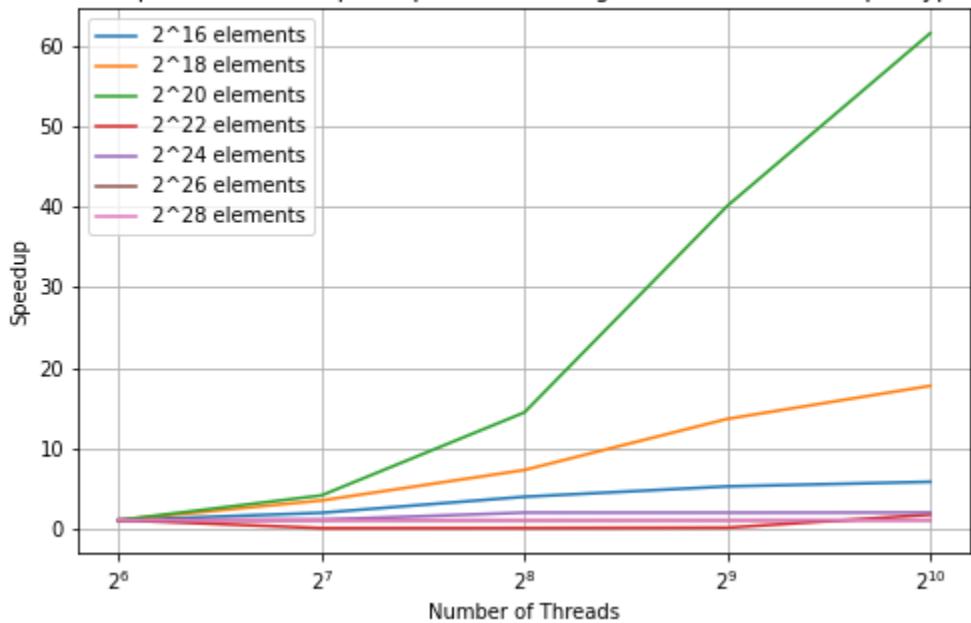
SampleSort, CUDA: Weak scaling of "main" region with "Sorted" input type



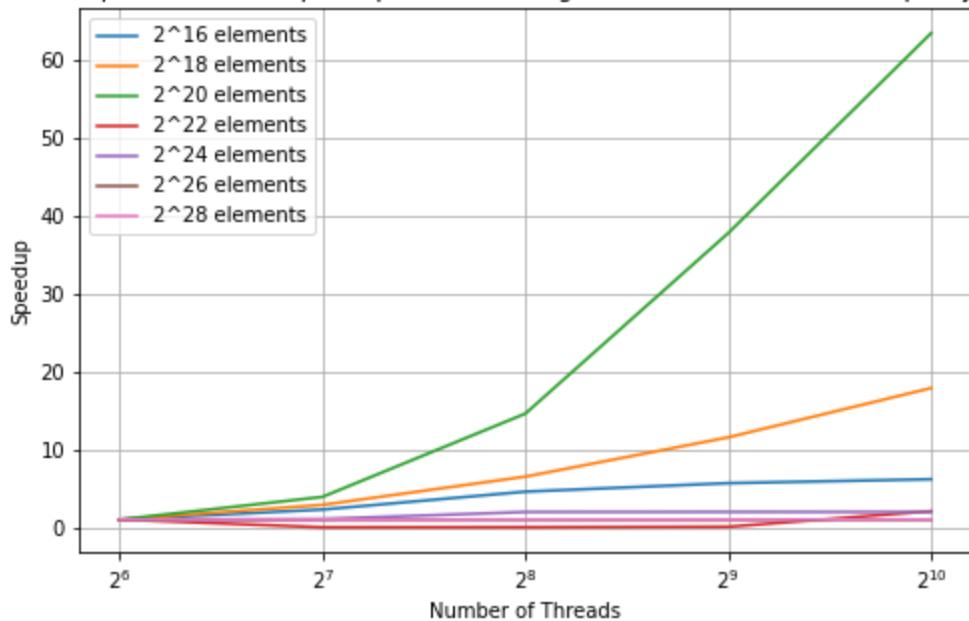
SampleSort, CUDA: Speedup of "comm" region with "1perturbed" input type



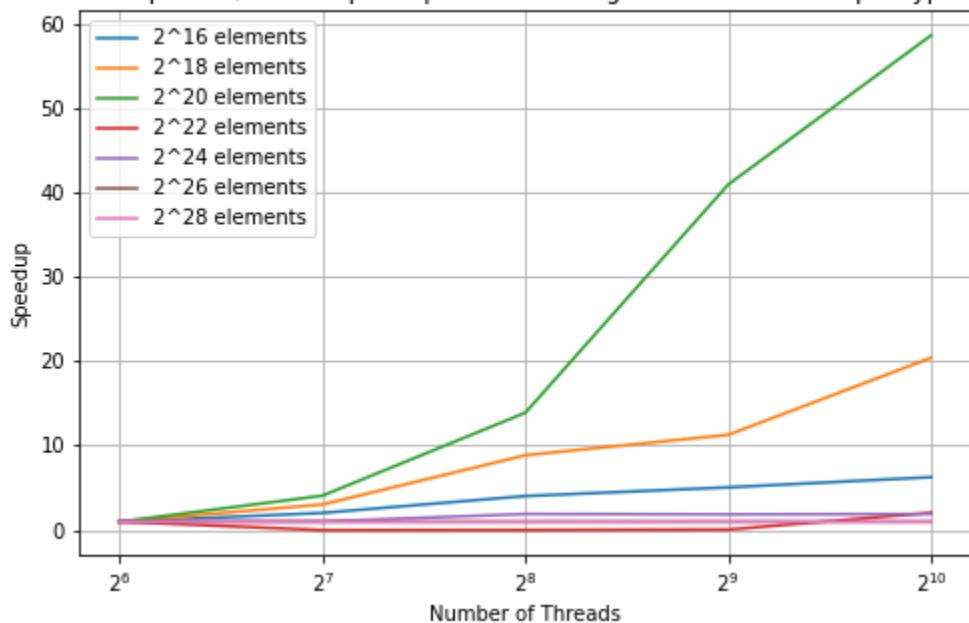
SampleSort, CUDA: Speedup of "comm" region with "Random" input type



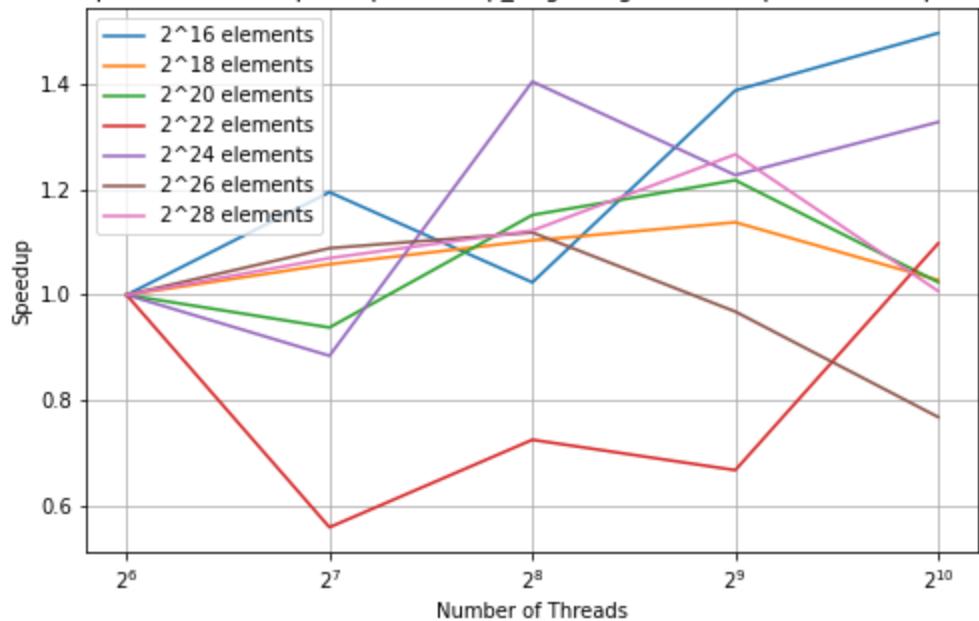
SampleSort, CUDA: Speedup of "comm" region with "ReverseSorted" input type



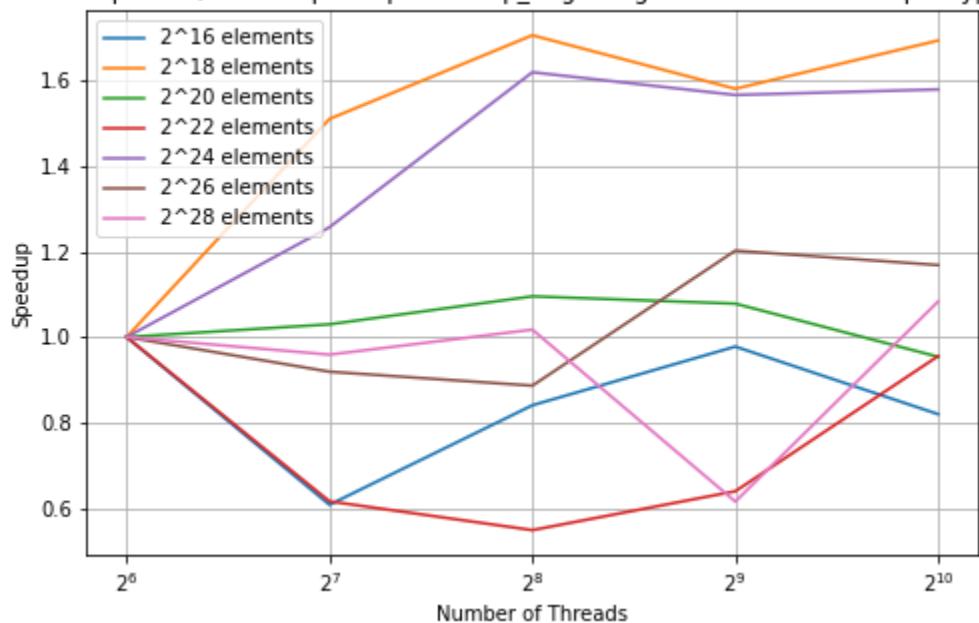
SampleSort, CUDA: Speedup of "comm" region with "Sorted" input type



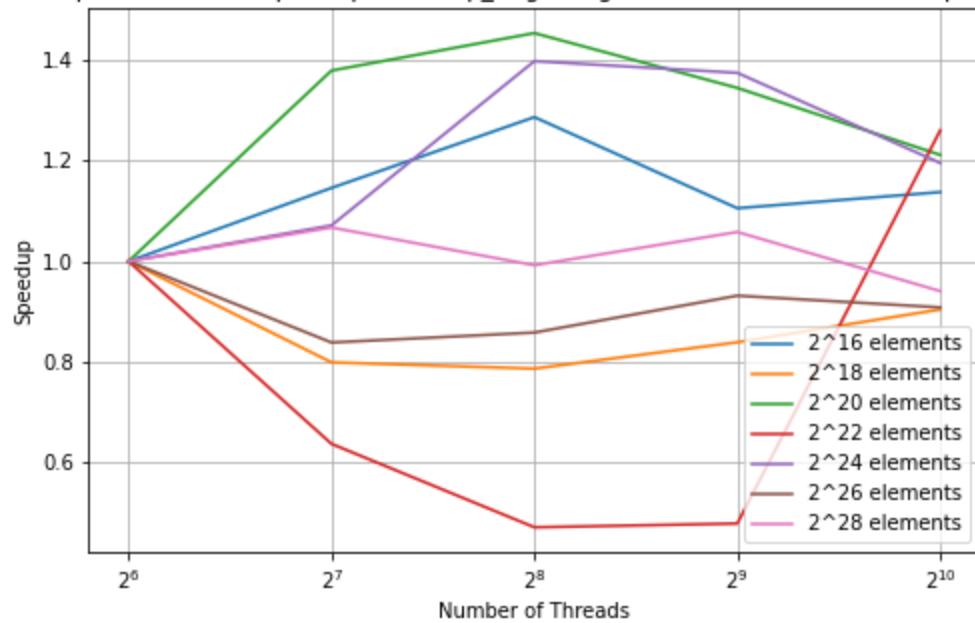
SampleSort, CUDA: Speedup of "comp_large" region with "1perturbed" input type



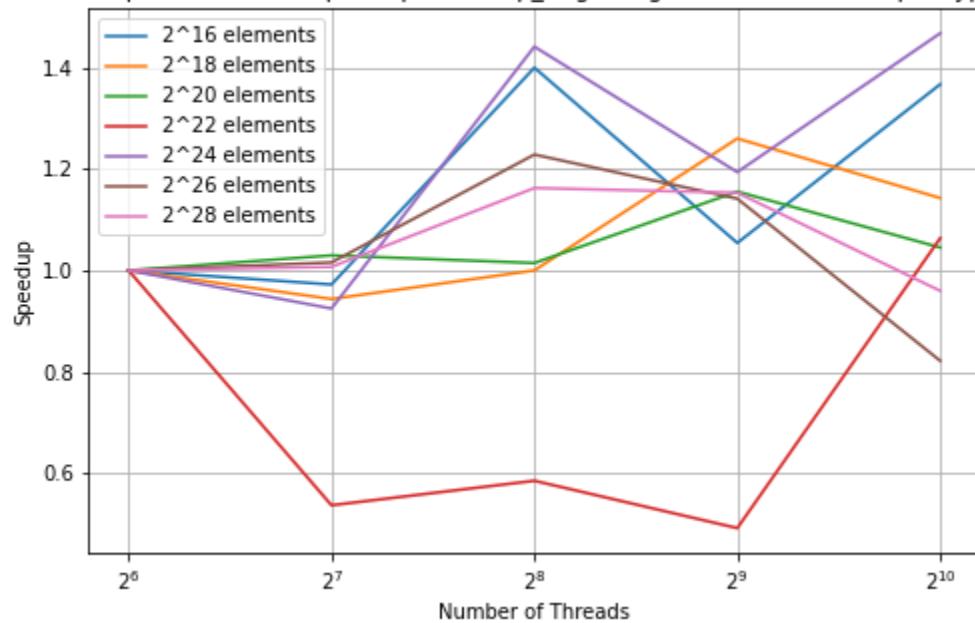
SampleSort, CUDA: Speedup of "comp_large" region with "Random" input type



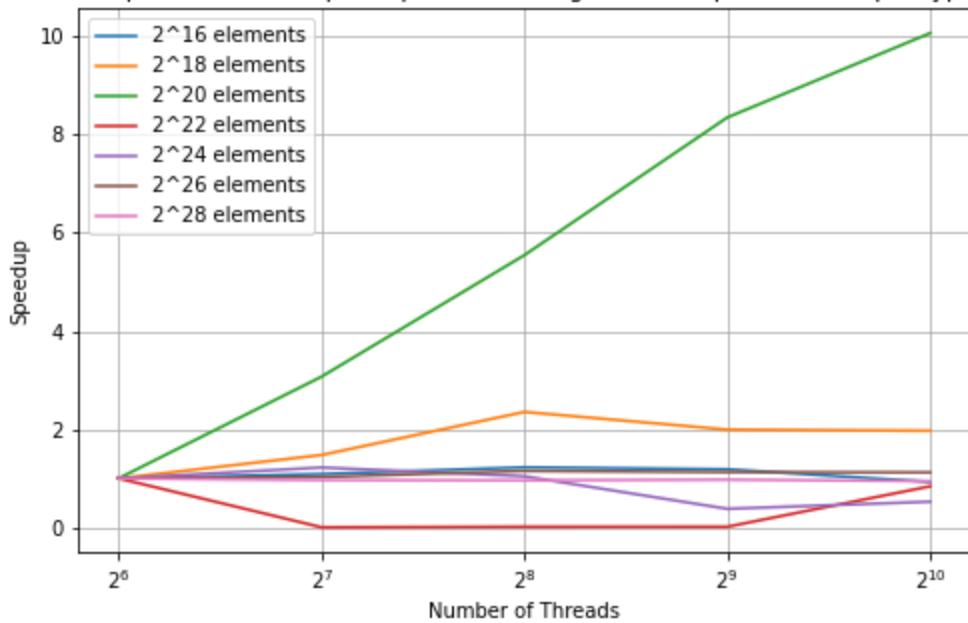
SampleSort, CUDA: Speedup of "comp_large" region with "ReverseSorted" input type



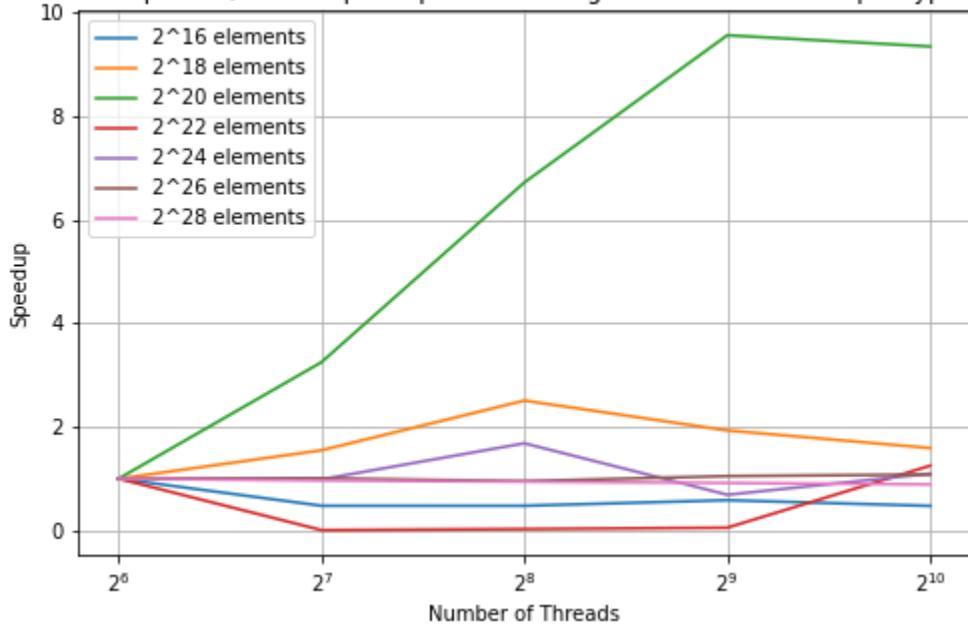
SampleSort, CUDA: Speedup of "comp_large" region with "Sorted" input type



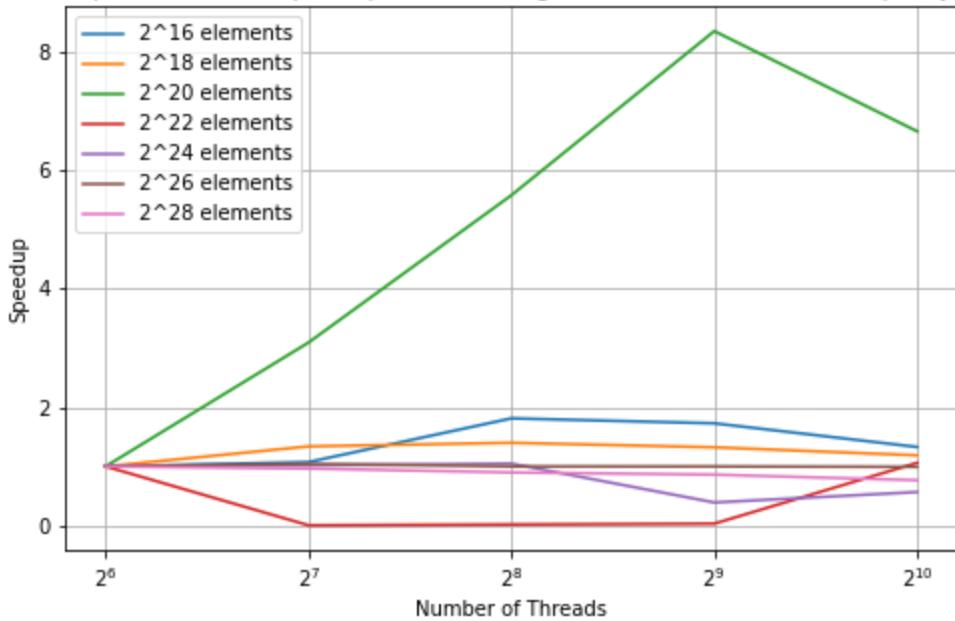
SampleSort, CUDA: Speedup of "main" region with "1perturbed" input type



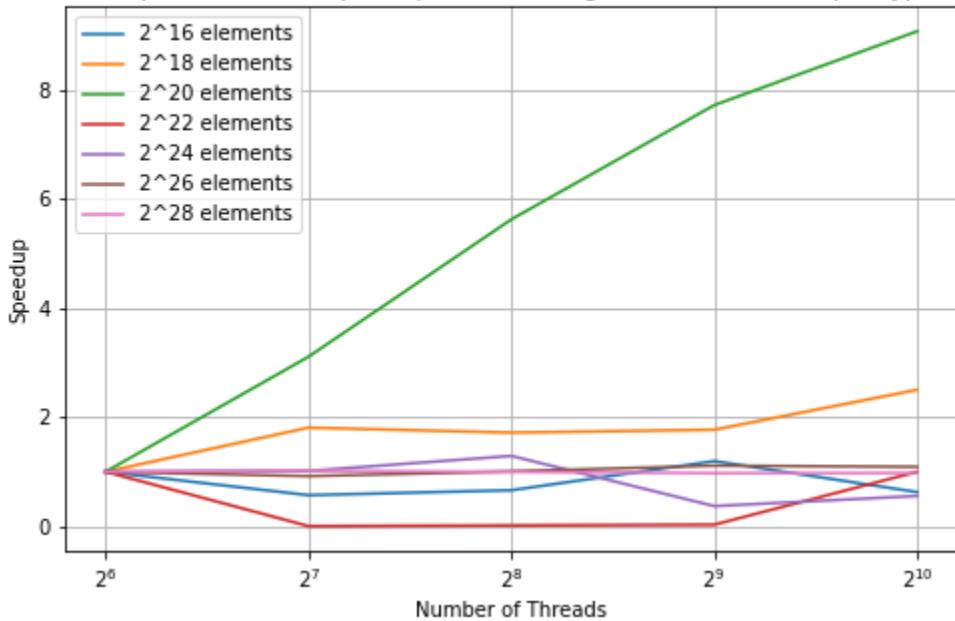
SampleSort, CUDA: Speedup of "main" region with "Random" input type



SampleSort, CUDA: Speedup of "main" region with "ReverseSorted" input type



SampleSort, CUDA: Speedup of "main" region with "Sorted" input type



Analysis

The strong scaling comm graphs show an overall trend of computation time going down as the number of threads increases. This is reflected in the speedup graph of comm as well with a speedup of nearly 60x for 2²⁰ elements.

On the other hand, computation seems to be completely independent of the number of threads and stays relatively constant. This is shown by all the comp graphs – strong, weak, and speedup – showing no overall trend, and speedup graphs showing only a slight variance. This still shows good scalability in that although computation times don't decrease, they don't increase as well when the number of threads increases.

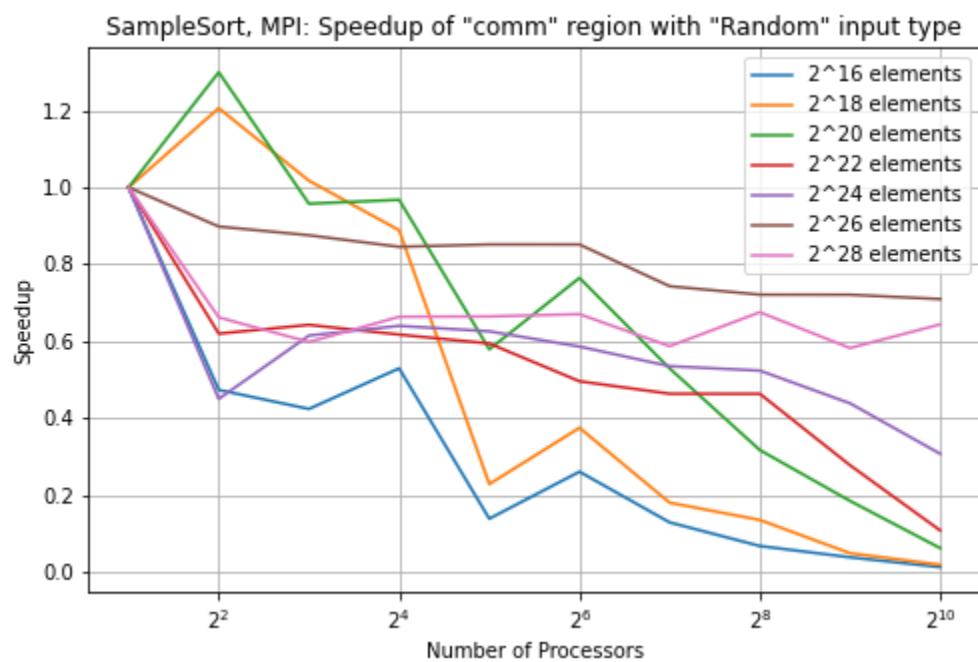
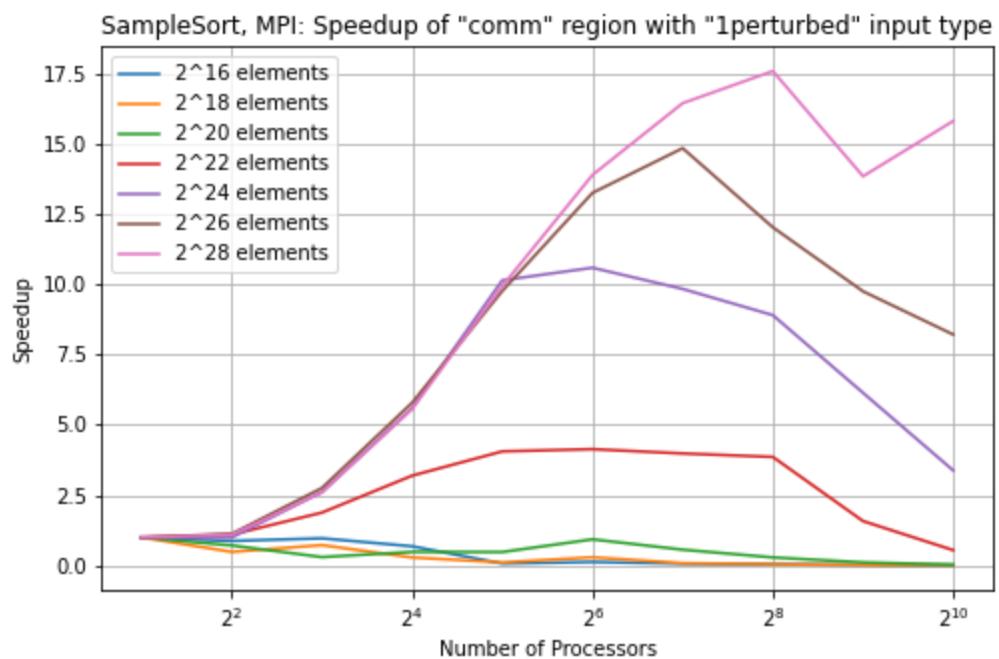
There doesn't appear to be any difference in times between different input types. This is because the sample sort algorithm I implemented will still have to go through all the stages of sample sort from communication to computation no matter how sorted the original input data already is.

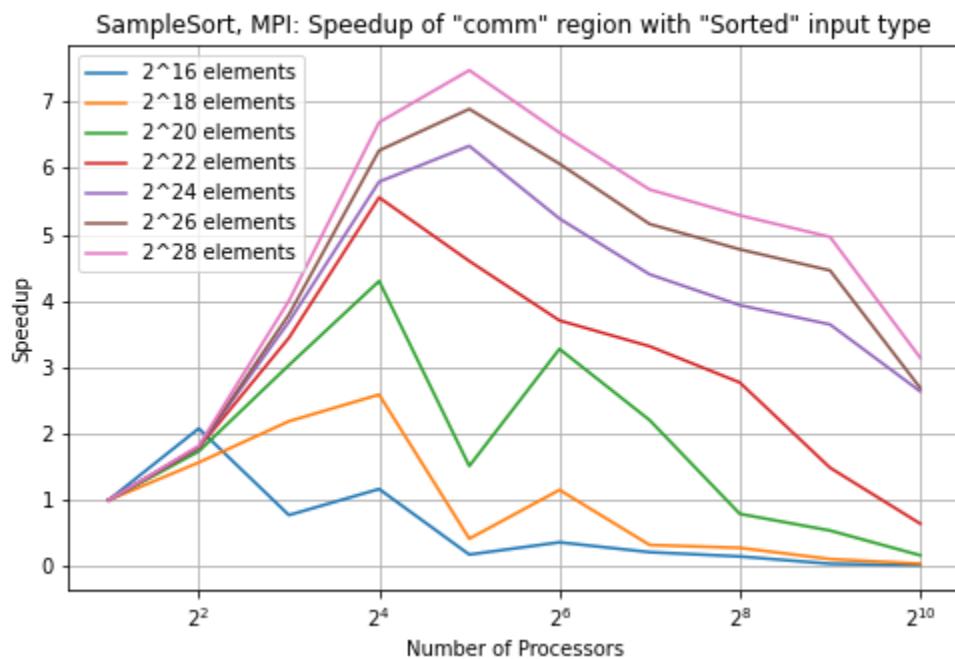
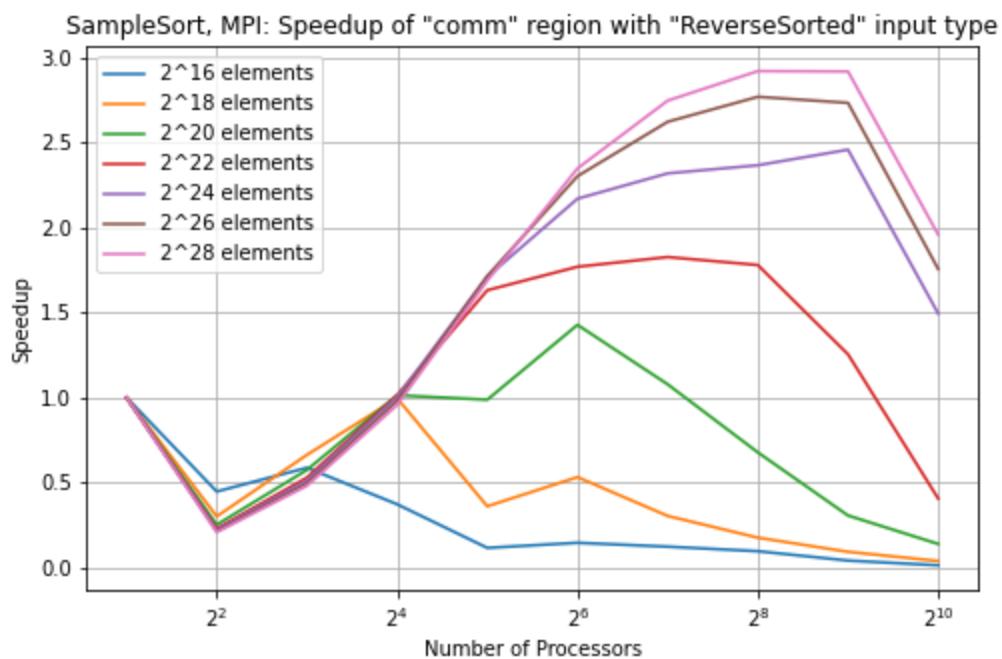
Overall, there is a significant speedup in the algorithm as the number of threads increases as shown by the main speedup graphs.

Sample Sort MPI

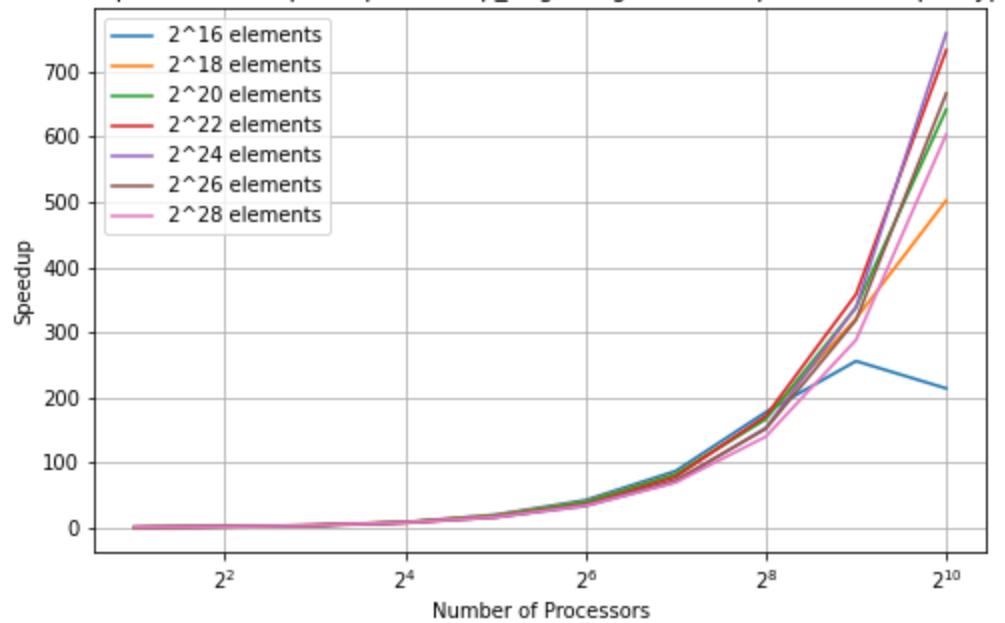
Graphs:

- Speed Up:

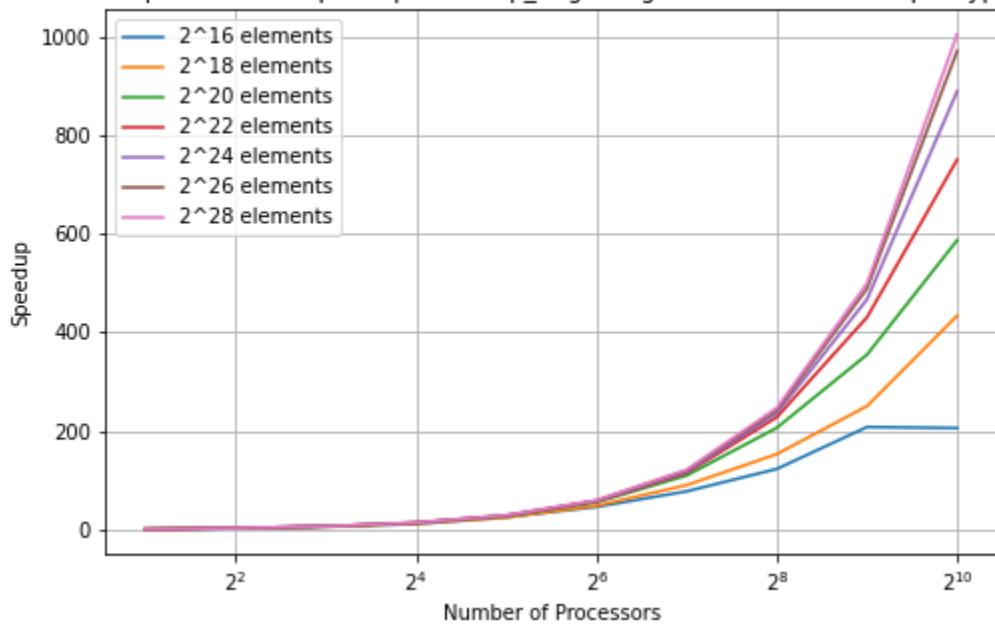




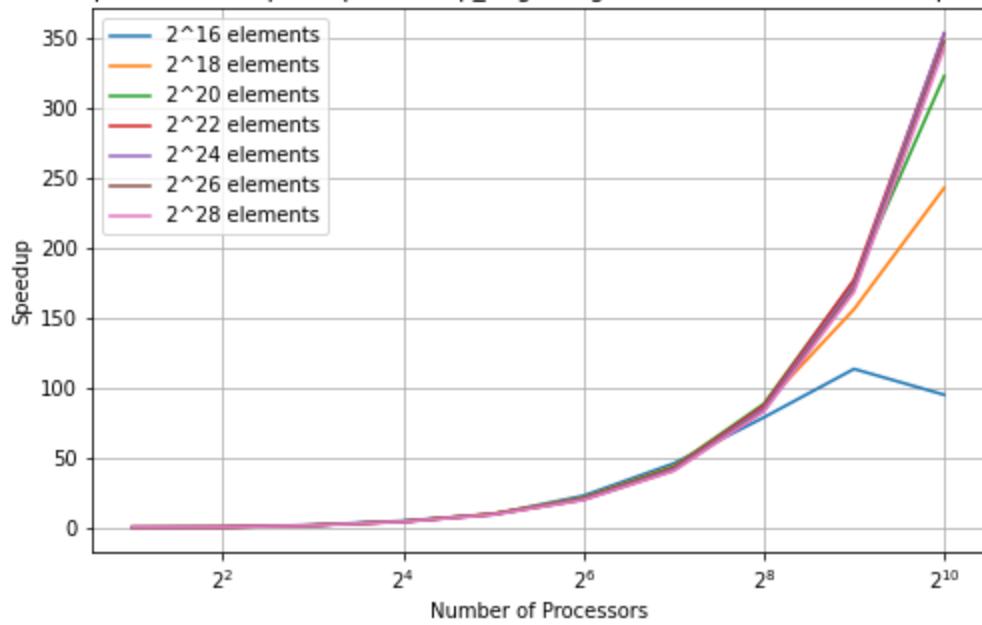
SampleSort, MPI: Speedup of "comp_large" region with "1perturbed" input type



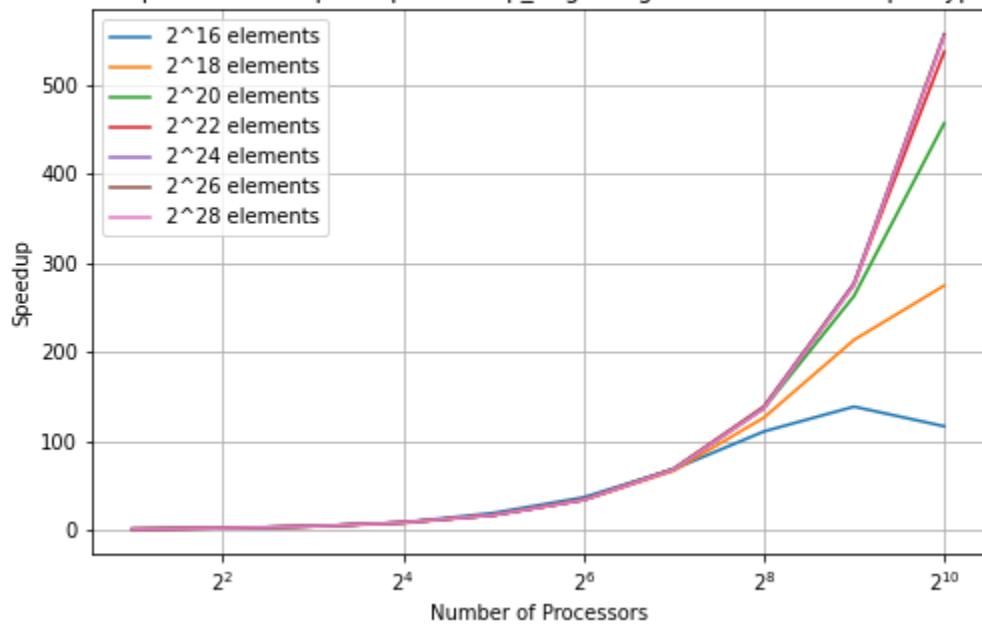
SampleSort, MPI: Speedup of "comp_large" region with "Random" input type



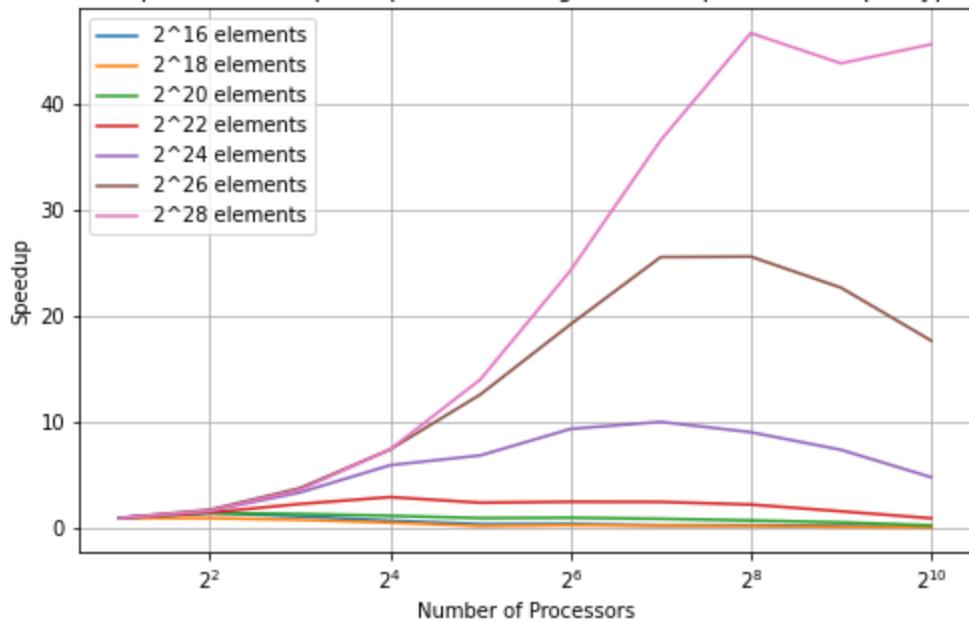
SampleSort, MPI: Speedup of "comp_large" region with "ReverseSorted" input type



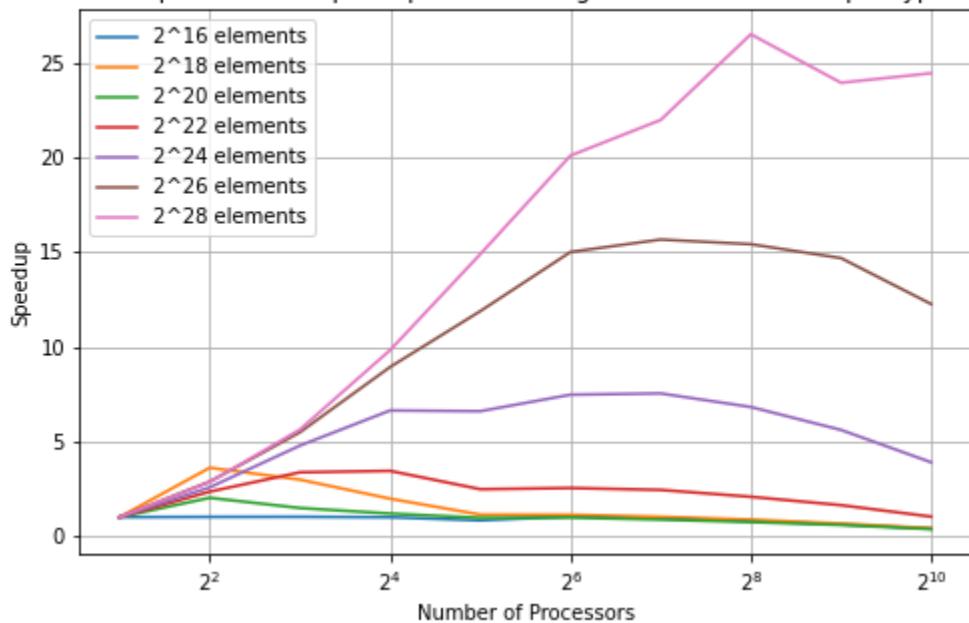
SampleSort, MPI: Speedup of "comp_large" region with "Sorted" input type

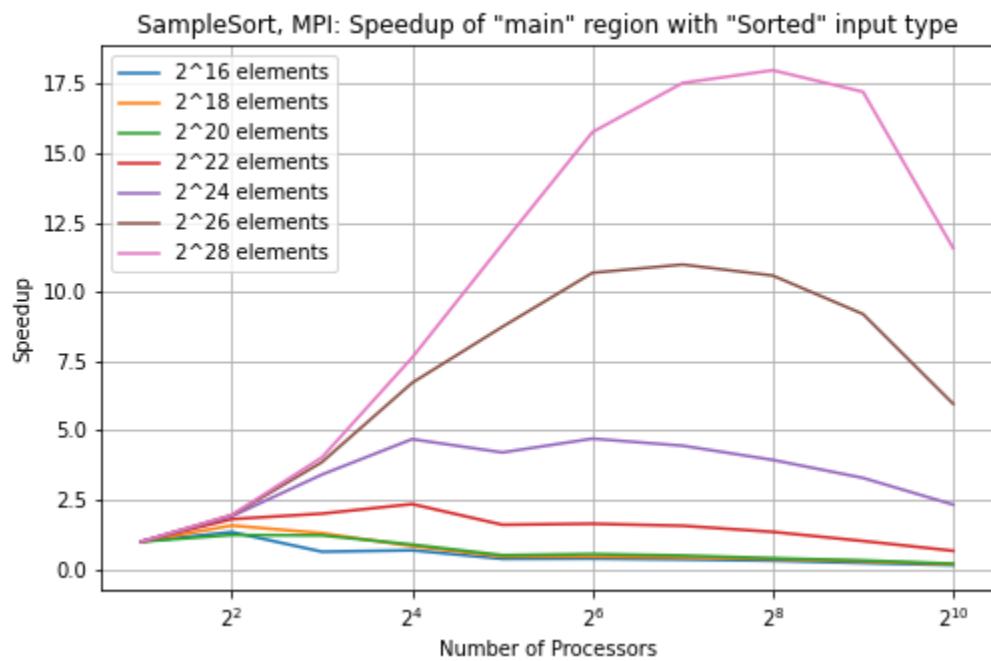
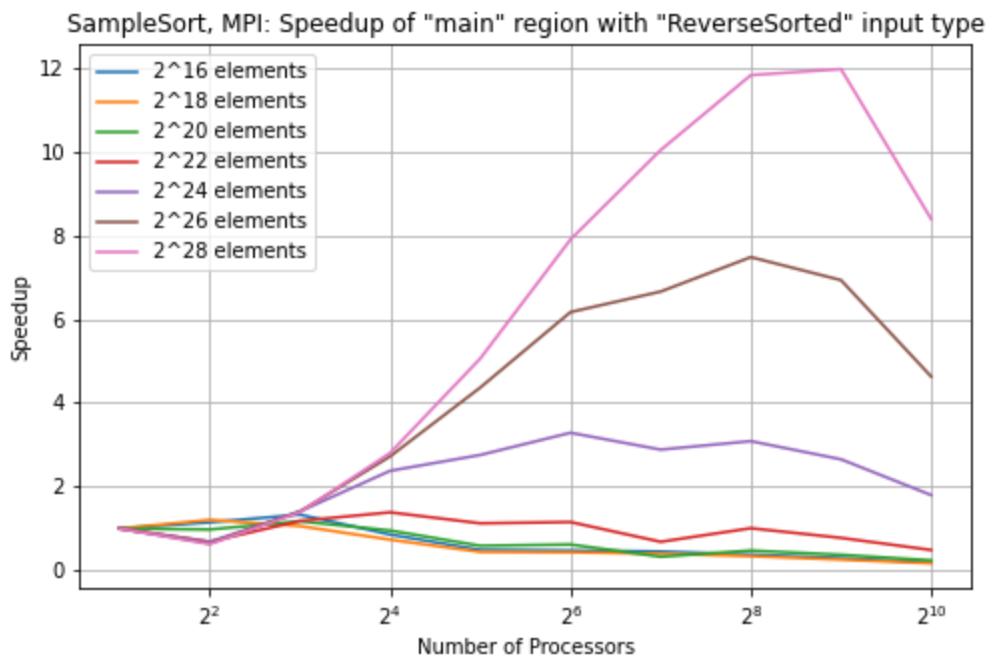


SampleSort, MPI: Speedup of "main" region with "1perturbed" input type

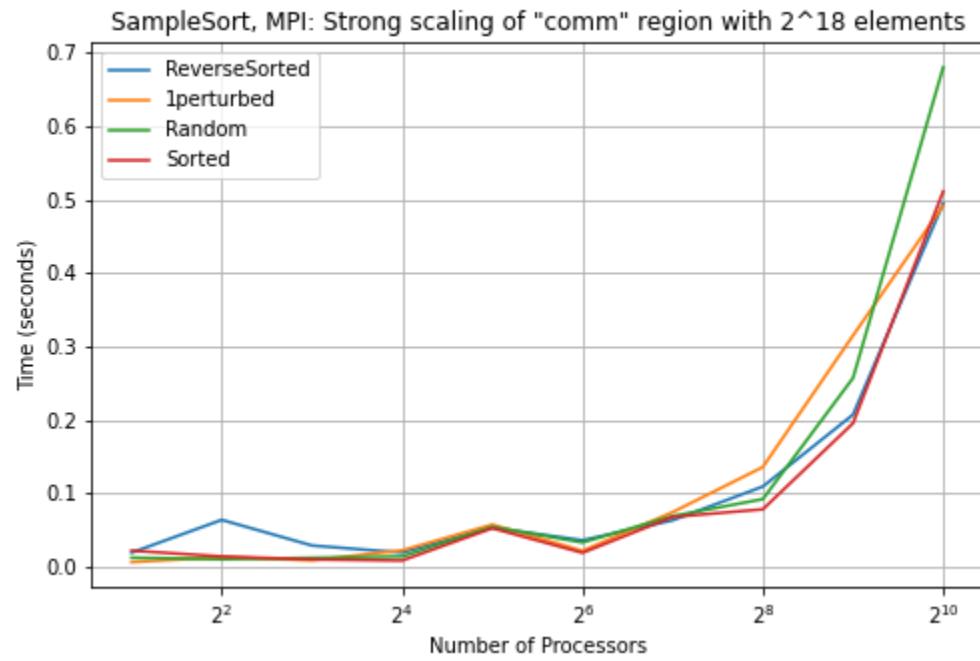
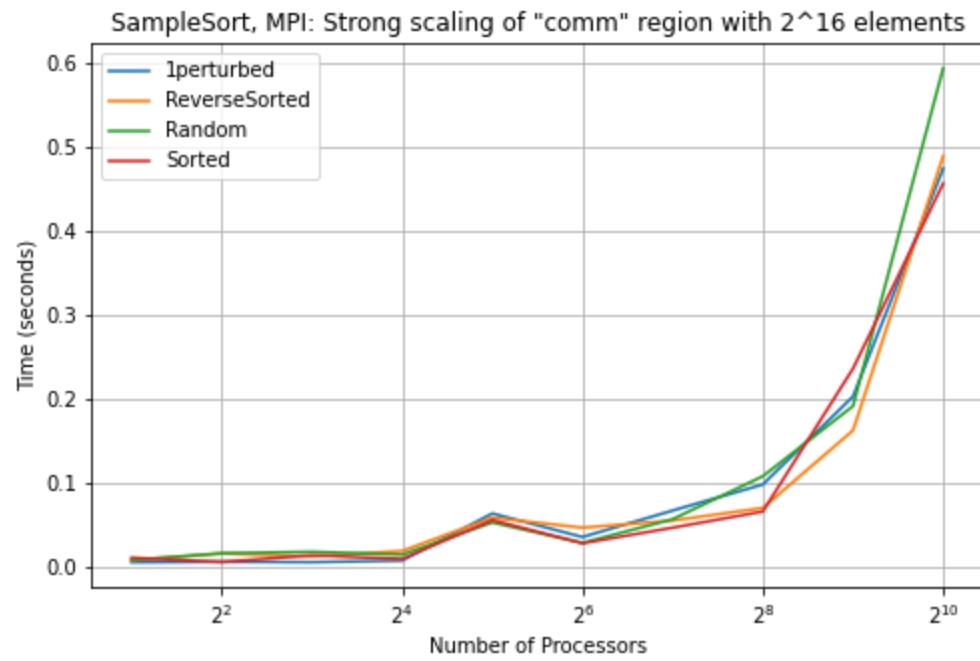


SampleSort, MPI: Speedup of "main" region with "Random" input type

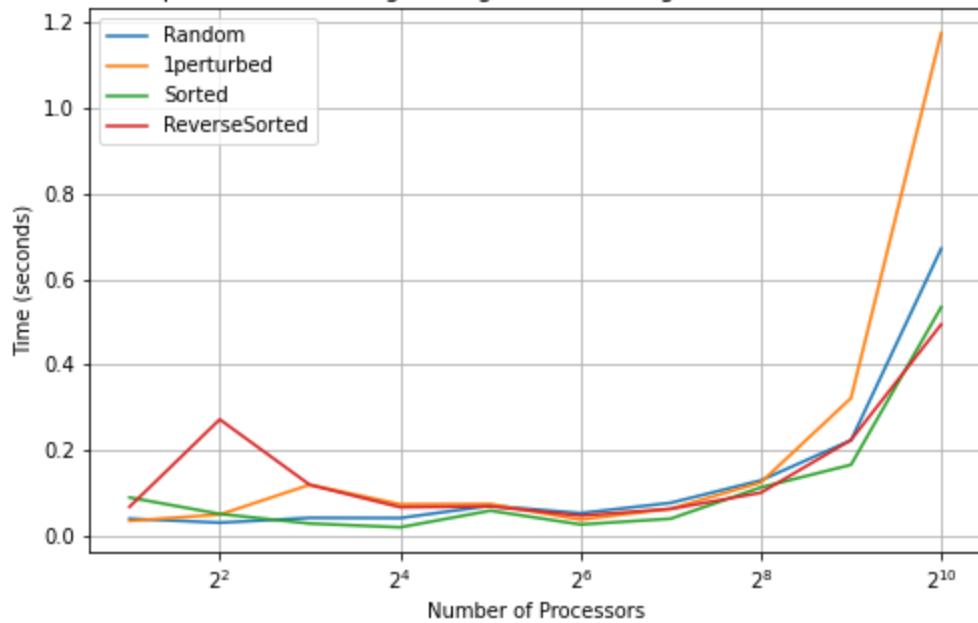




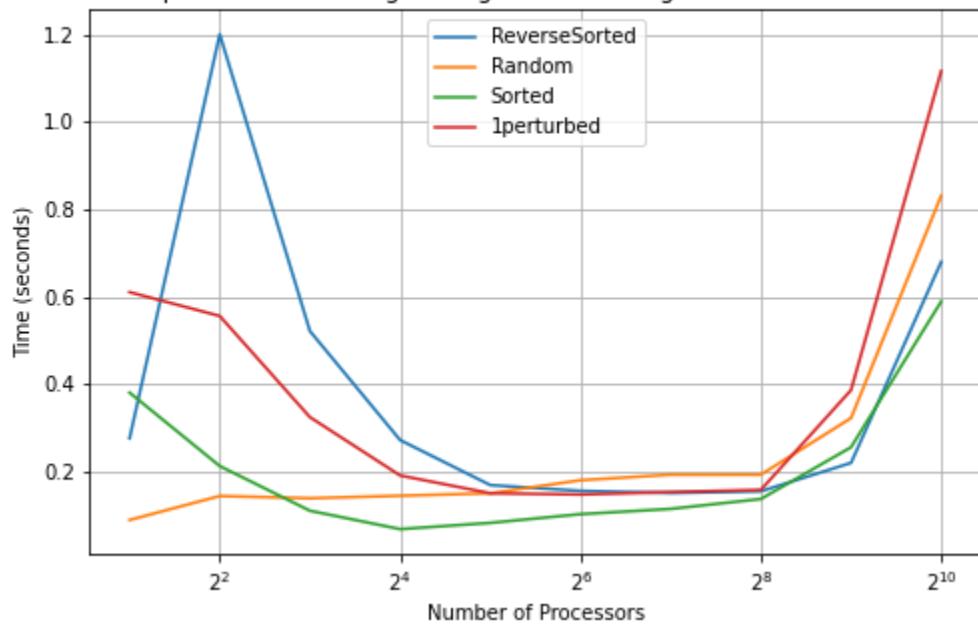
- **Strong Scaling:**



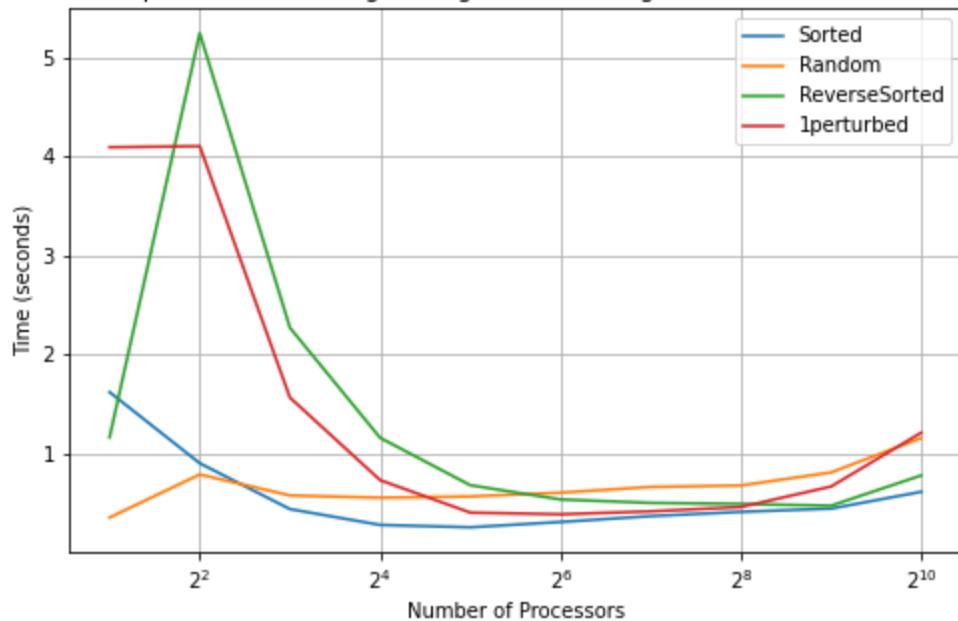
SampleSort, MPI: Strong scaling of "comm" region with 2^{20} elements



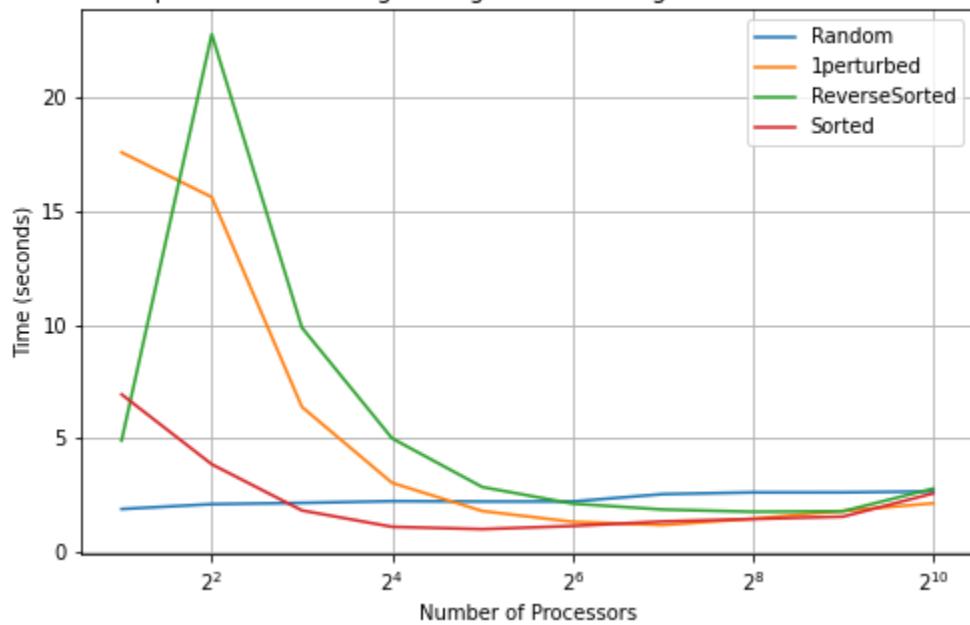
SampleSort, MPI: Strong scaling of "comm" region with 2^{22} elements

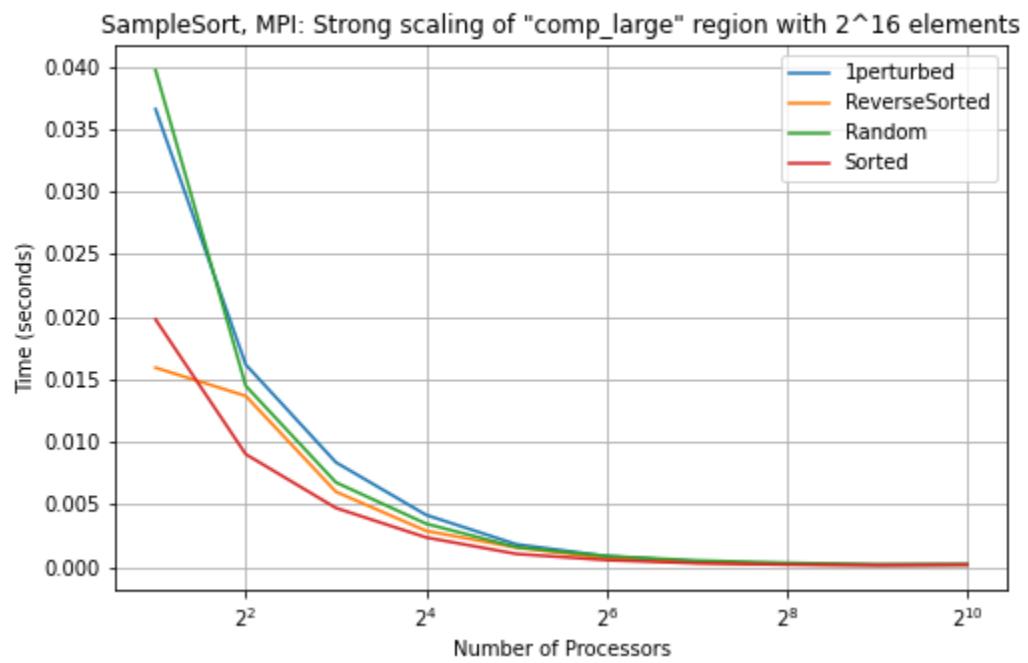
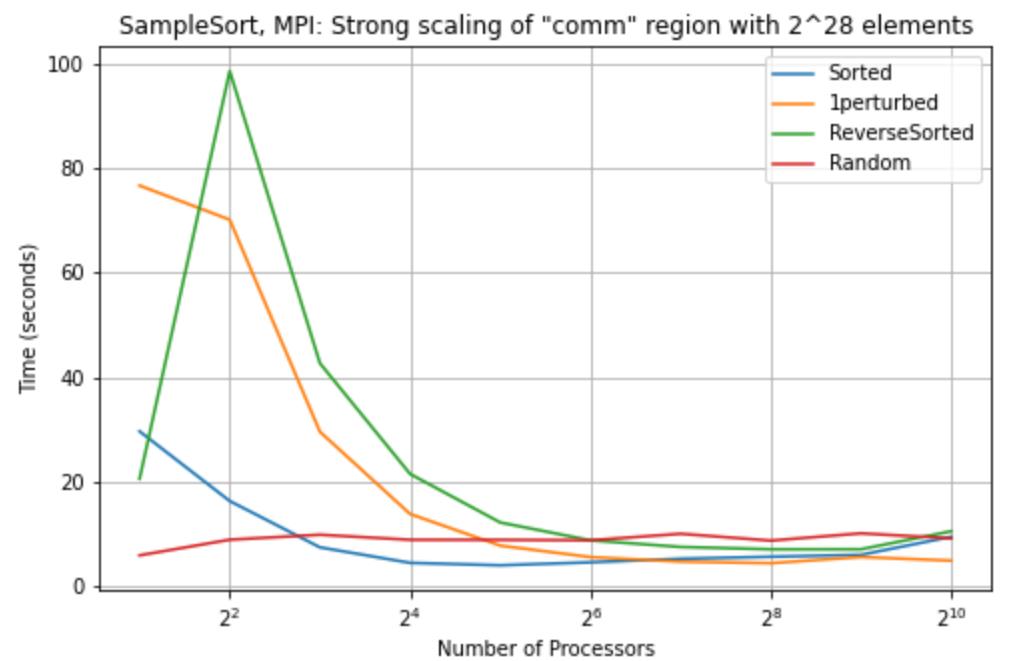


SampleSort, MPI: Strong scaling of "comm" region with 2^{24} elements

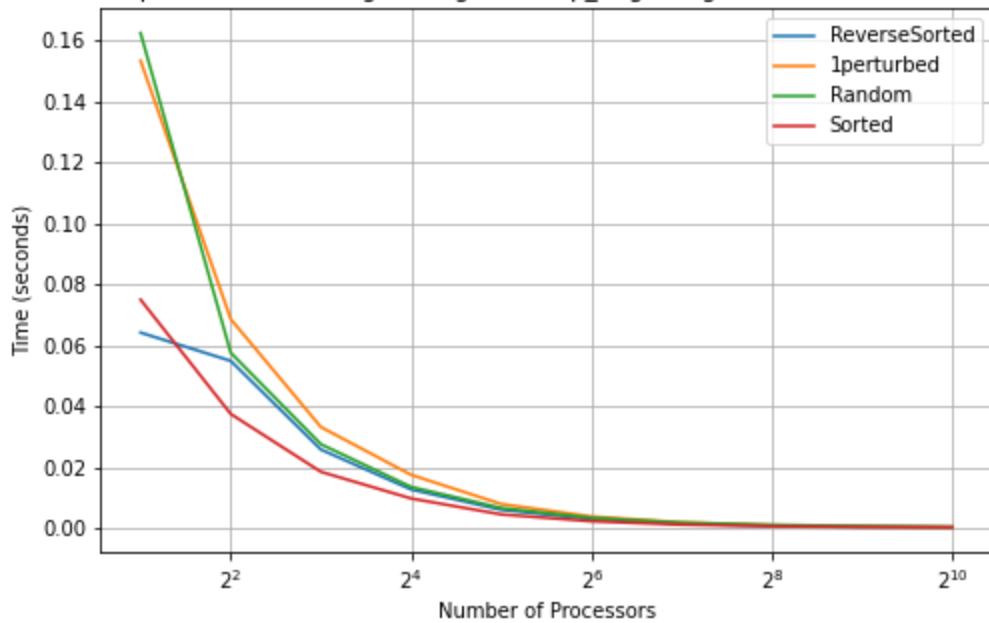


SampleSort, MPI: Strong scaling of "comm" region with 2^{26} elements

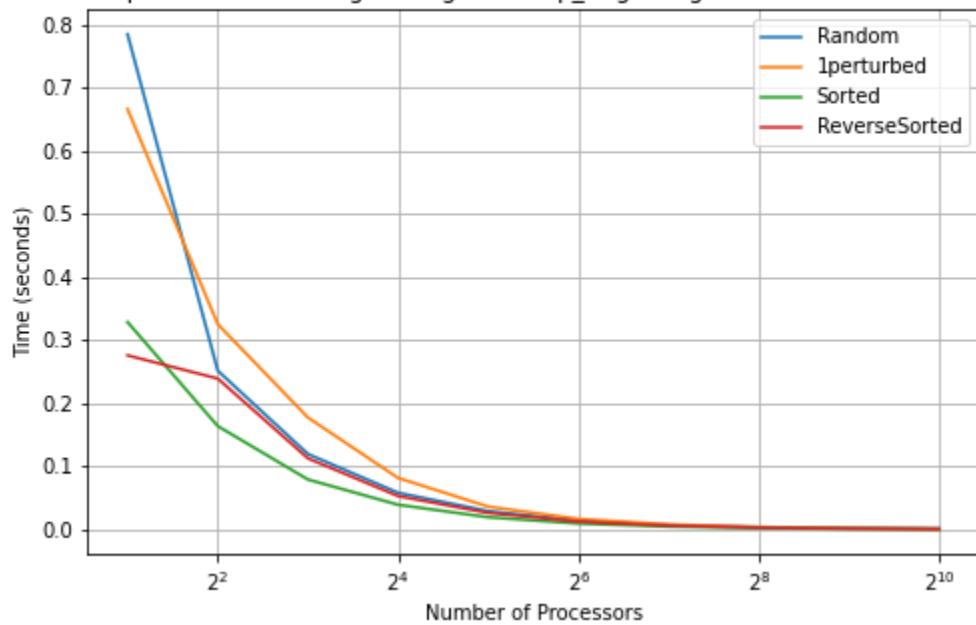


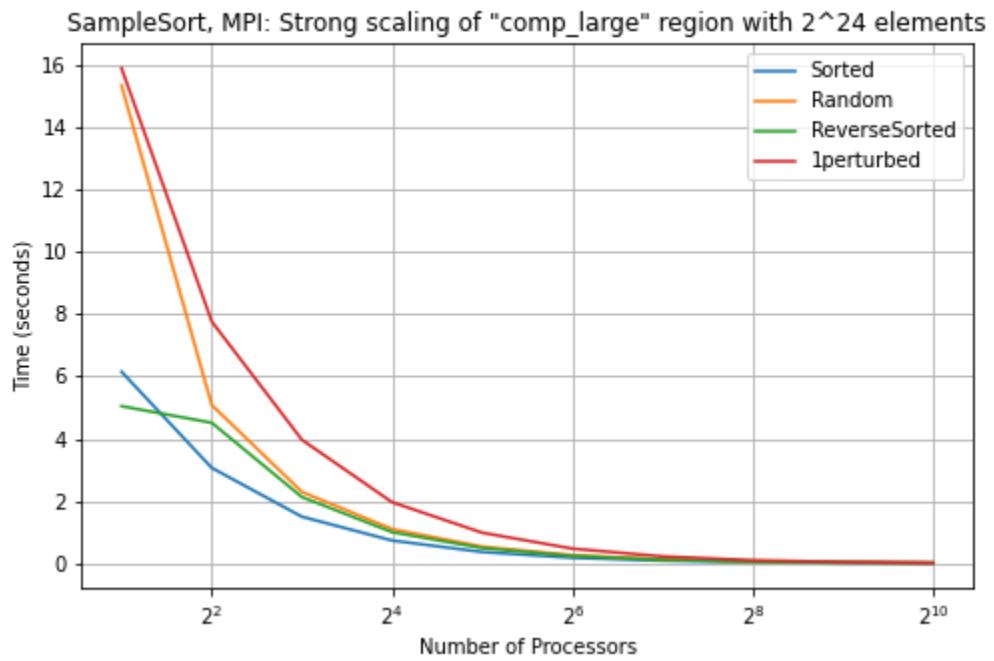
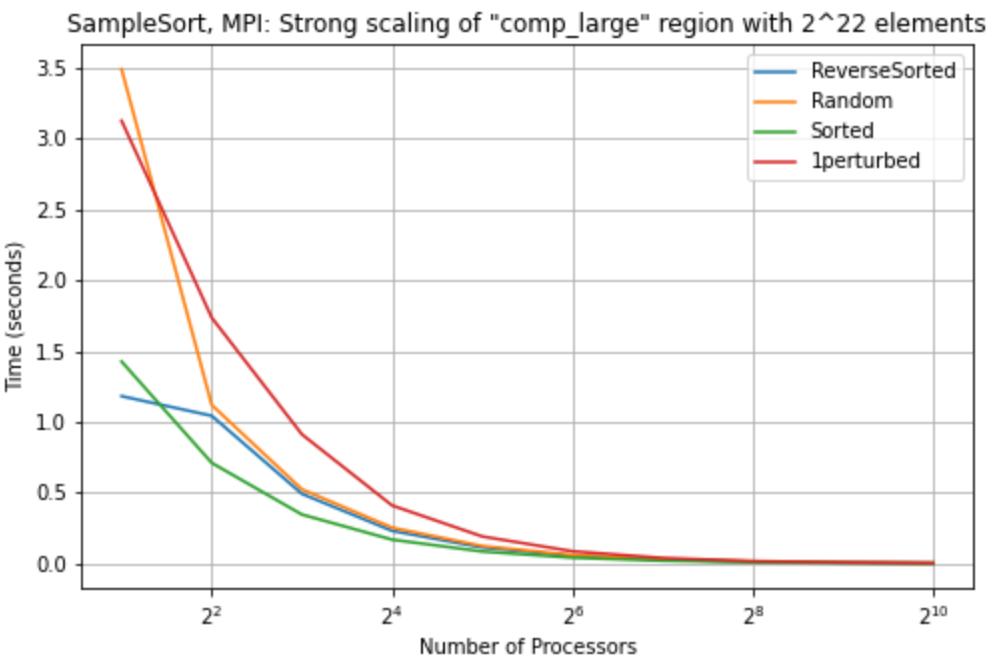


SampleSort, MPI: Strong scaling of "comp_large" region with 2^{18} elements

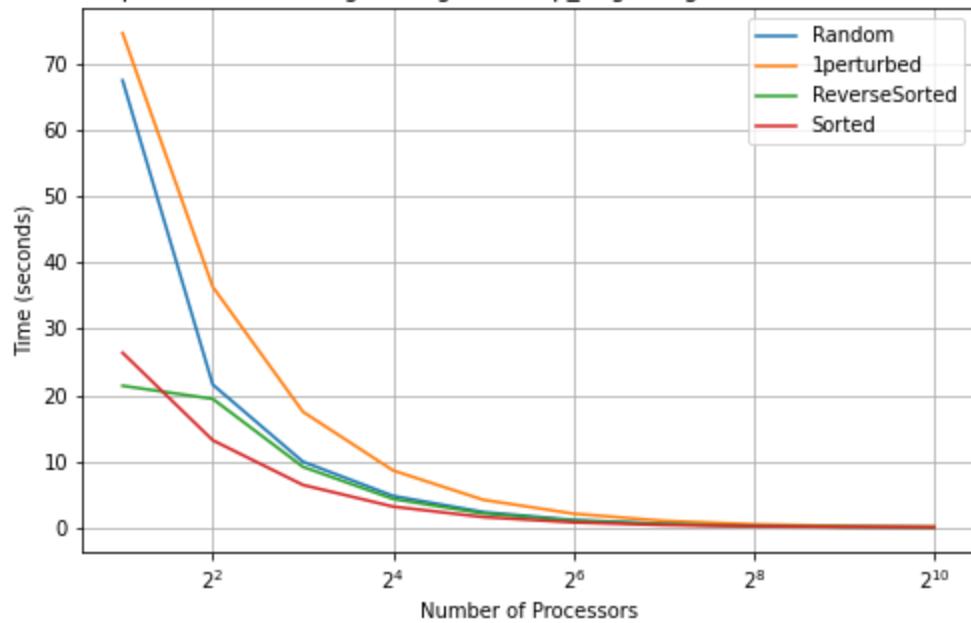


SampleSort, MPI: Strong scaling of "comp_large" region with 2^{20} elements

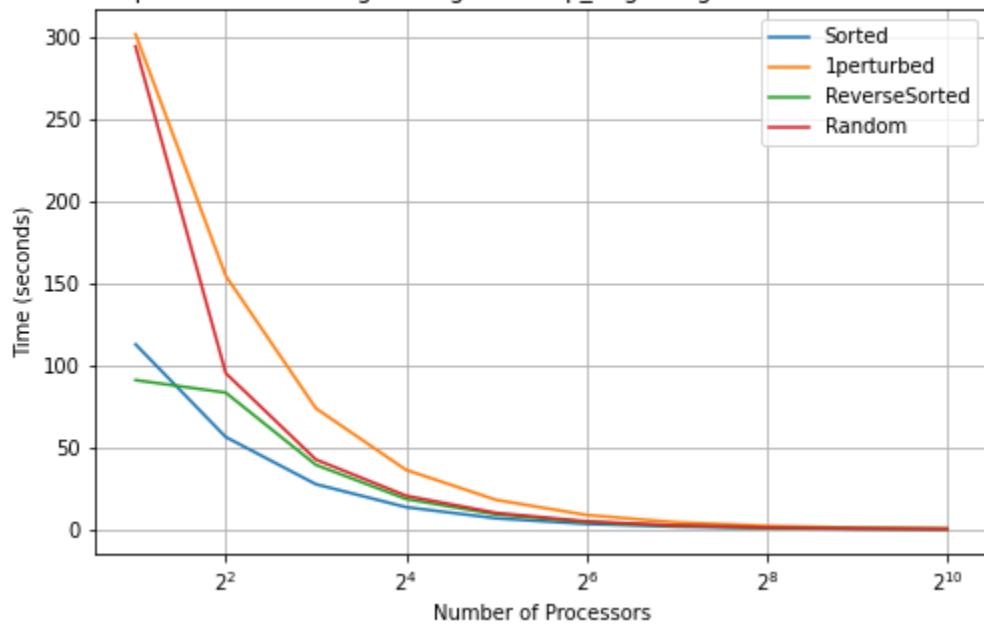


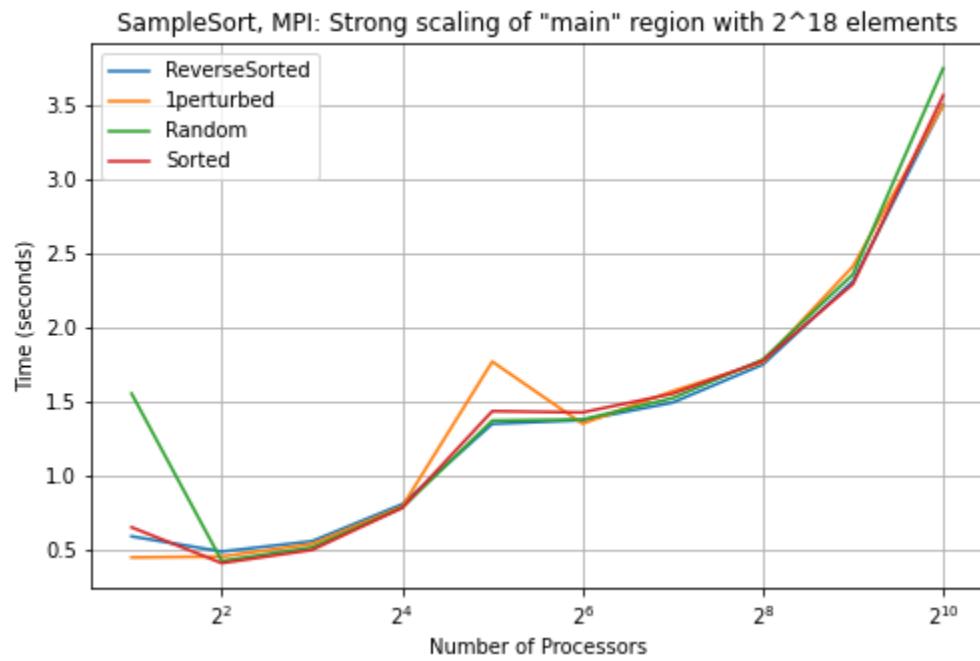
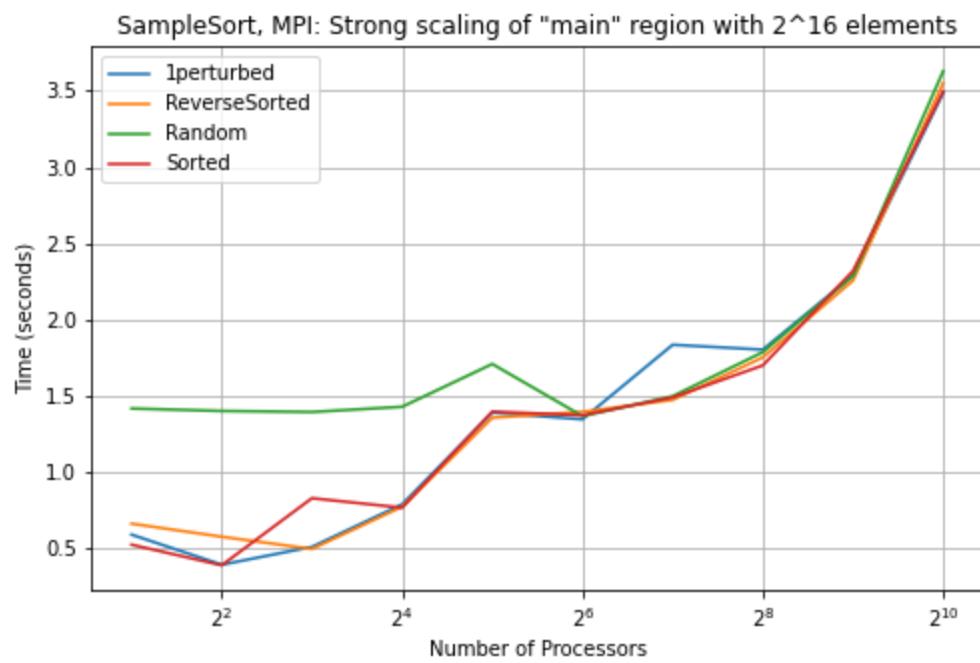


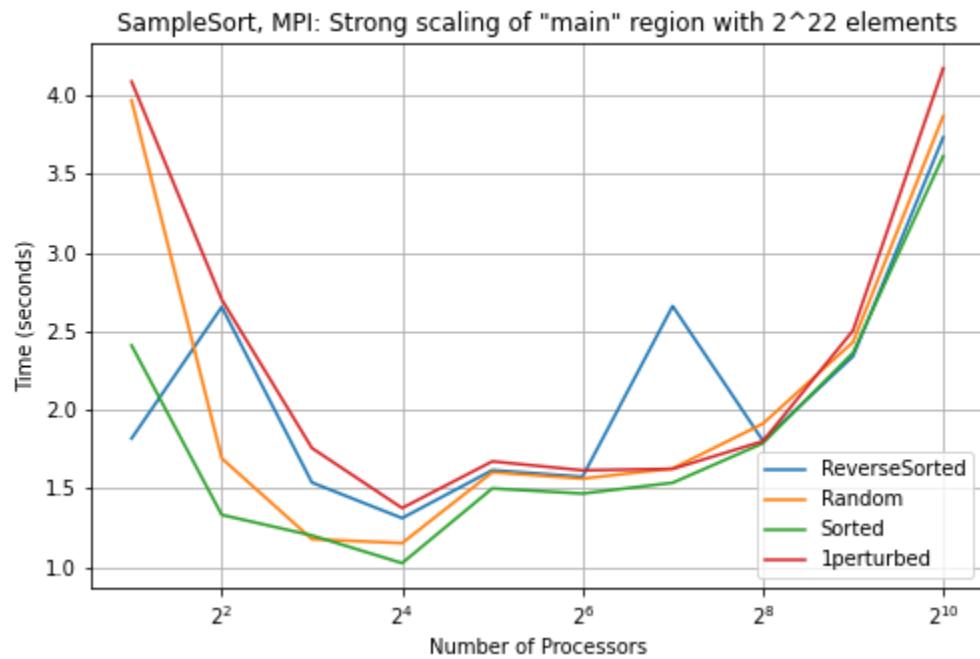
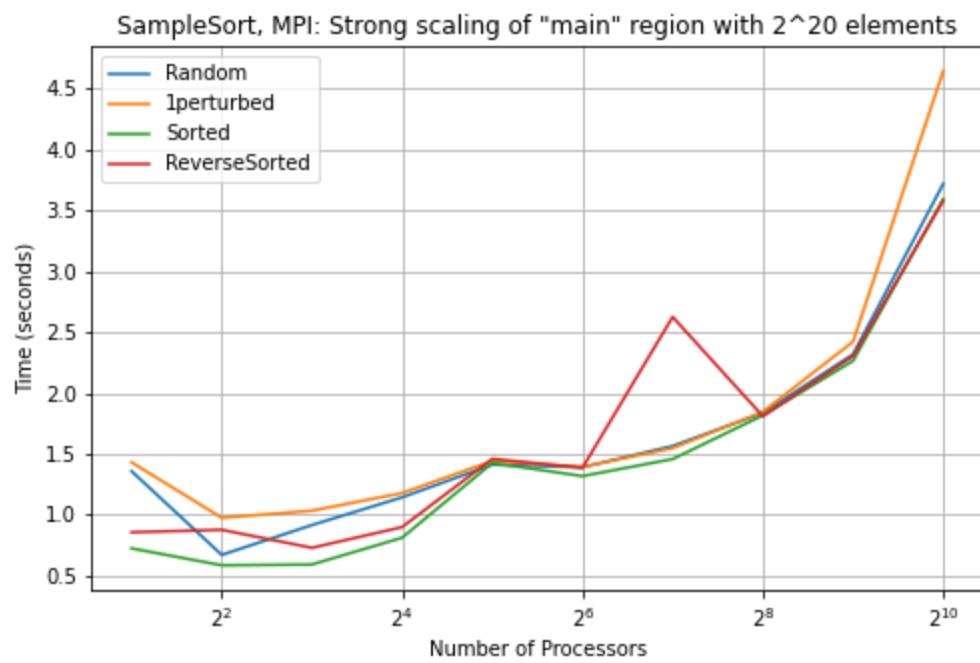
SampleSort, MPI: Strong scaling of "comp_large" region with 2^{26} elements



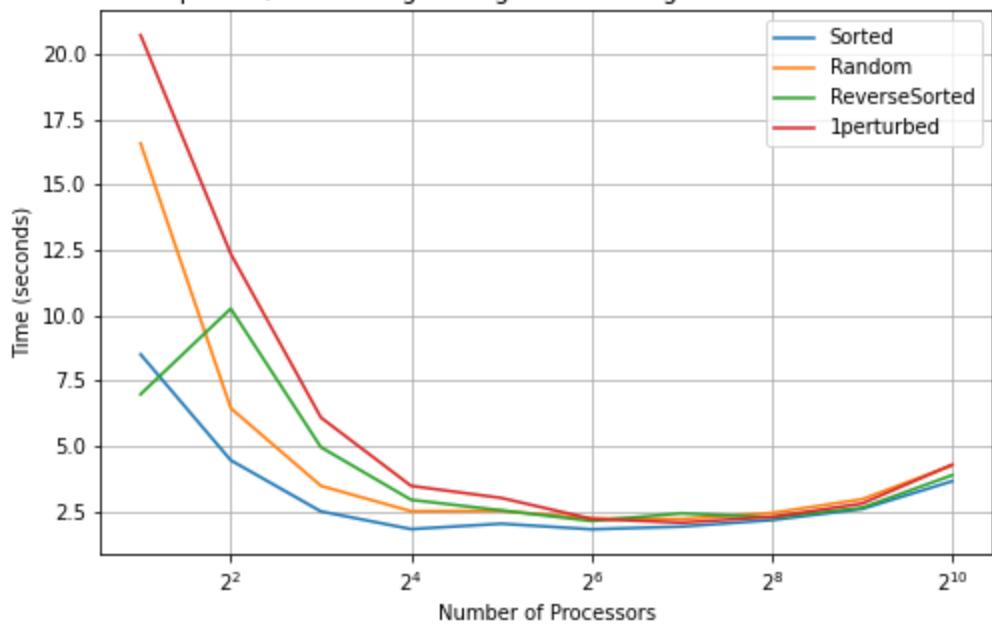
SampleSort, MPI: Strong scaling of "comp_large" region with 2^{28} elements



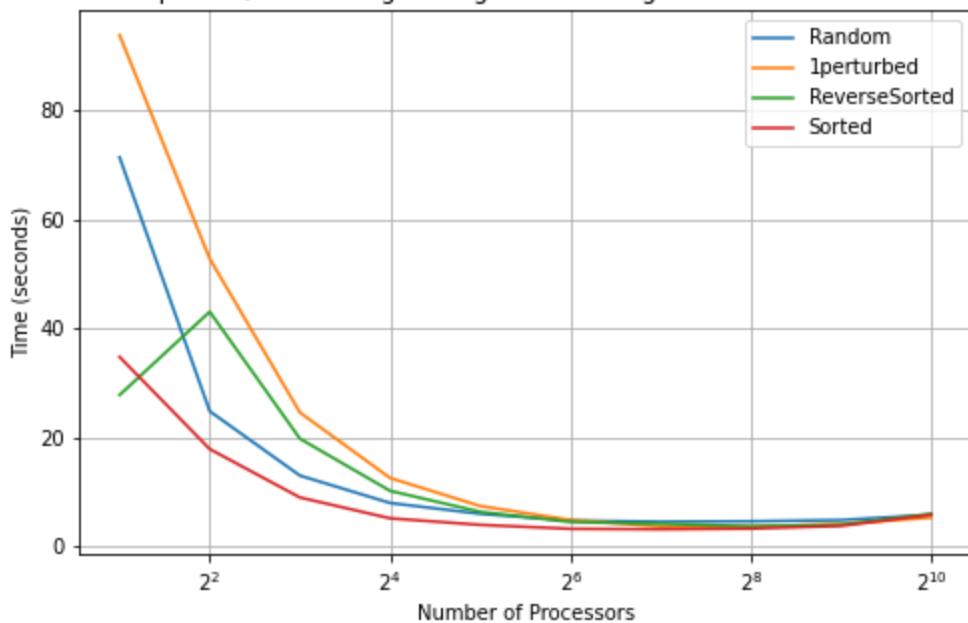




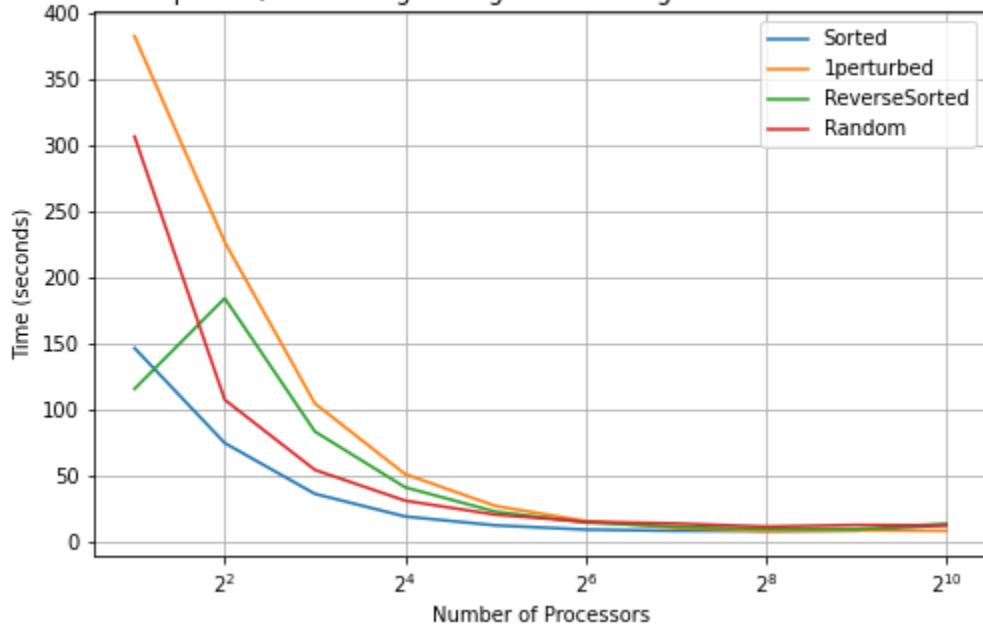
SampleSort, MPI: Strong scaling of "main" region with 2^{24} elements



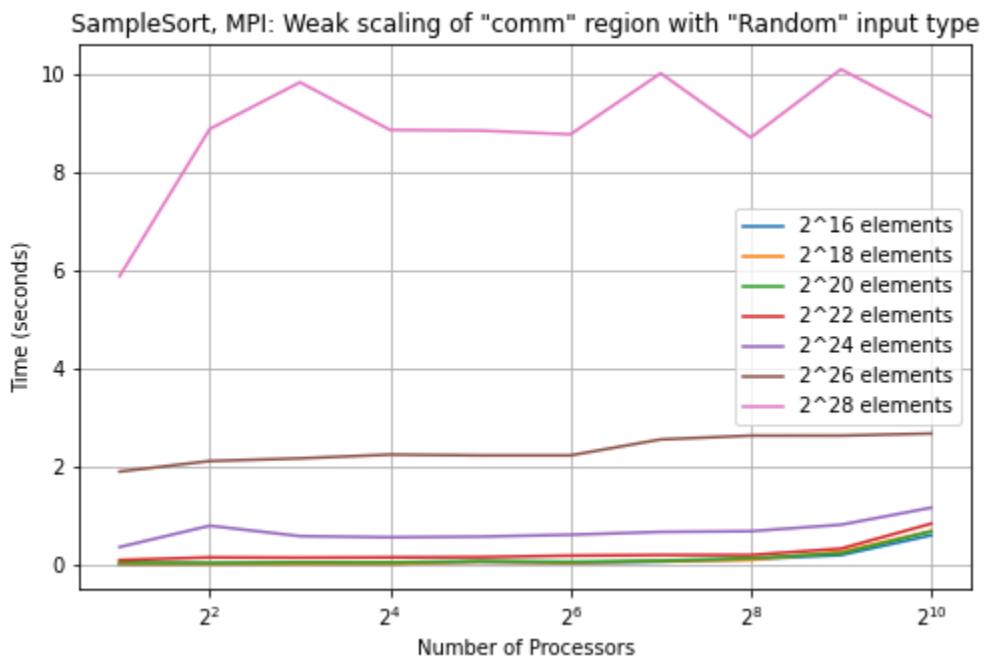
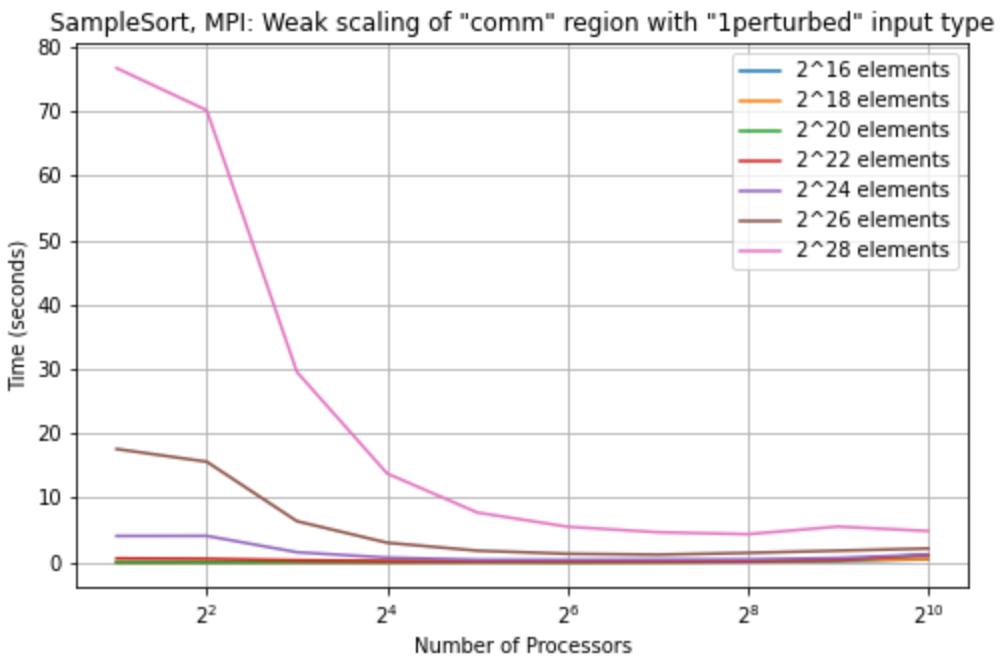
SampleSort, MPI: Strong scaling of "main" region with 2^{26} elements



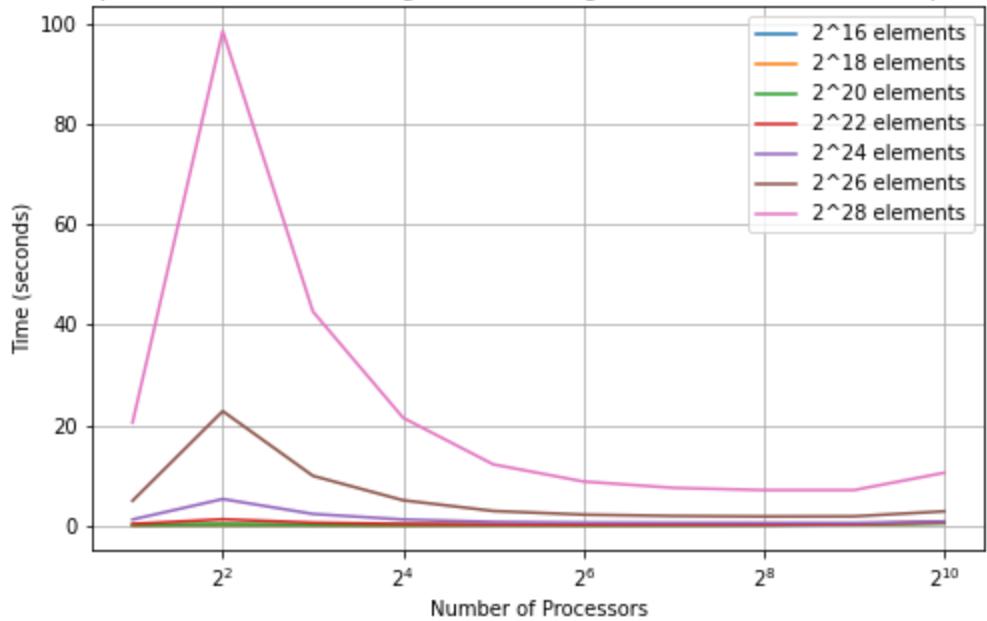
SampleSort, MPI: Strong scaling of "main" region with 2^{28} elements



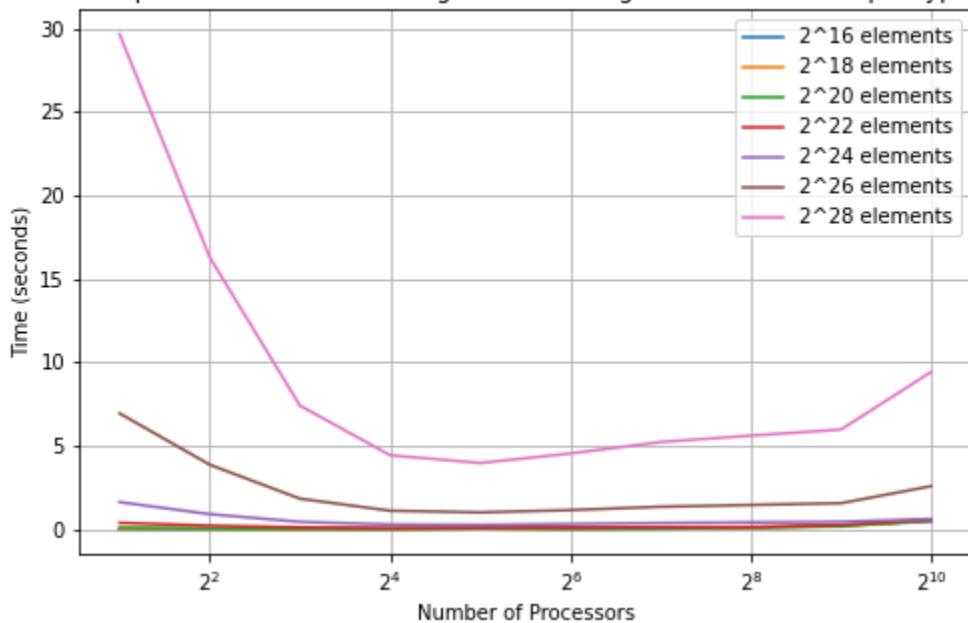
- **Weak Scaling:**



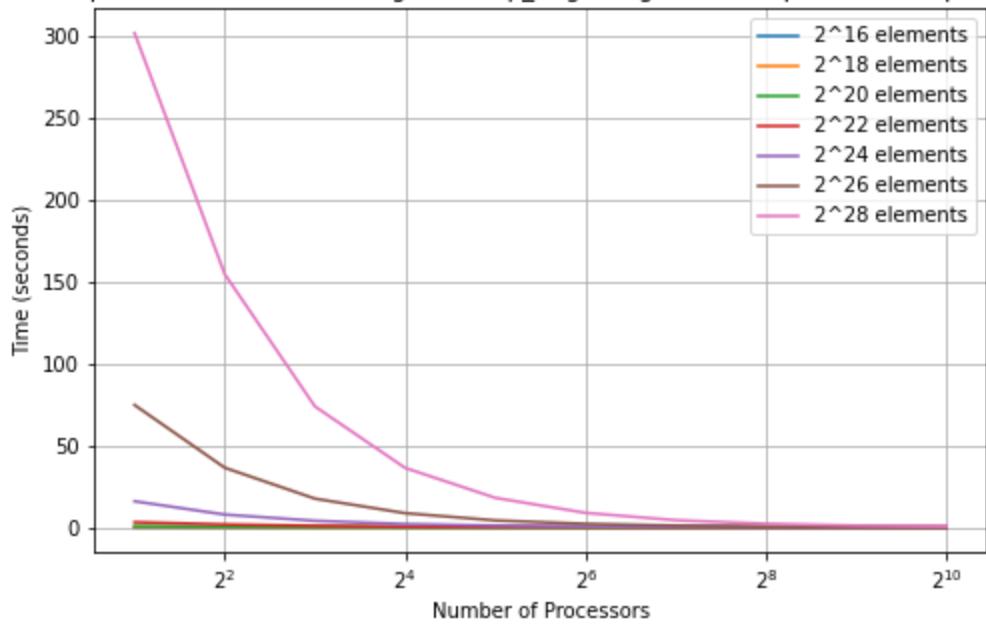
SampleSort, MPI: Weak scaling of "comm" region with "ReverseSorted" input type



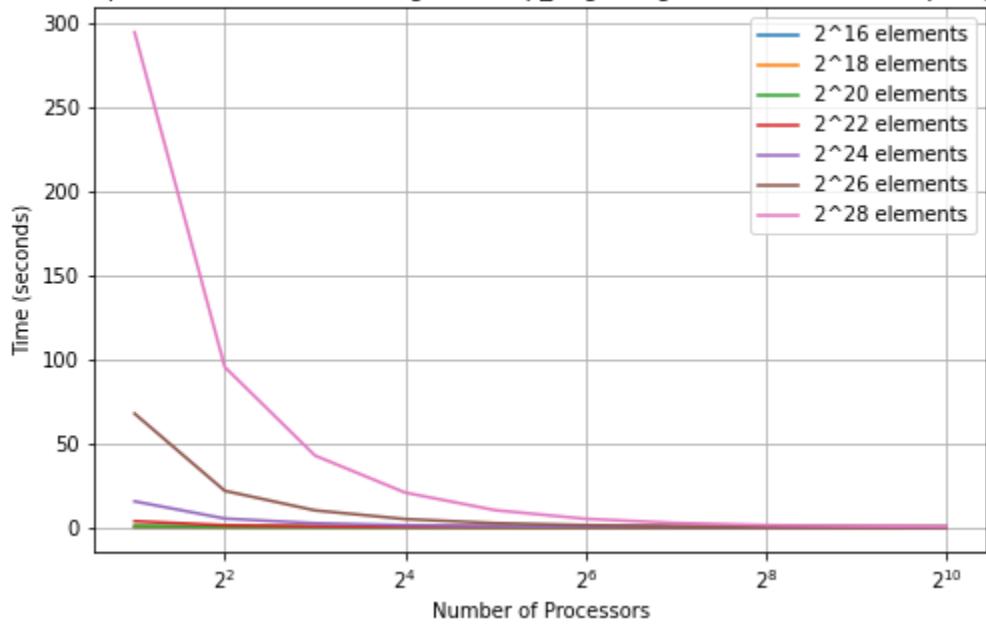
SampleSort, MPI: Weak scaling of "comm" region with "Sorted" input type



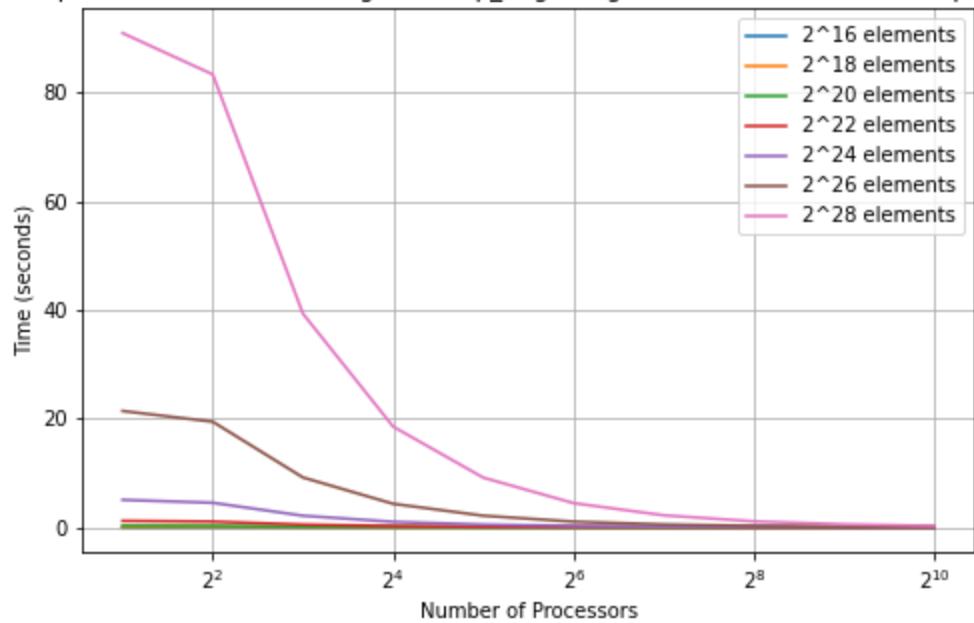
SampleSort, MPI: Weak scaling of "comp_large" region with "1perturbed" input type



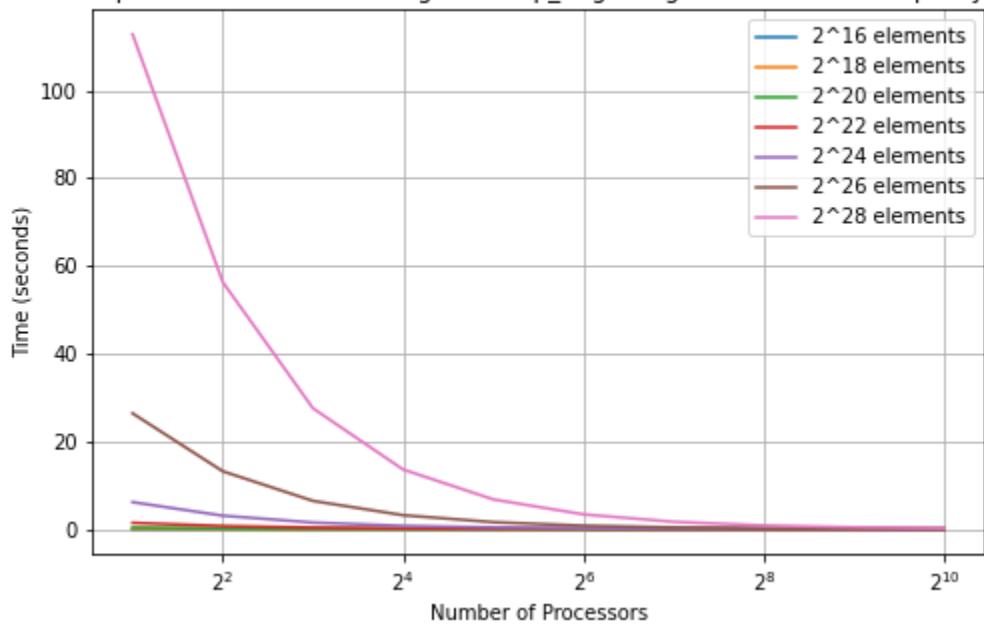
SampleSort, MPI: Weak scaling of "comp_large" region with "Random" input type

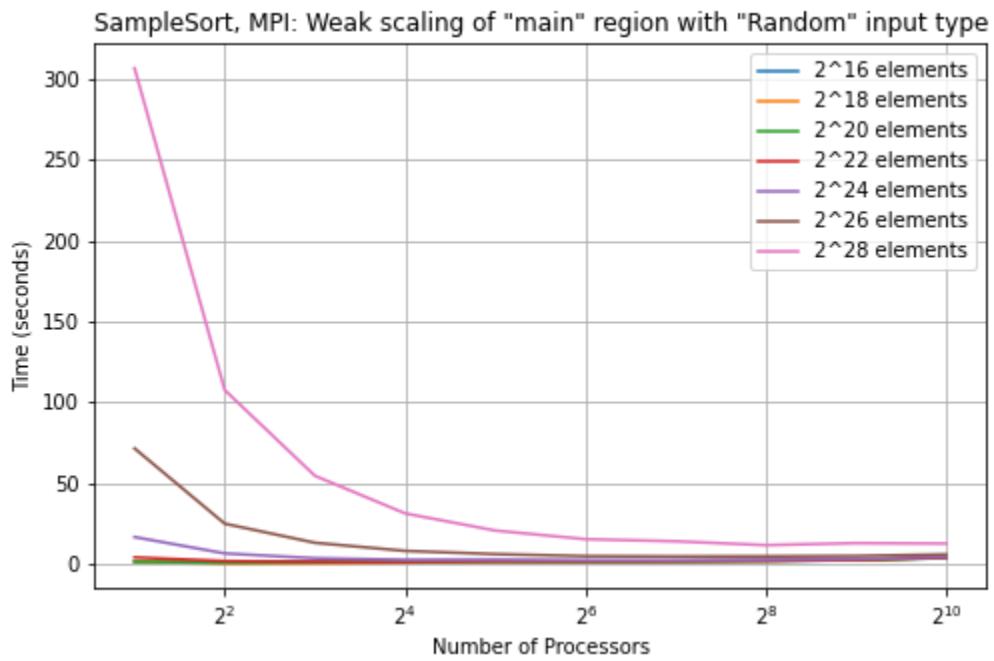
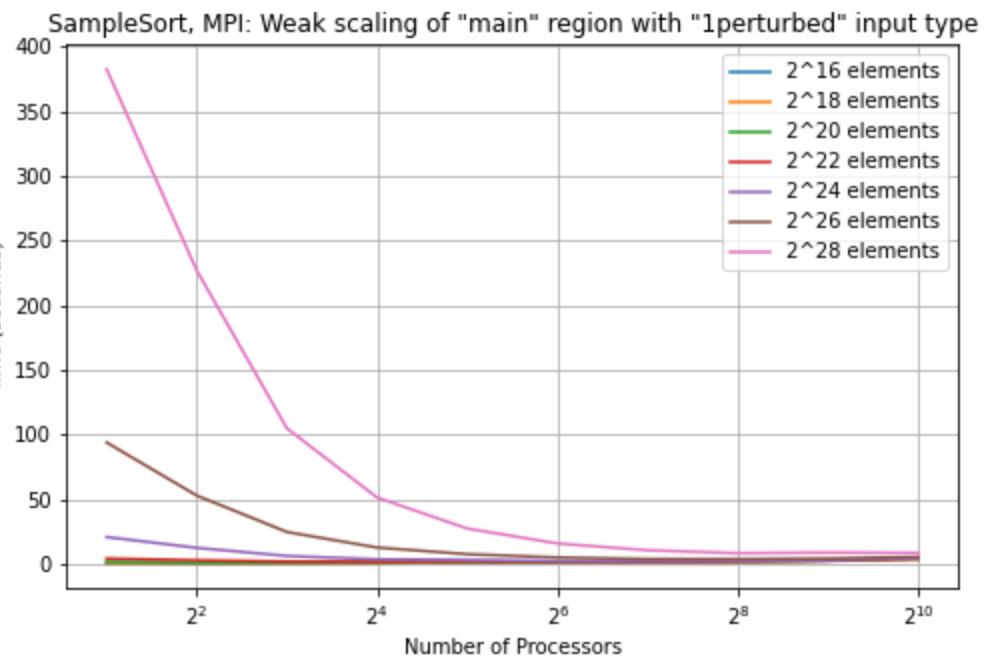


SampleSort, MPI: Weak scaling of "comp_large" region with "ReverseSorted" input type

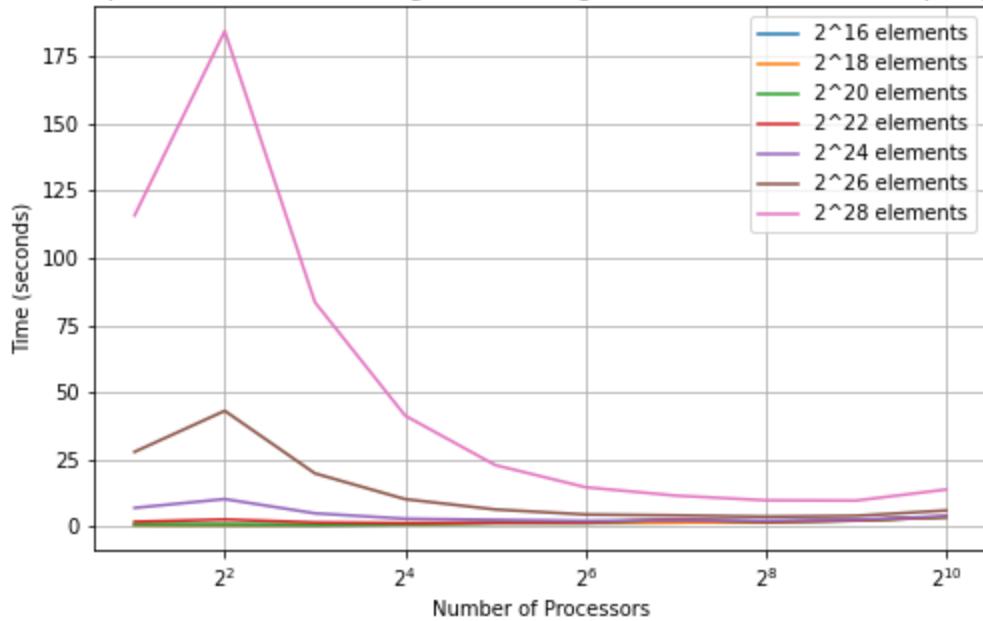


SampleSort, MPI: Weak scaling of "comp_large" region with "Sorted" input type

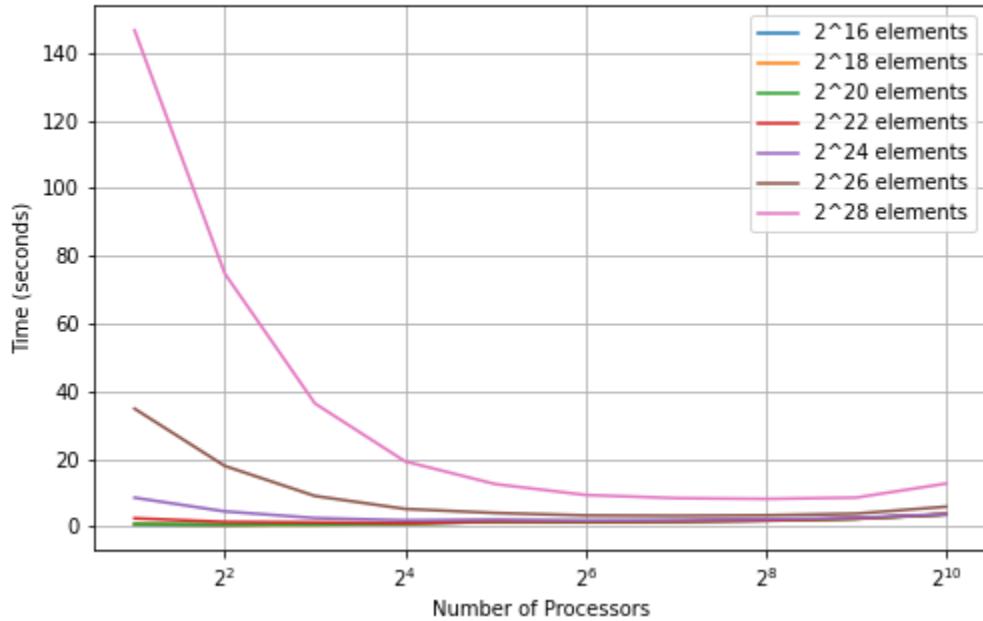




SampleSort, MPI: Weak scaling of "main" region with "ReverseSorted" input type



SampleSort, MPI: Weak scaling of "main" region with "Sorted" input type



Analysis:

One can see from the strong scaling graphs, that with an increasing number of processors, the execution time of the large computation decreased for all input types and all input sizes. The decrease in time was more pronounced with greater input sizes.

At the extreme, it took 300 seconds to finish large computation with two processors for Random and 1% Perturbed and it took less than one second to do the same with 1024 processors.

One can see from the strong scaling graphs, that with an increasing number of processors, the execution time of the whole program increased for input sizes of 2^{16} , 2^{18} , and 2^{20} . There was an initial decrease and then increase (the saddle point being 2^6 processors) for input size of 2^{22} . There was a decrease for 2^{24} , 2^{26} , and 2^{28} elements. This makes sense as the communication overhead for higher processors is expensive and outweighs the benefits of parallel computation for smaller input sizes.

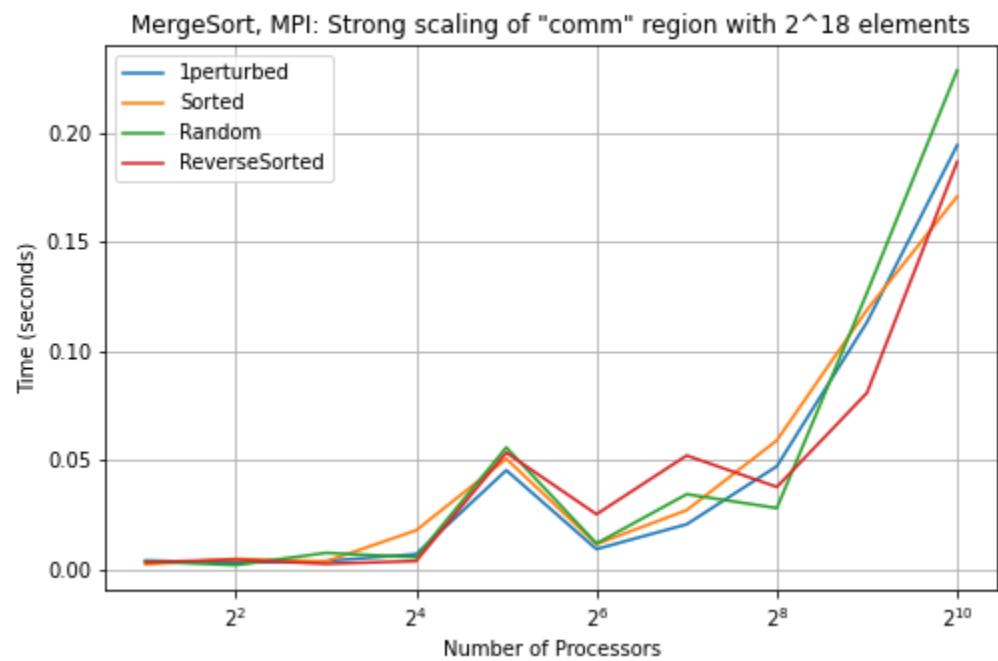
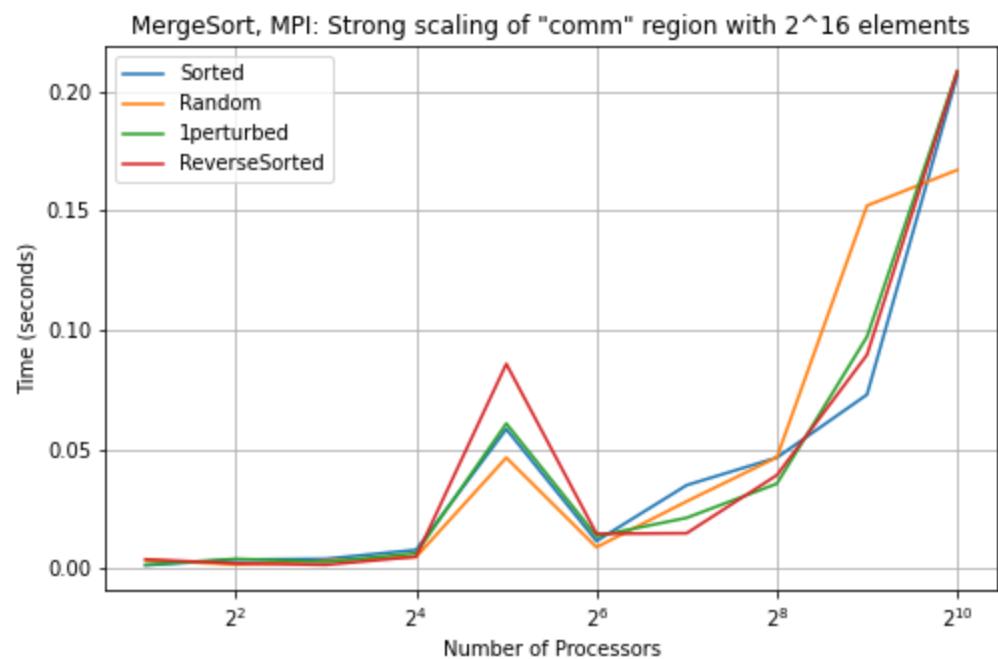
One can see from the strong scaling graphs, that with an increasing number of processors, the execution time of the communication increased for input sizes of 2^{16} , 2^{18} , and 2^{20} . There was an initial decrease and then increase (the saddle point being 2^6 processors) for input size of 2^{22} . There was a decrease for 2^{24} , 2^{26} , and 2^{28} elements. This makes sense as the communication overhead for higher processors is expensive and outweighs the benefits of parallel computation for smaller input sizes.

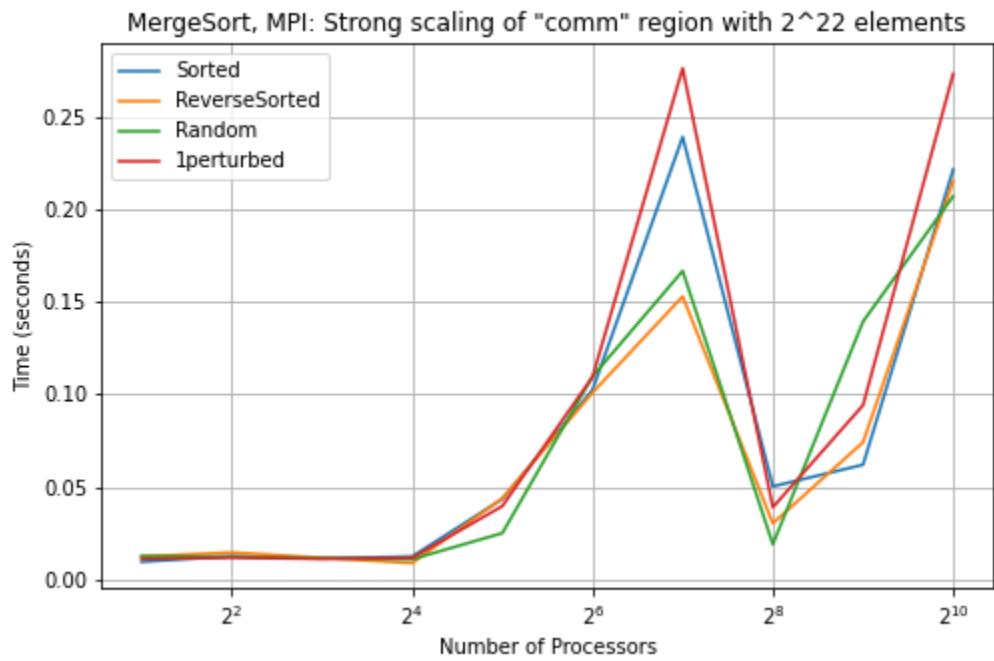
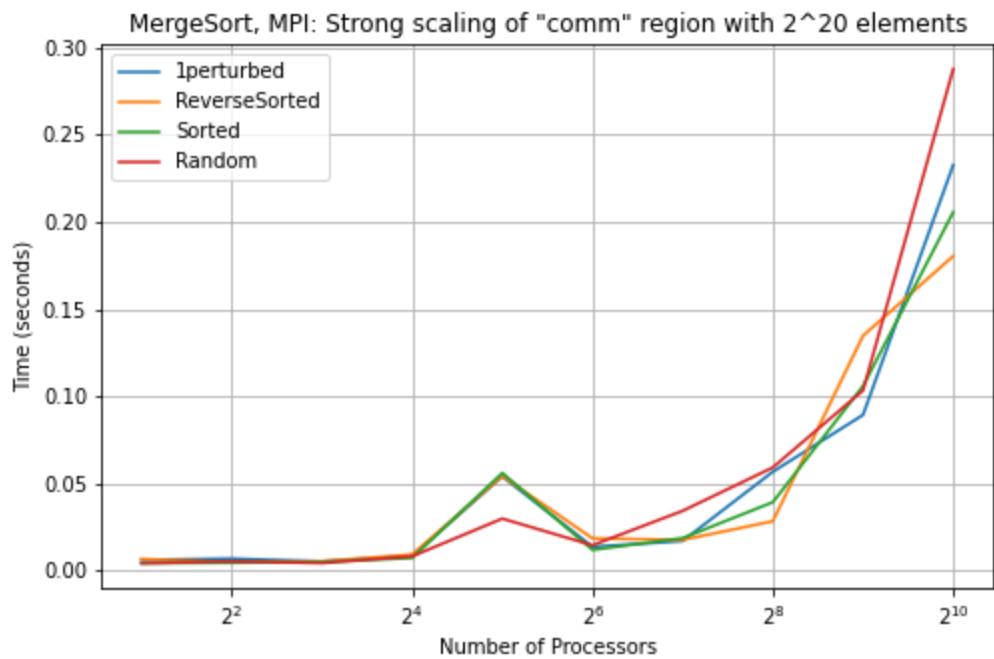
One can see from the weak scaling graphs, that with an increasing number of processors, the execution time of both the large computation and whole program decreased for all input types and all input sizes. The decrease was more pronounced with more greater input sizes.

One can see from the speed up graphs, that with an increasing number of processors, the execution time of both the large computation and whole program increased for all input types but only large input sizes like 2^{28} and 2^{26} . We can conclude that the speed up benefits only occurred for input sizes where the benefits of parallel computation outweighed the communication overhead in a higher number of processors. For the whole program speed-up, depending on the type of input type, I was achieving up to 40x speed up. Considering the ideal speed up is 512, I am content with this speed up. For large computational speed up, while I did achieve speed-up times of up to 1000, the explanation for this occurrence is because our baseline time was 300 seconds (with two processors) and the time with 1024 processors was less than a second. I expect this to be because Sample Sort works extremely ineffectively with a small sample size and since my sample size was always one less than the number of processors, my baseline time used one sample. Having an extremely large time for my baseline would create unusually high speed up times.

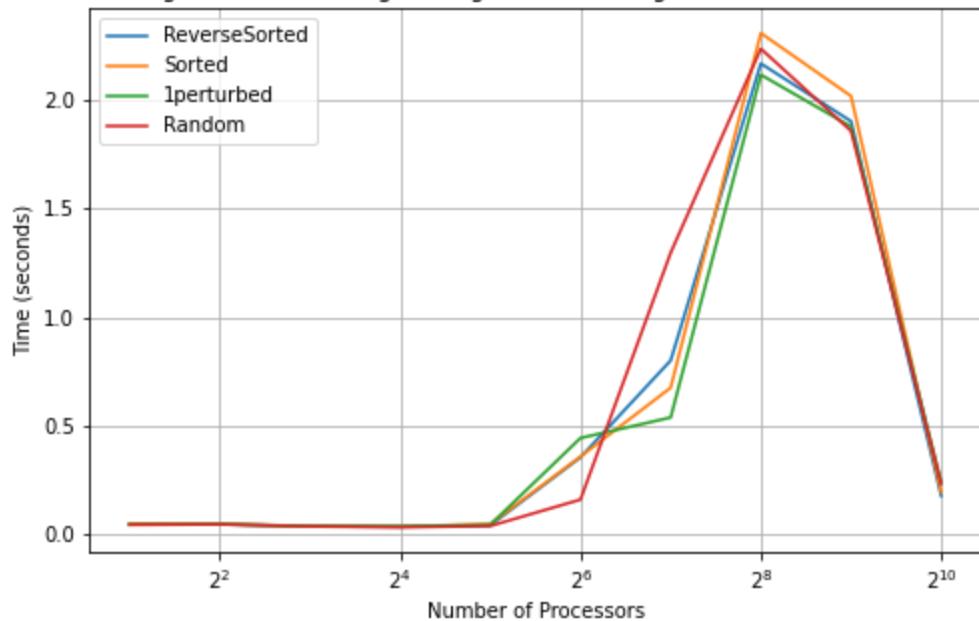
Merge Sort MPI

Graphs:

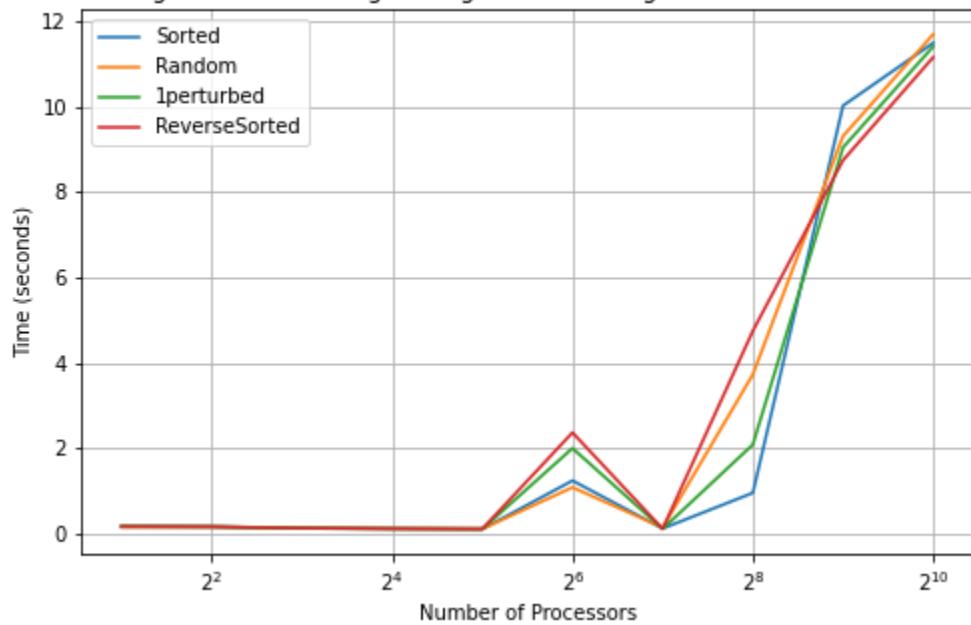




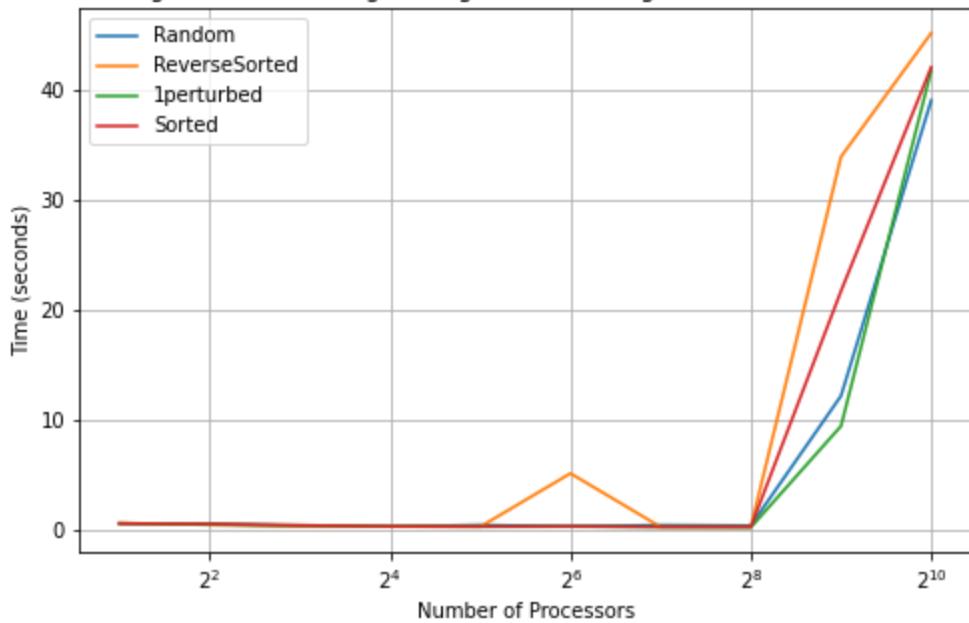
MergeSort, MPI: Strong scaling of "comm" region with 2^{24} elements



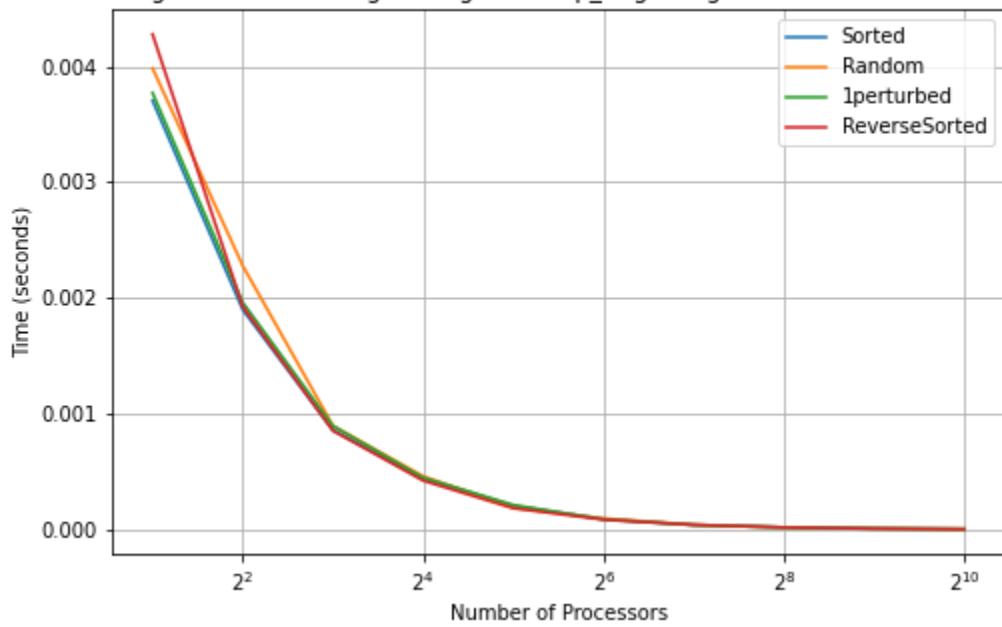
MergeSort, MPI: Strong scaling of "comm" region with 2^{26} elements

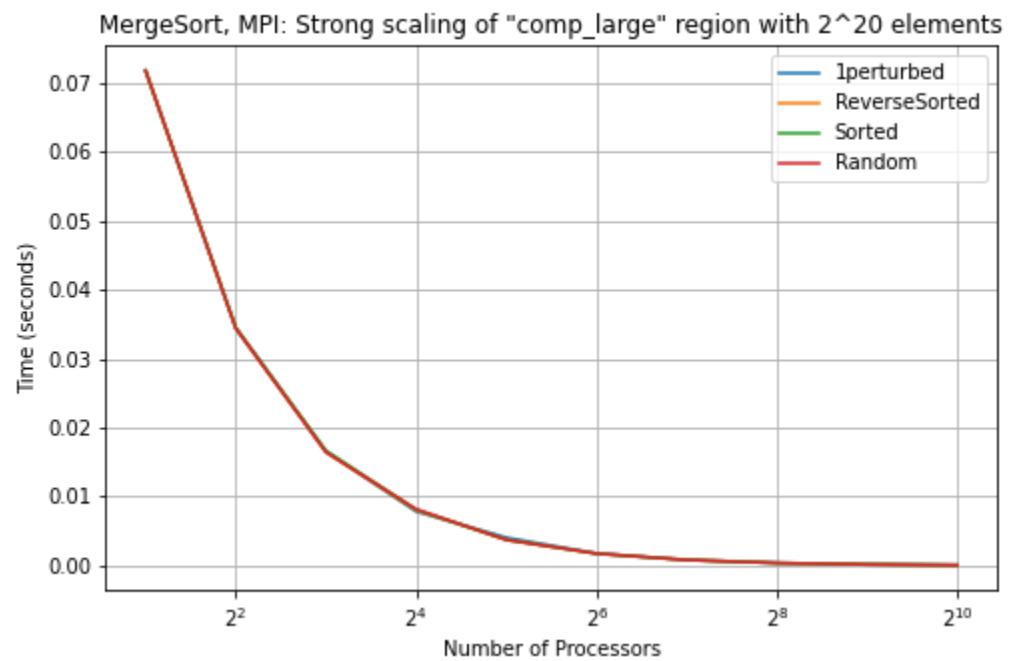
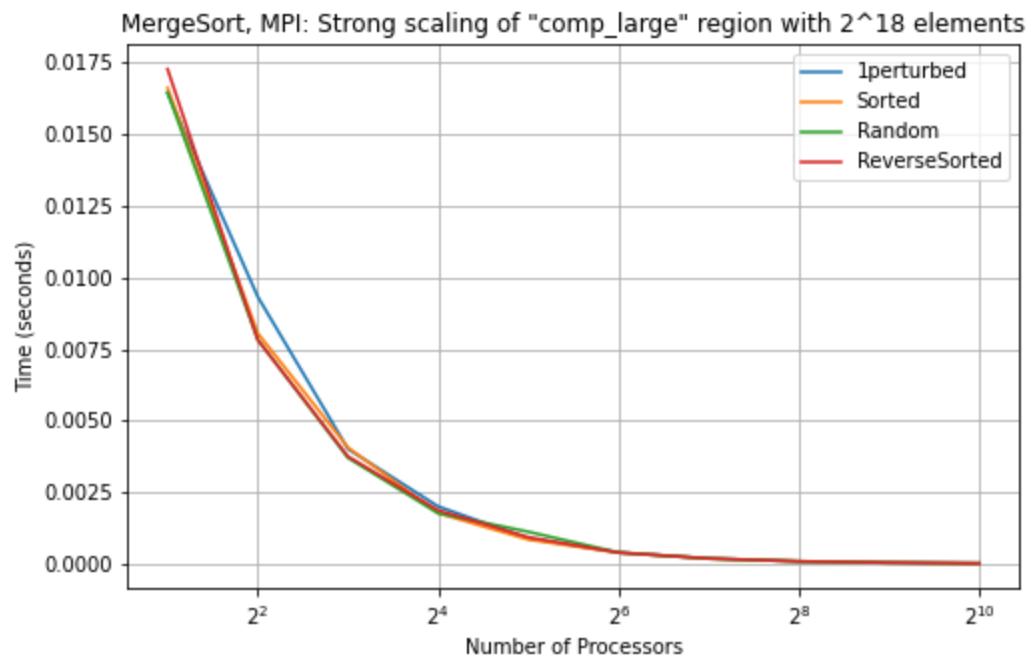


MergeSort, MPI: Strong scaling of "comm" region with 2^{28} elements

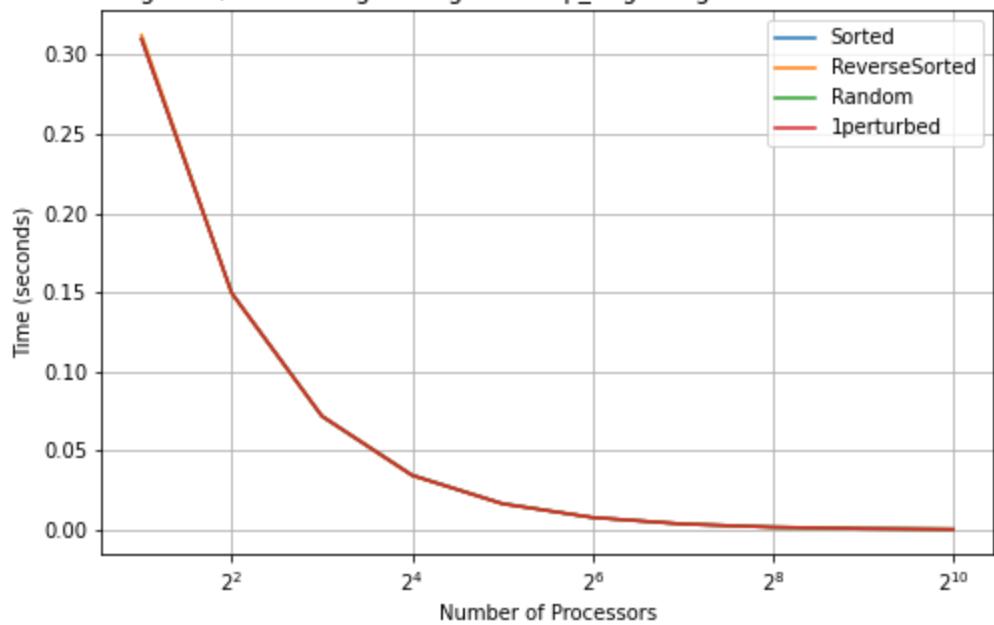


MergeSort, MPI: Strong scaling of "comp_large" region with 2^{16} elements

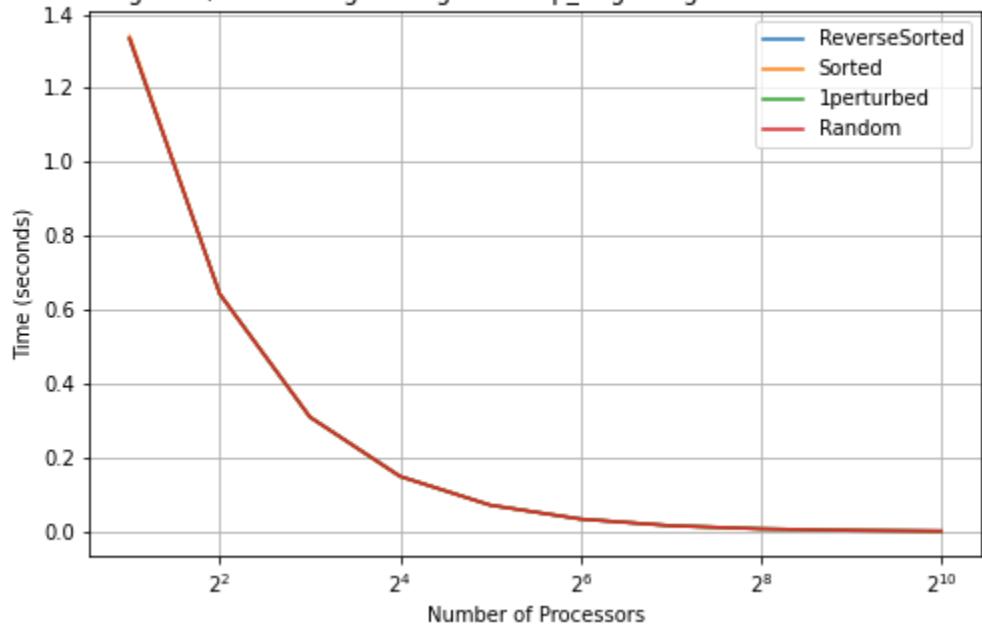


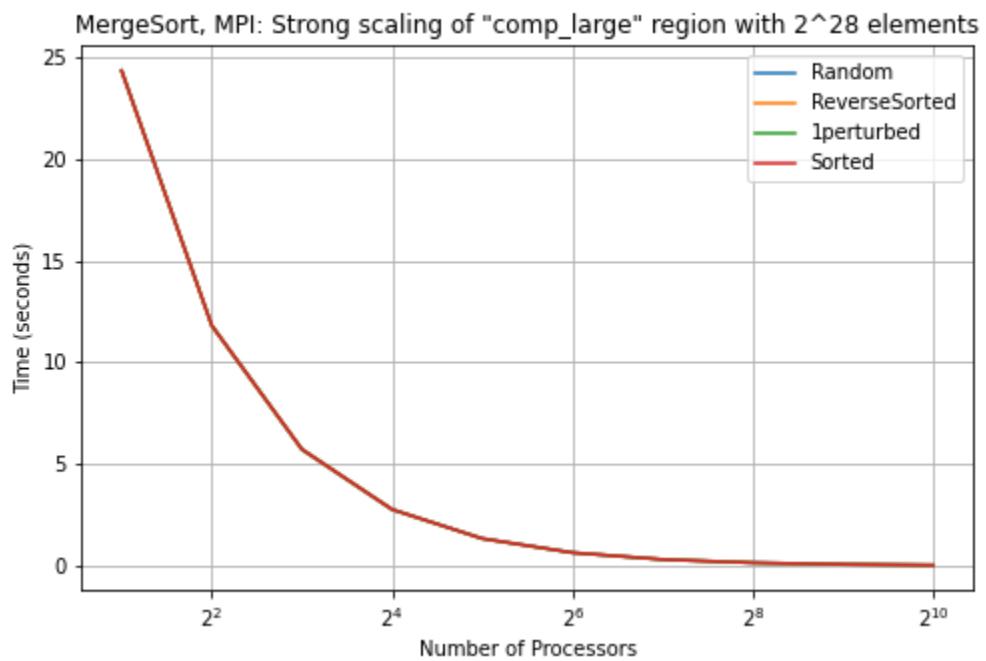
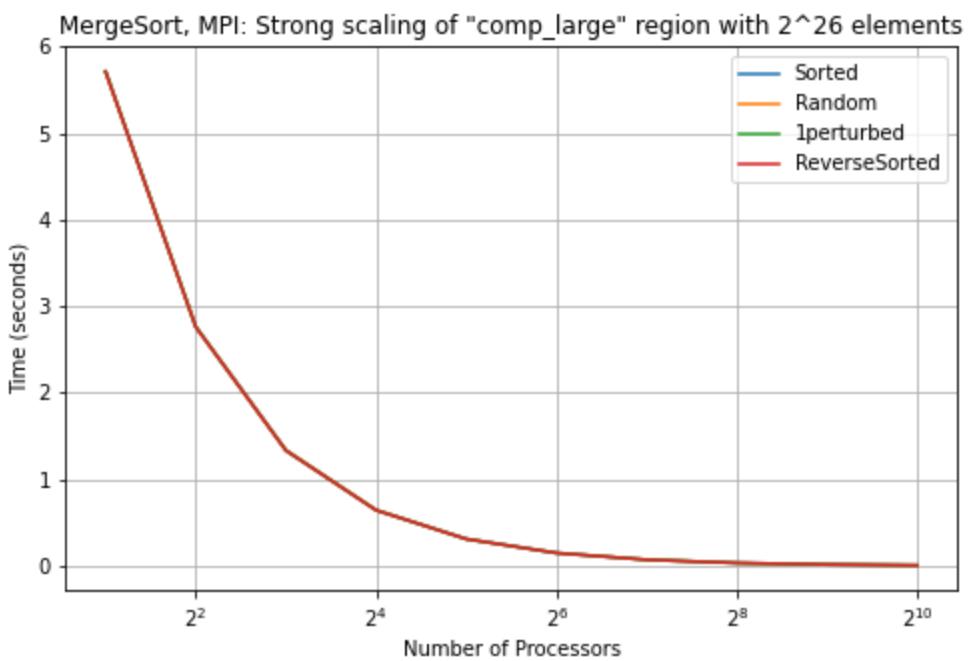


MergeSort, MPI: Strong scaling of "comp_large" region with 2^{22} elements

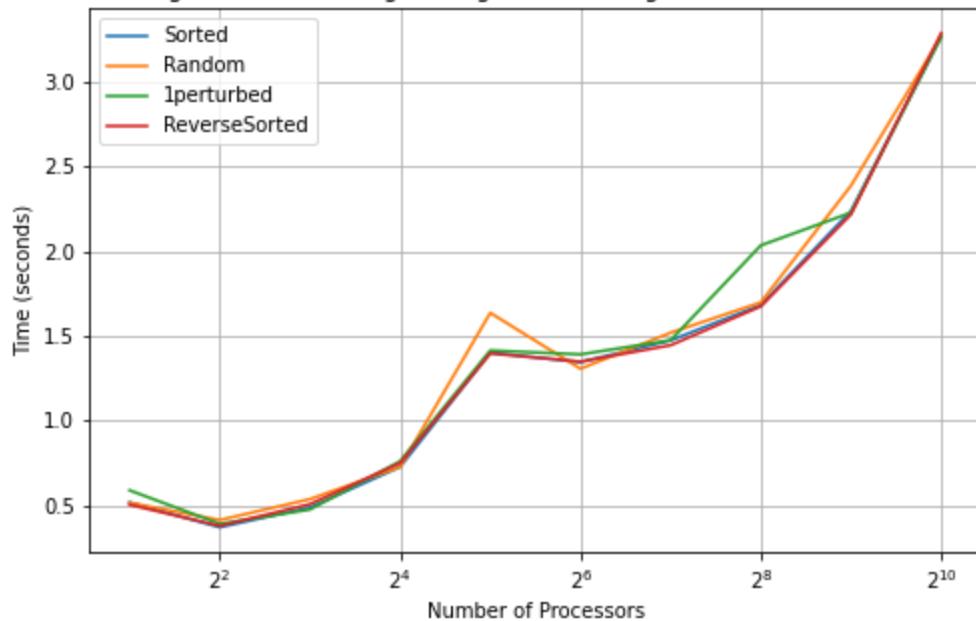


MergeSort, MPI: Strong scaling of "comp_large" region with 2^{24} elements

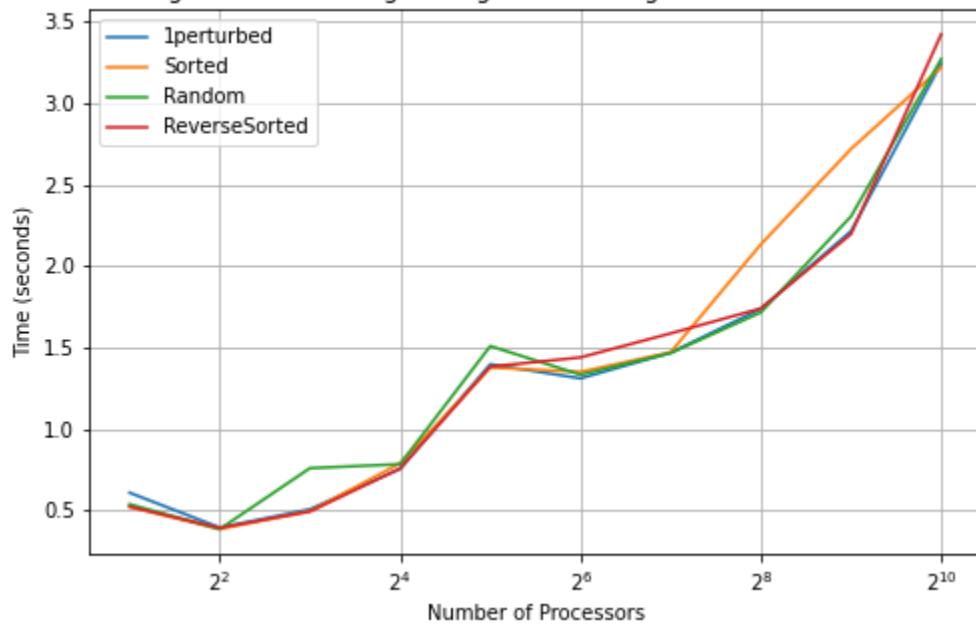


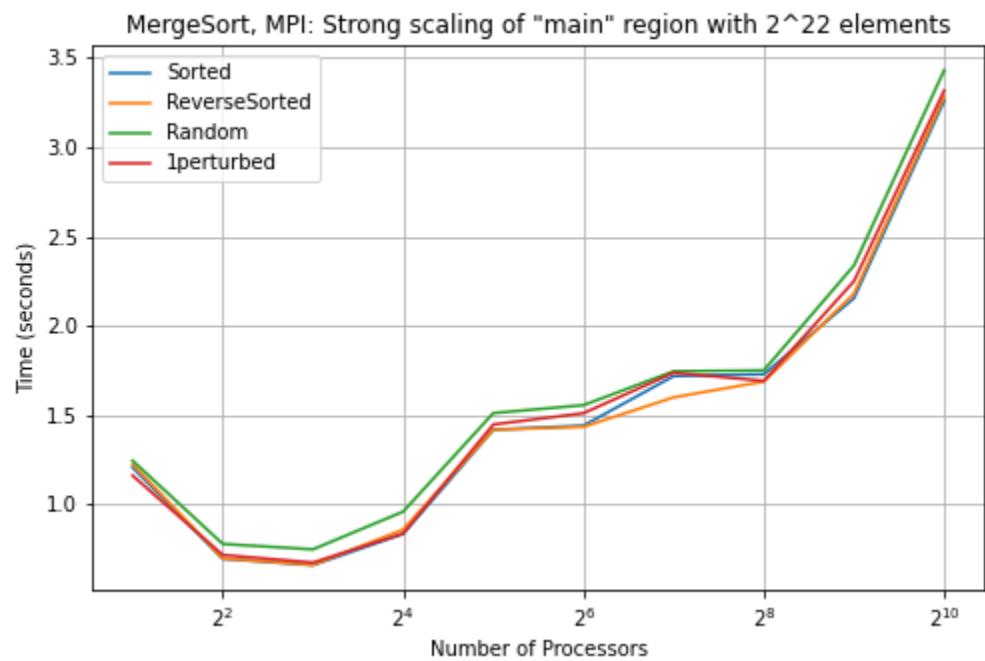
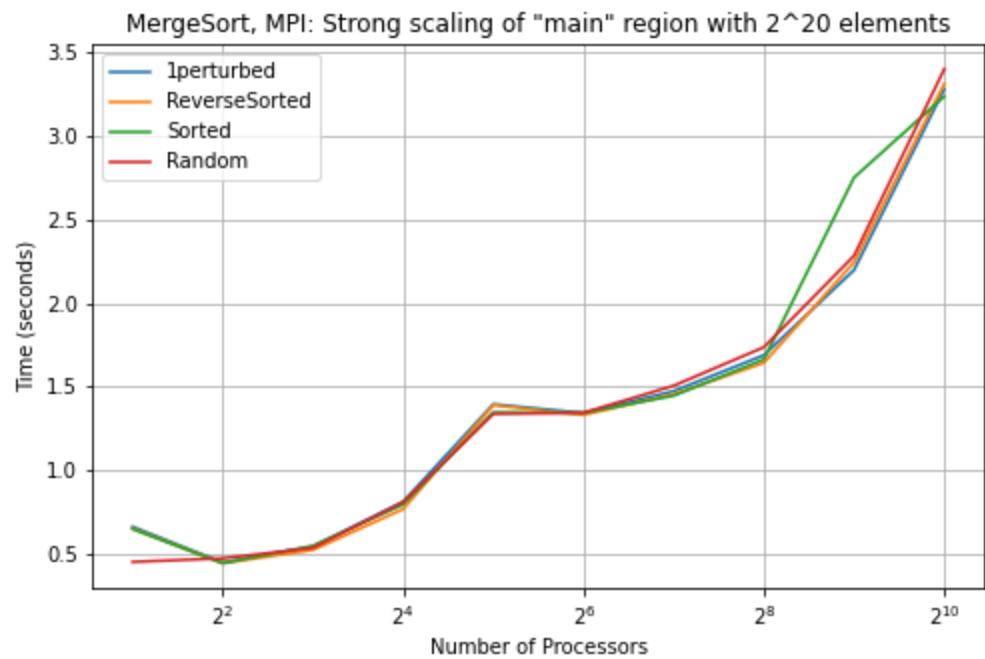


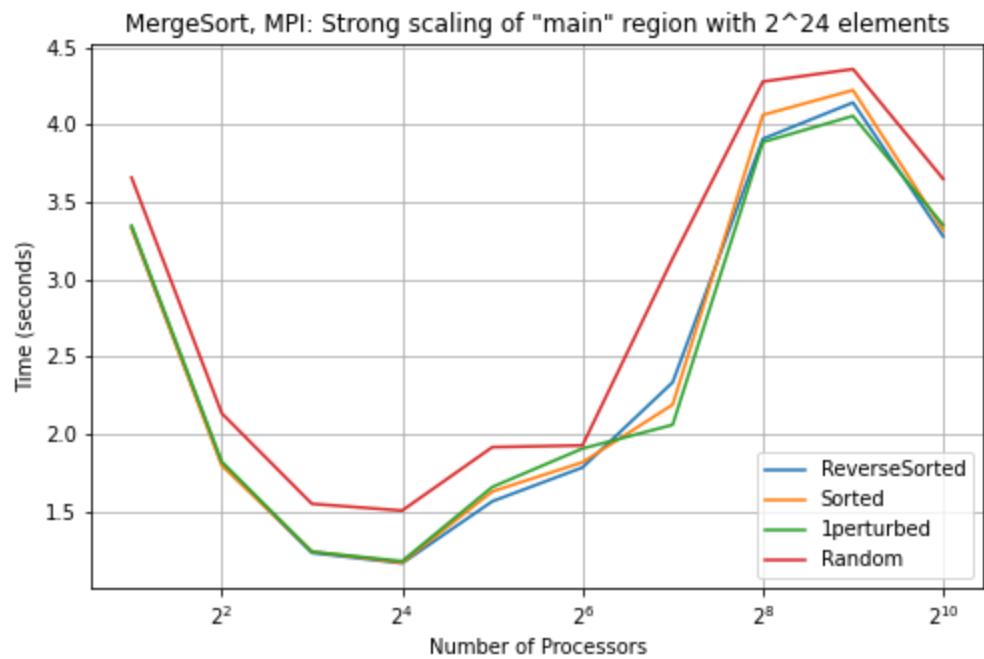
MergeSort, MPI: Strong scaling of "main" region with 2^{16} elements



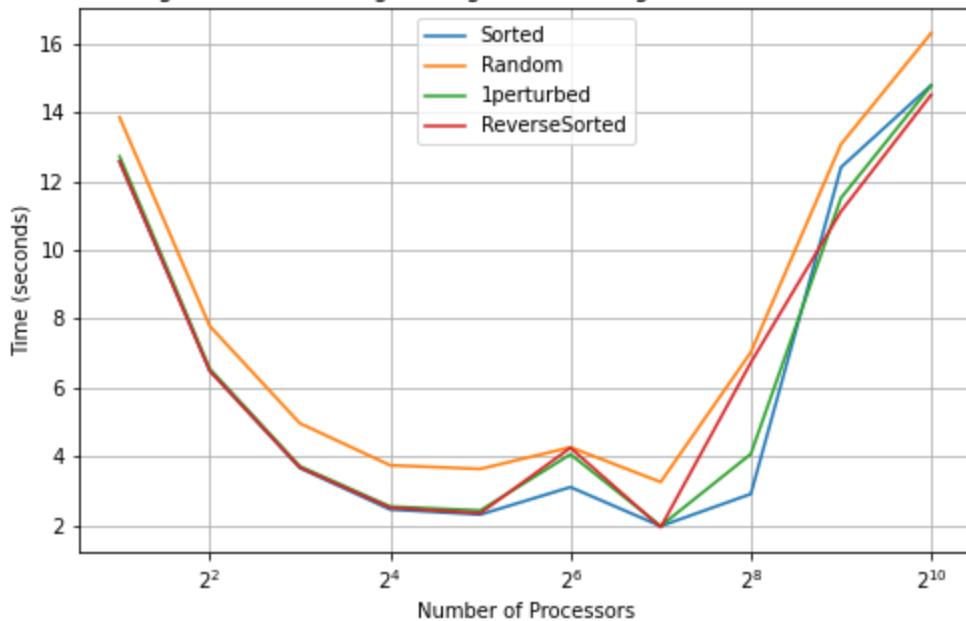
MergeSort, MPI: Strong scaling of "main" region with 2^{18} elements



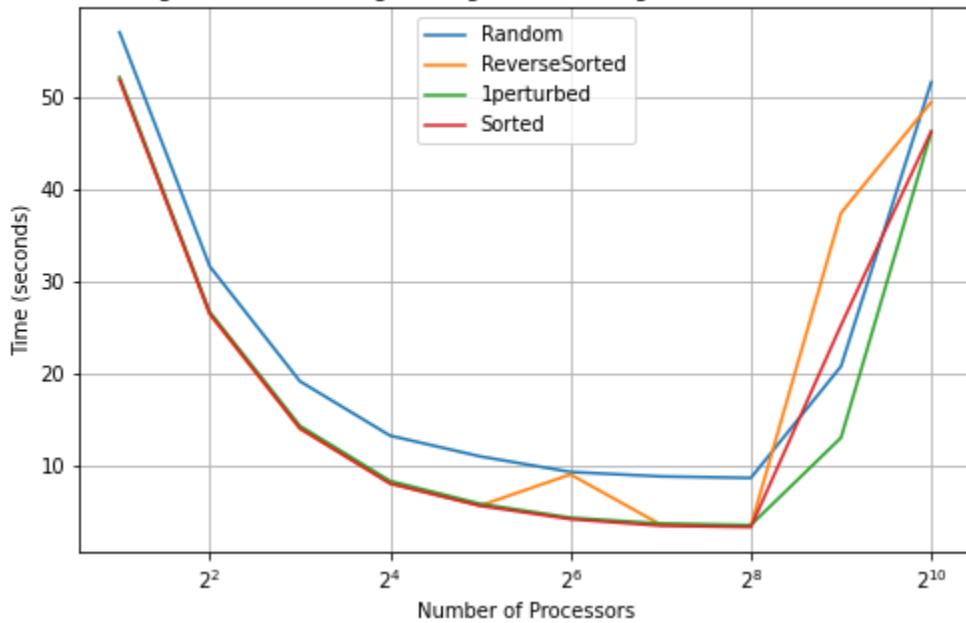




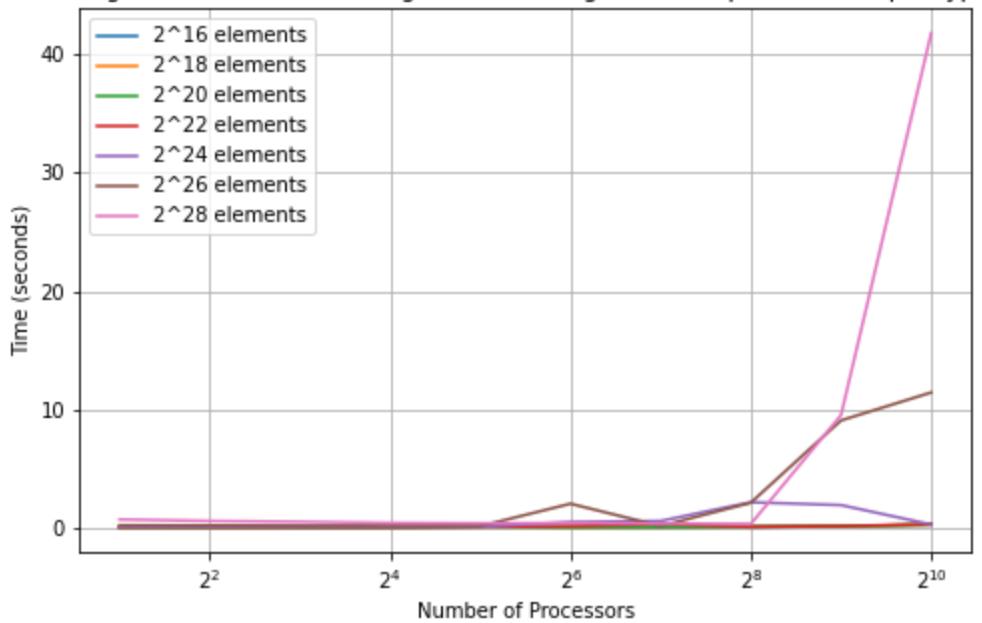
MergeSort, MPI: Strong scaling of "main" region with 2^{26} elements



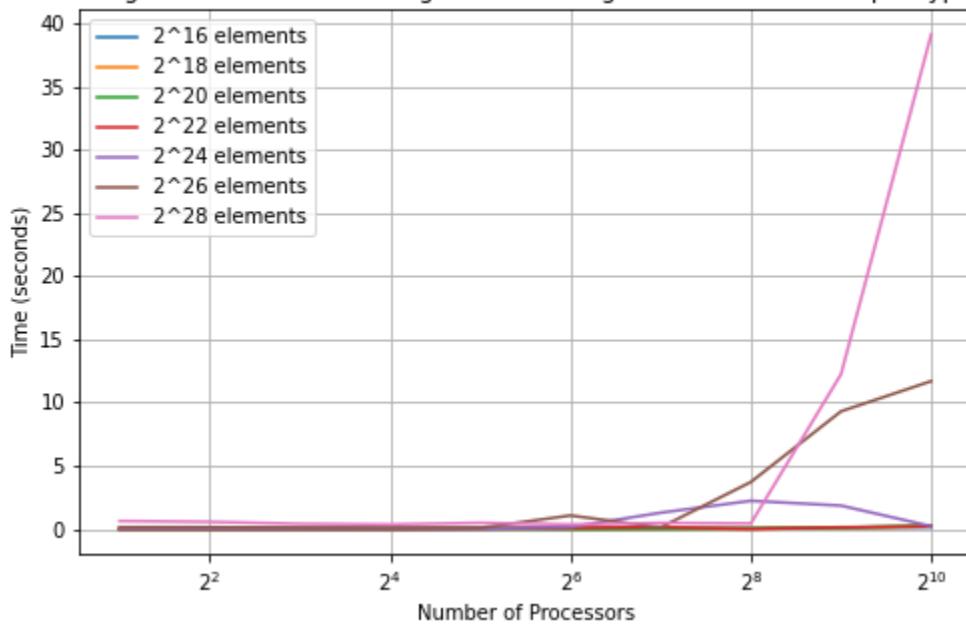
MergeSort, MPI: Strong scaling of "main" region with 2^{28} elements



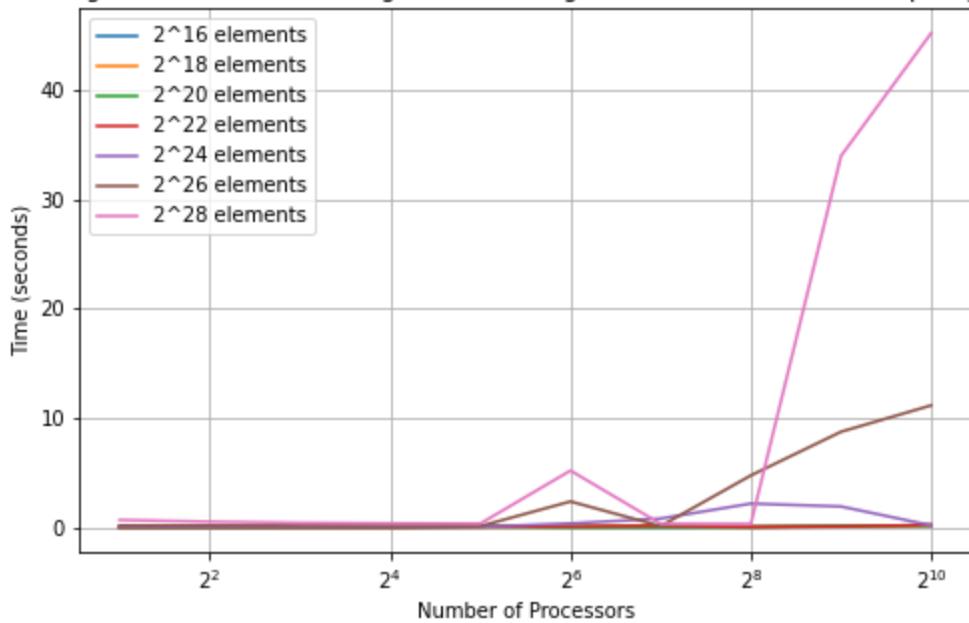
MergeSort, MPI: Weak scaling of "comm" region with "1perturbed" input type



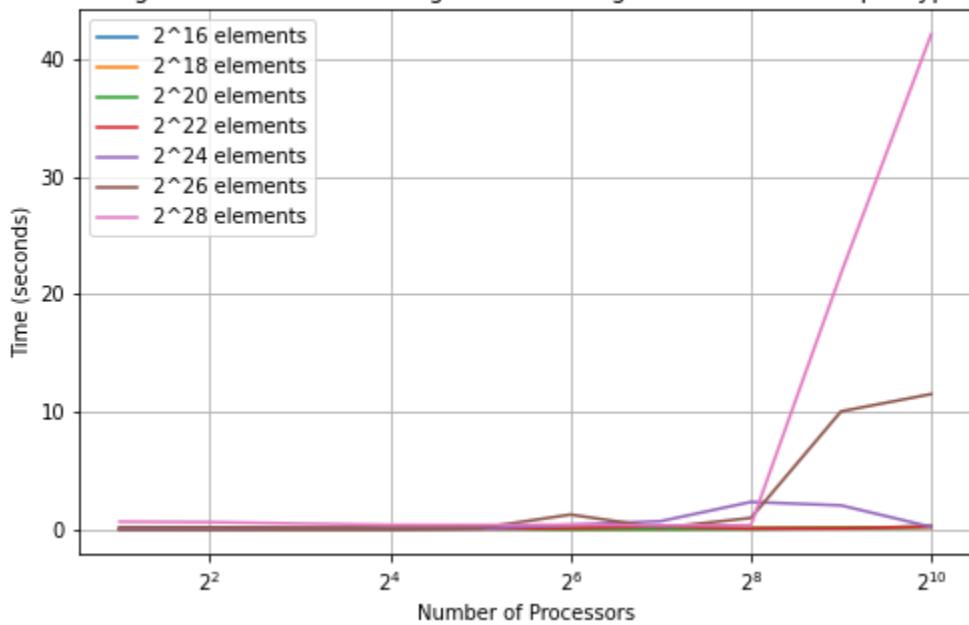
MergeSort, MPI: Weak scaling of "comm" region with "Random" input type



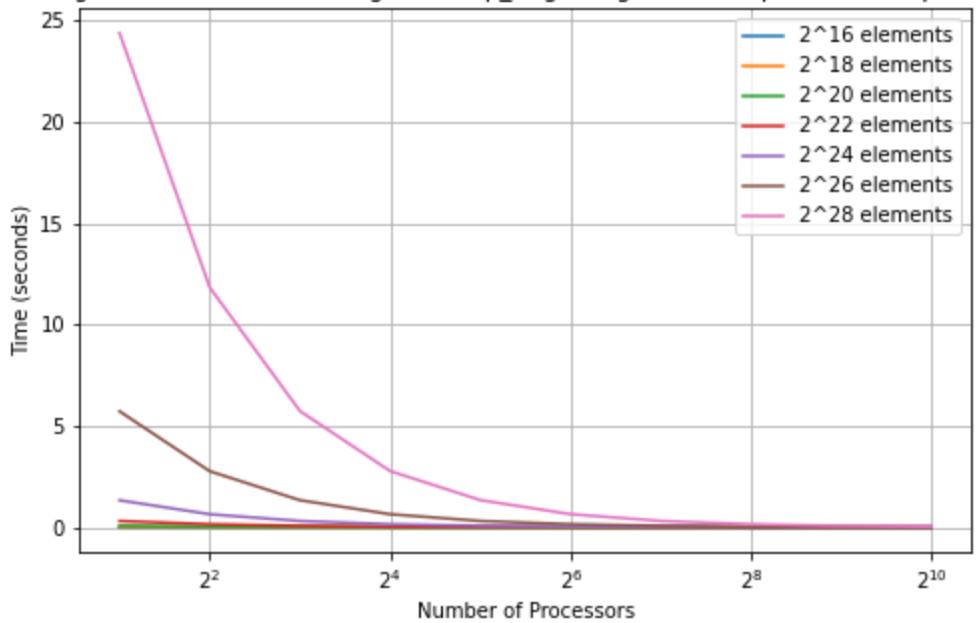
MergeSort, MPI: Weak scaling of "comm" region with "ReverseSorted" input type



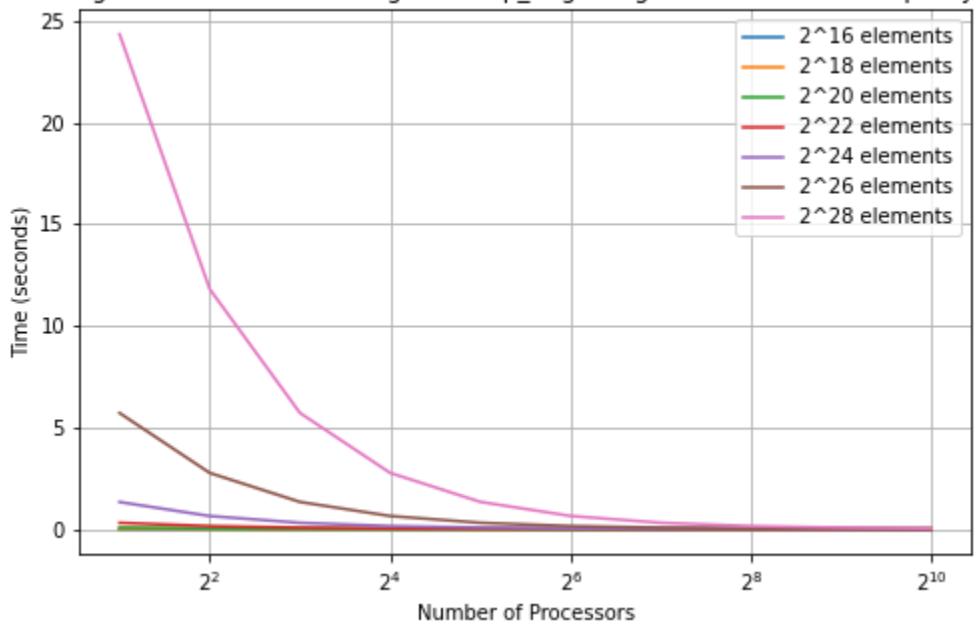
MergeSort, MPI: Weak scaling of "comm" region with "Sorted" input type



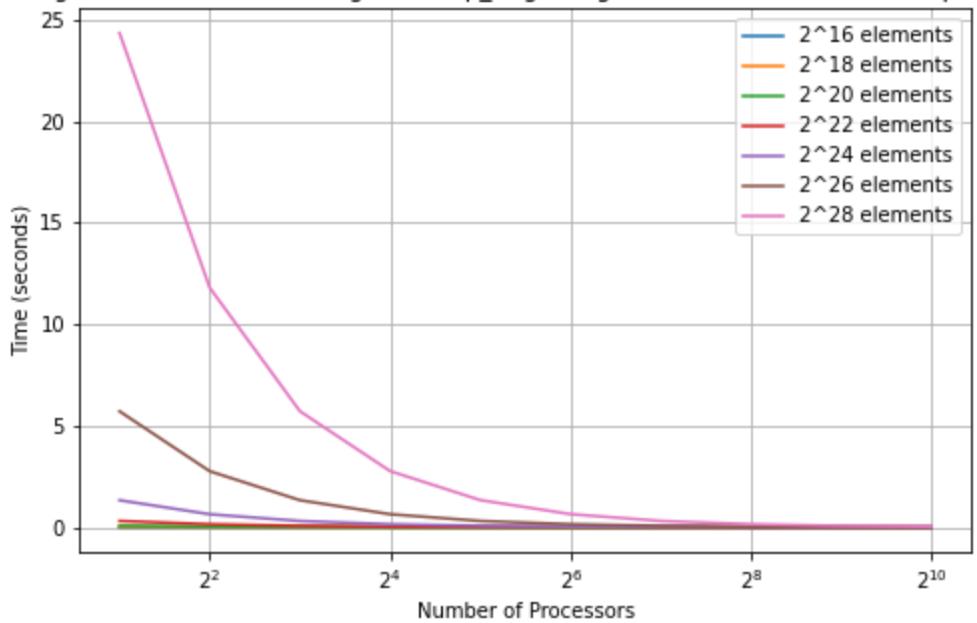
MergeSort, MPI: Weak scaling of "comp_large" region with "1perturbed" input type



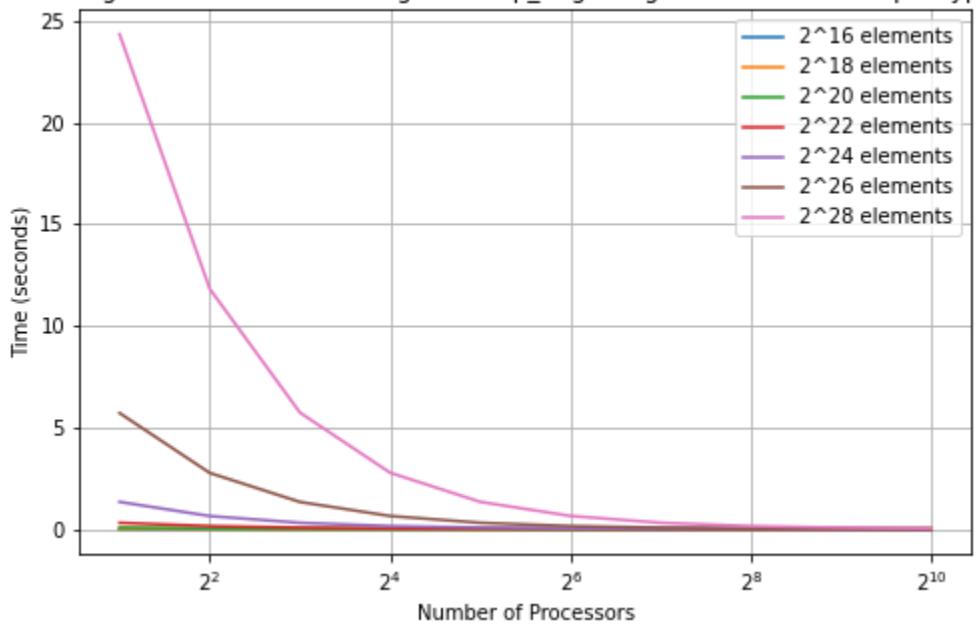
MergeSort, MPI: Weak scaling of "comp_large" region with "Random" input type



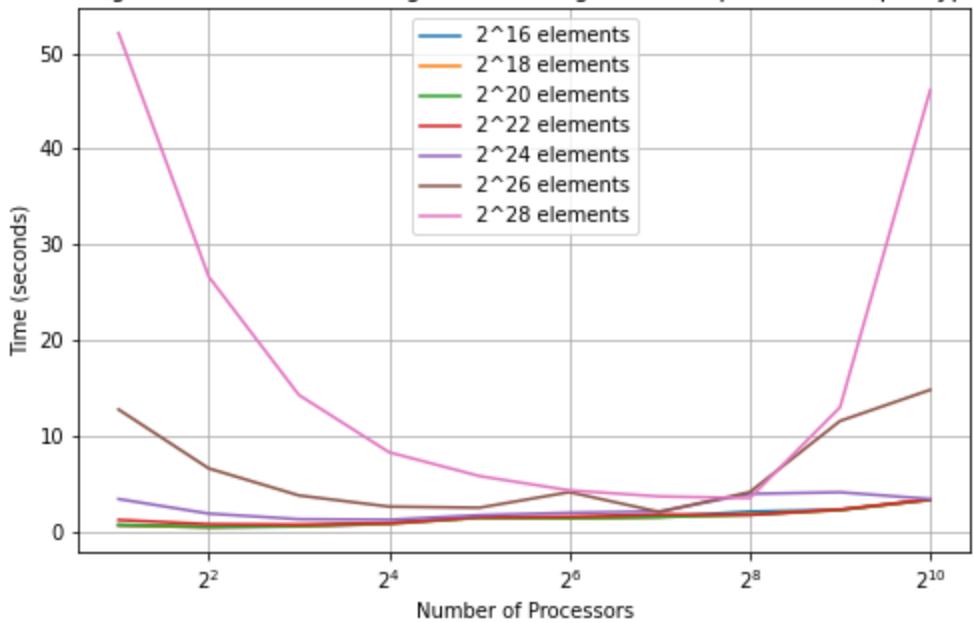
MergeSort, MPI: Weak scaling of "comp_large" region with "ReverseSorted" input type



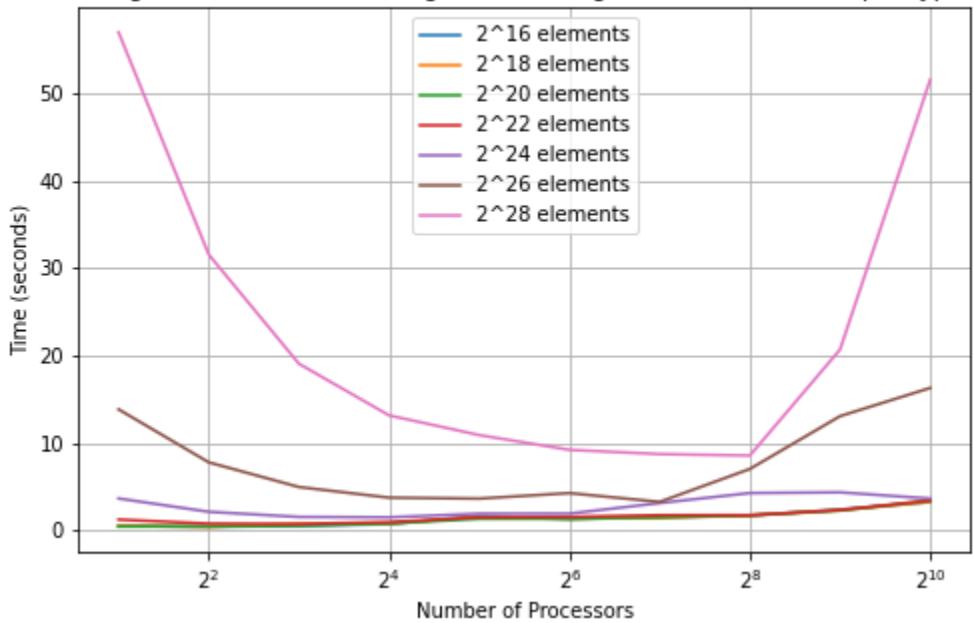
MergeSort, MPI: Weak scaling of "comp_large" region with "Sorted" input type



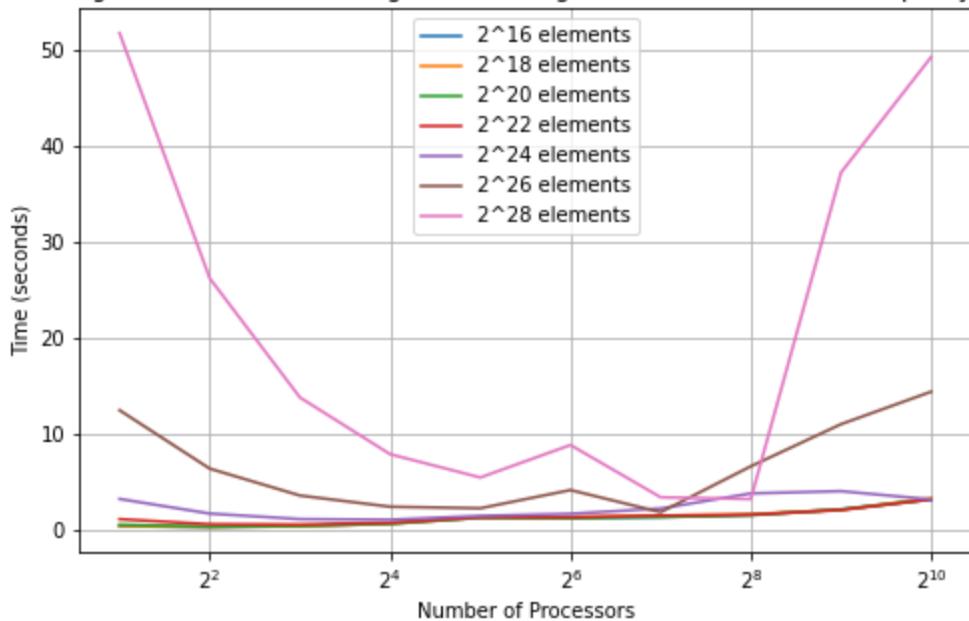
MergeSort, MPI: Weak scaling of "main" region with "1perturbed" input type



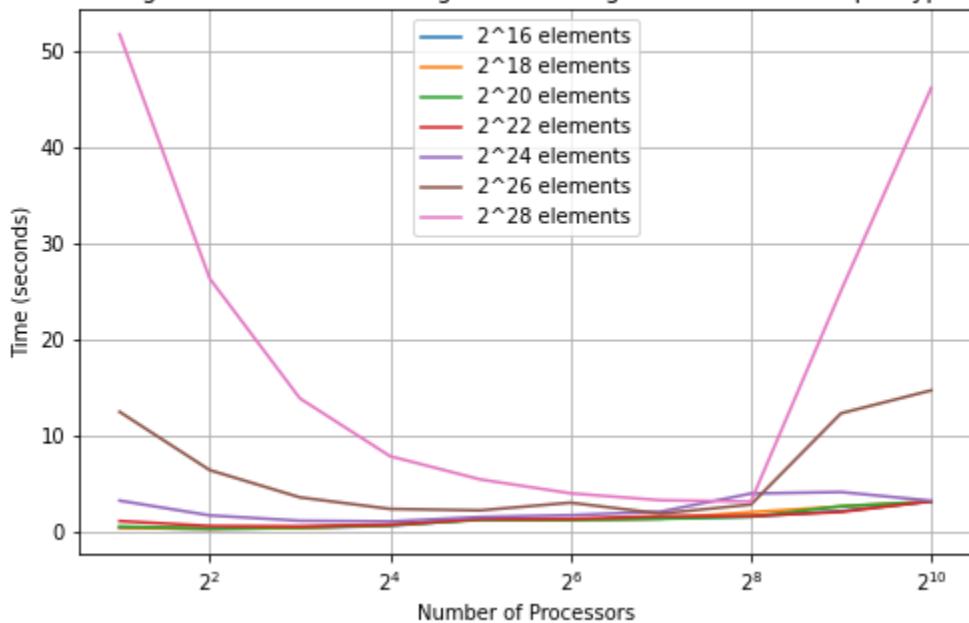
MergeSort, MPI: Weak scaling of "main" region with "Random" input type



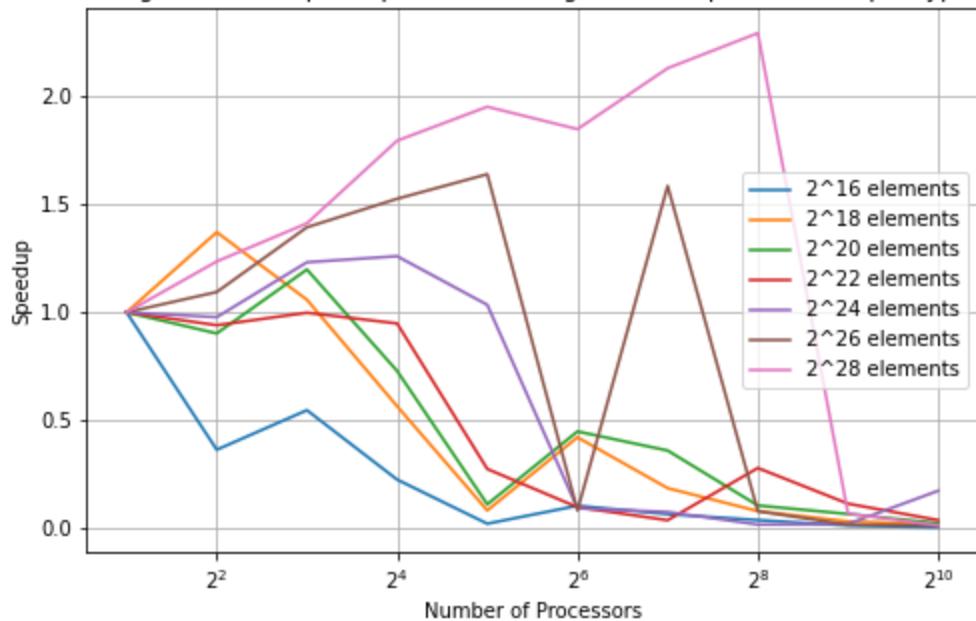
MergeSort, MPI: Weak scaling of "main" region with "ReverseSorted" input type



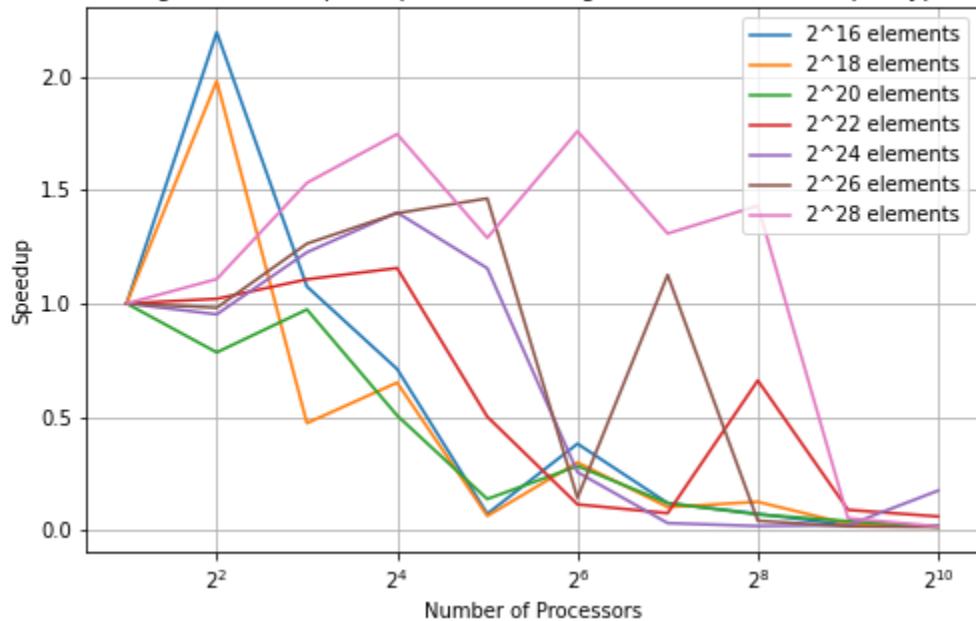
MergeSort, MPI: Weak scaling of "main" region with "Sorted" input type



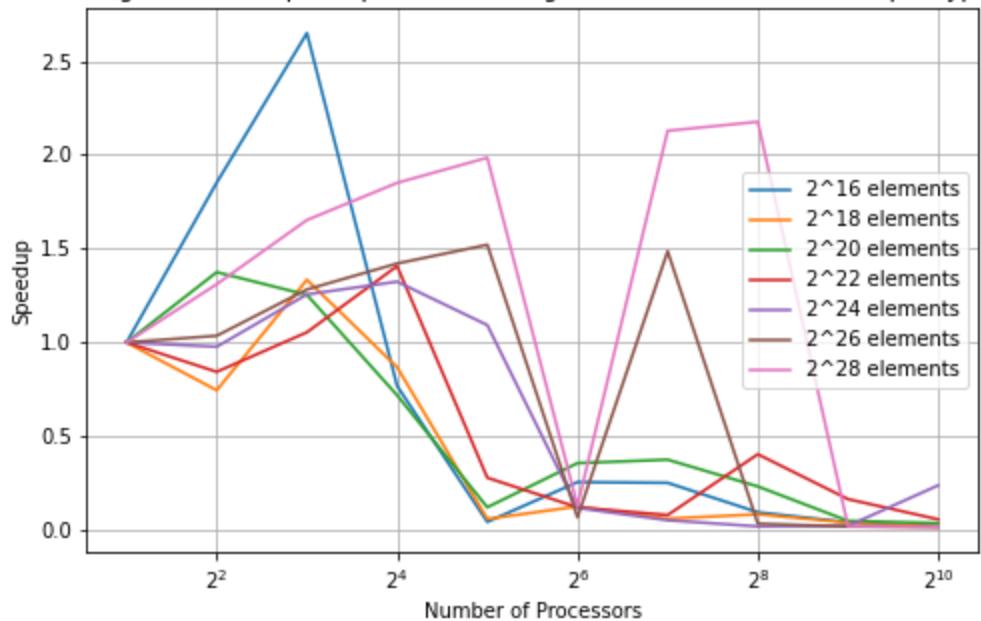
MergeSort, MPI: Speedup of "comm" region with "1perturbed" input type



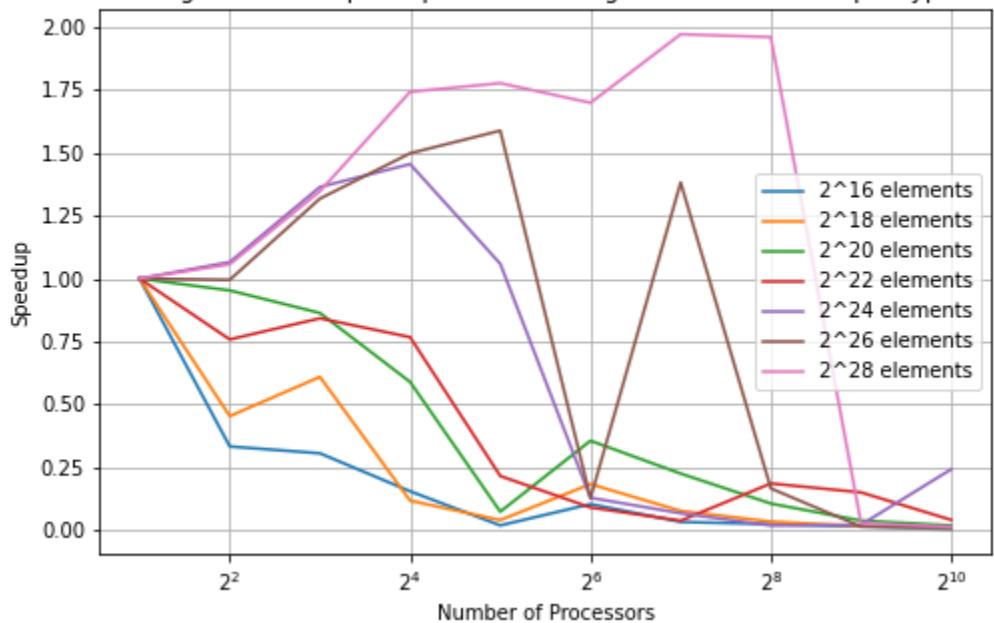
MergeSort, MPI: Speedup of "comm" region with "Random" input type

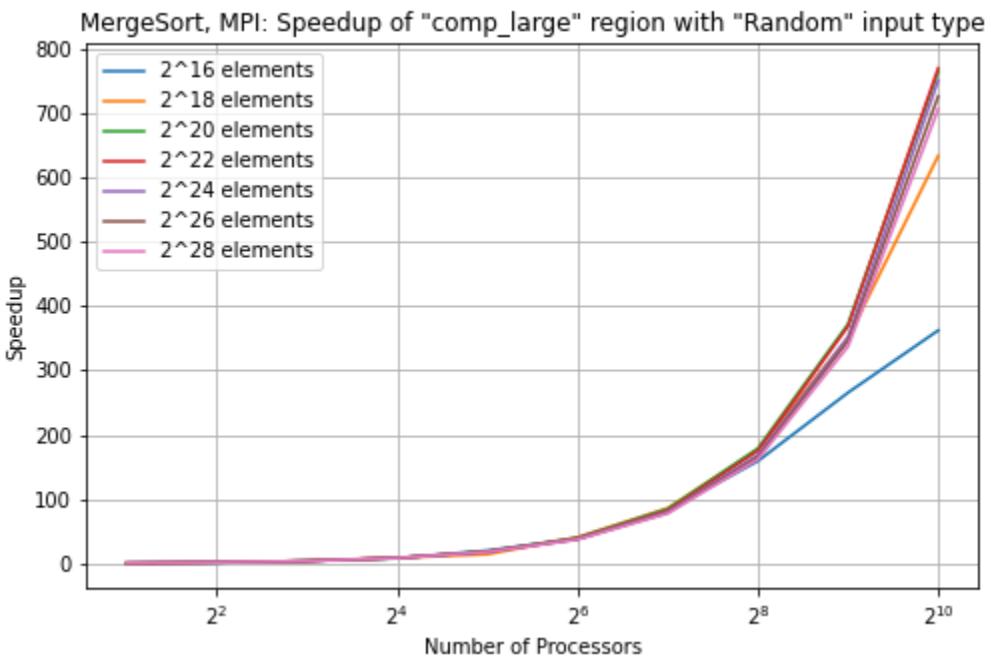
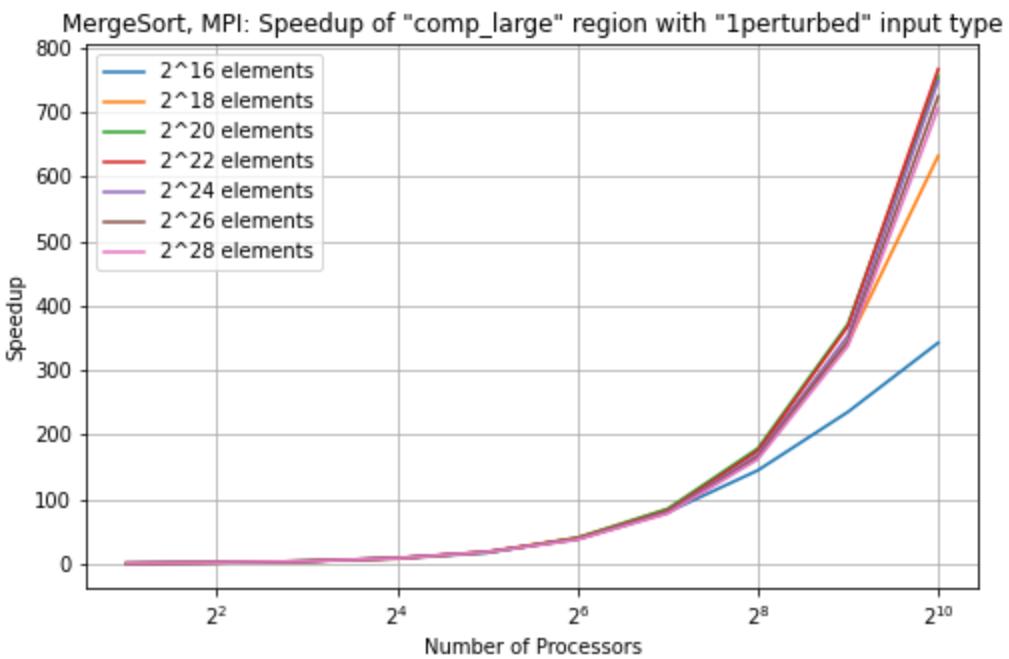


MergeSort, MPI: Speedup of "comm" region with "ReverseSorted" input type

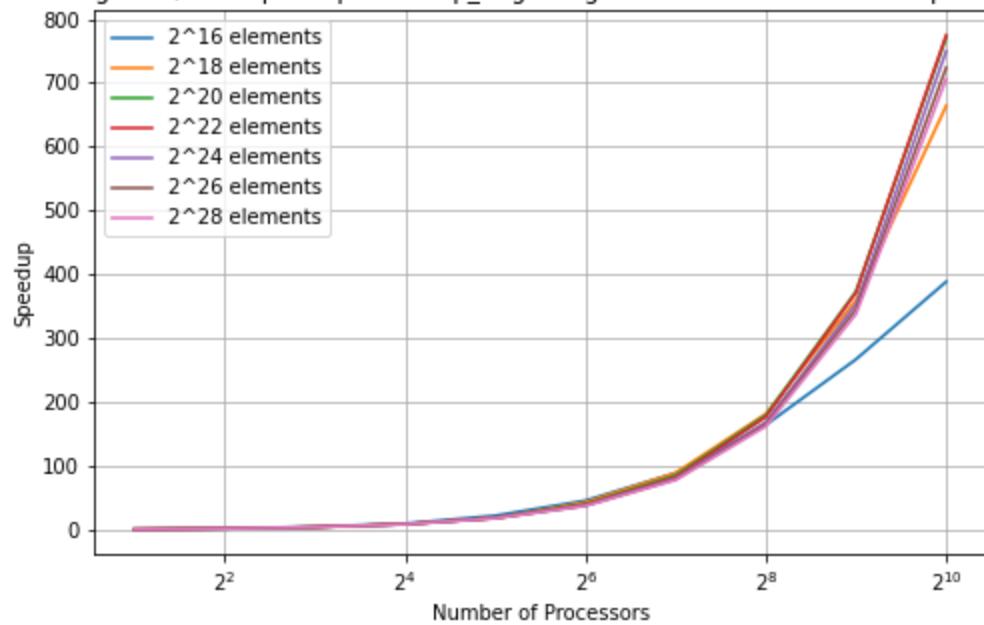


MergeSort, MPI: Speedup of "comm" region with "Sorted" input type

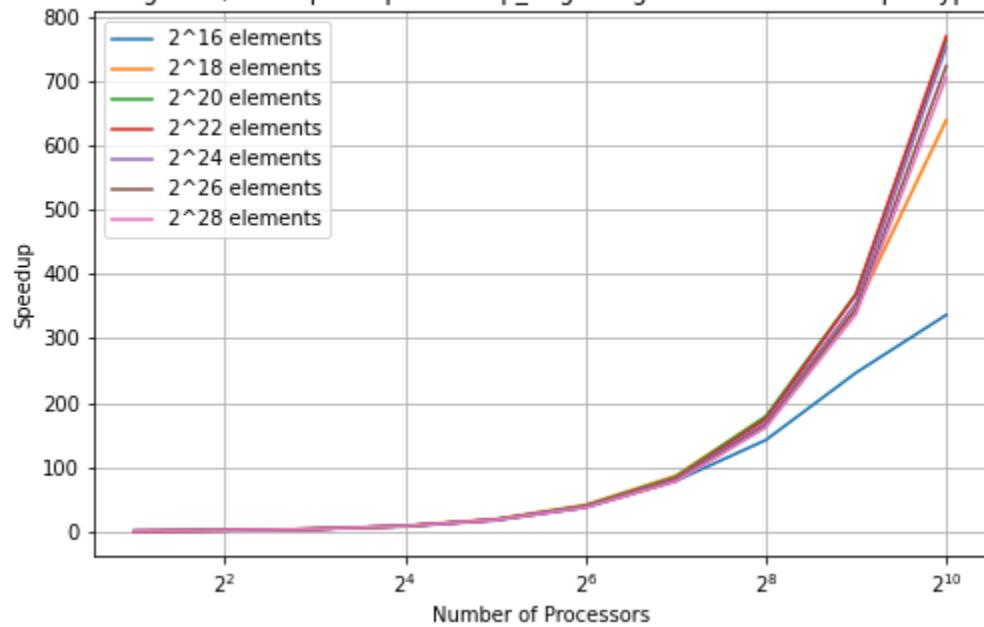




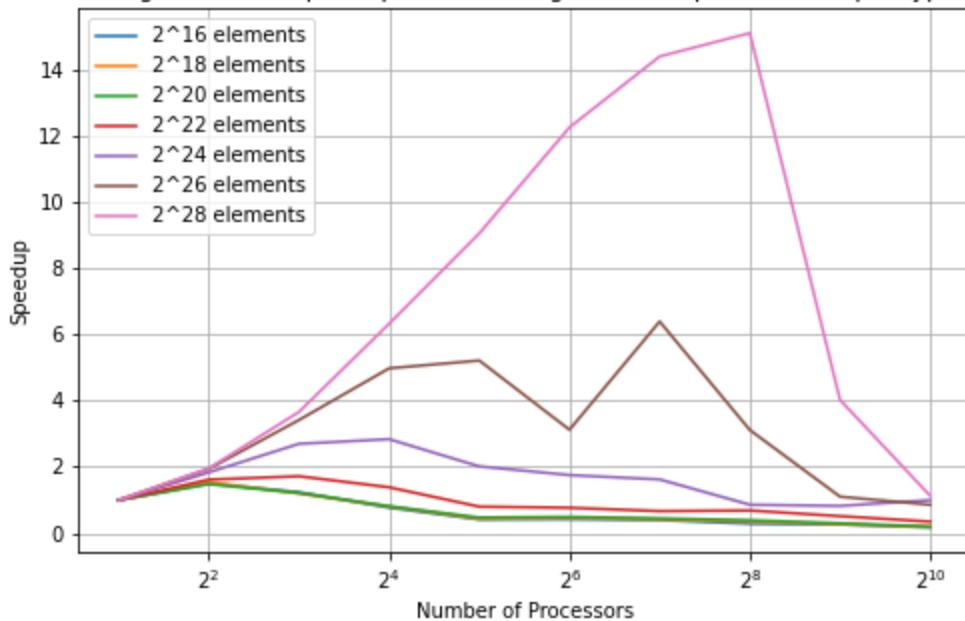
MergeSort, MPI: Speedup of "comp_large" region with "ReverseSorted" input type



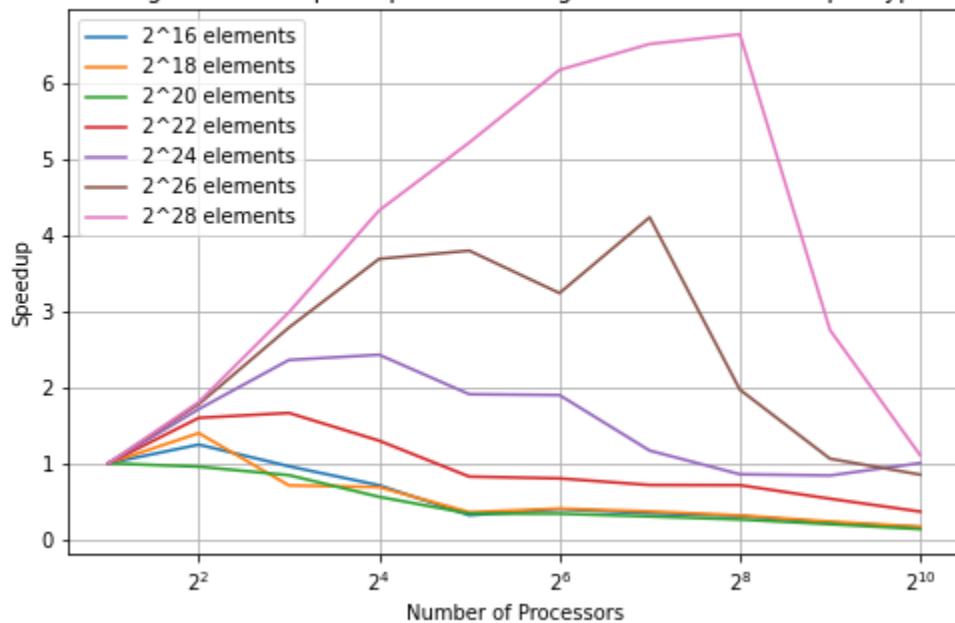
MergeSort, MPI: Speedup of "comp_large" region with "Sorted" input type

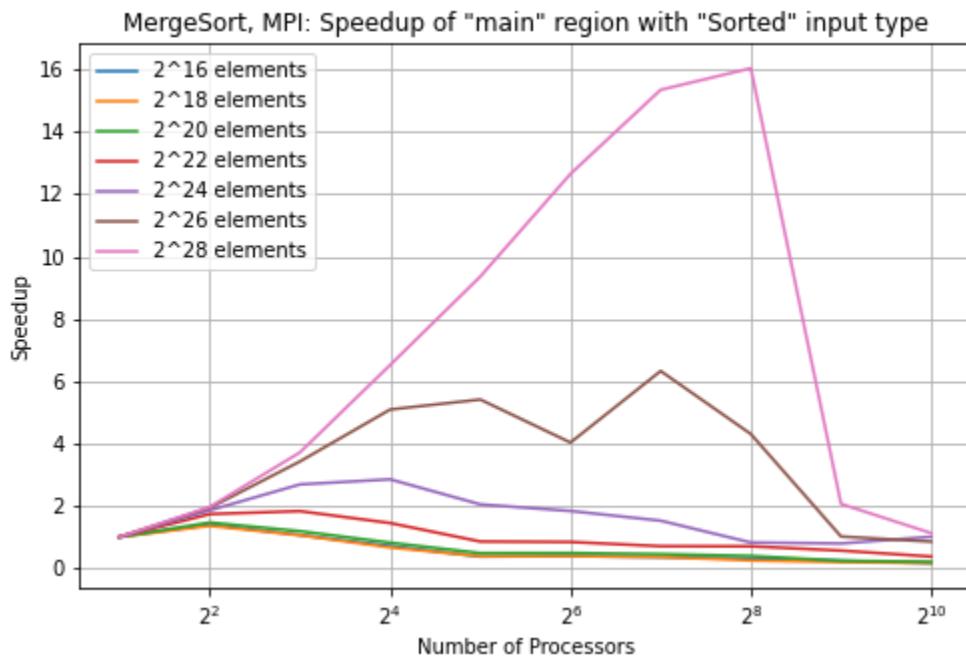
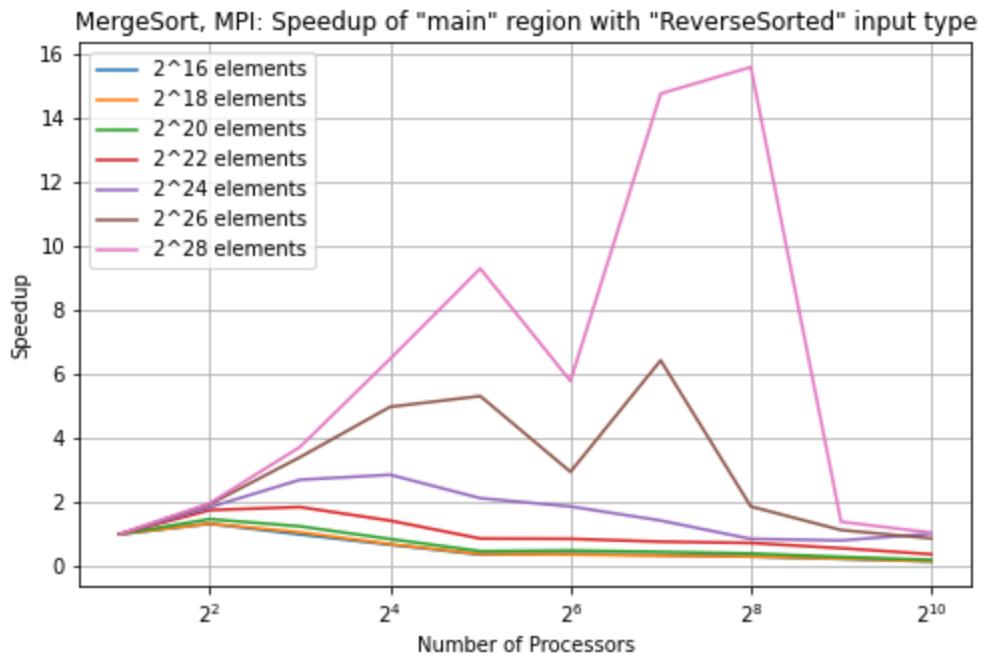


MergeSort, MPI: Speedup of "main" region with "1perturbed" input type



MergeSort, MPI: Speedup of "main" region with "Random" input type



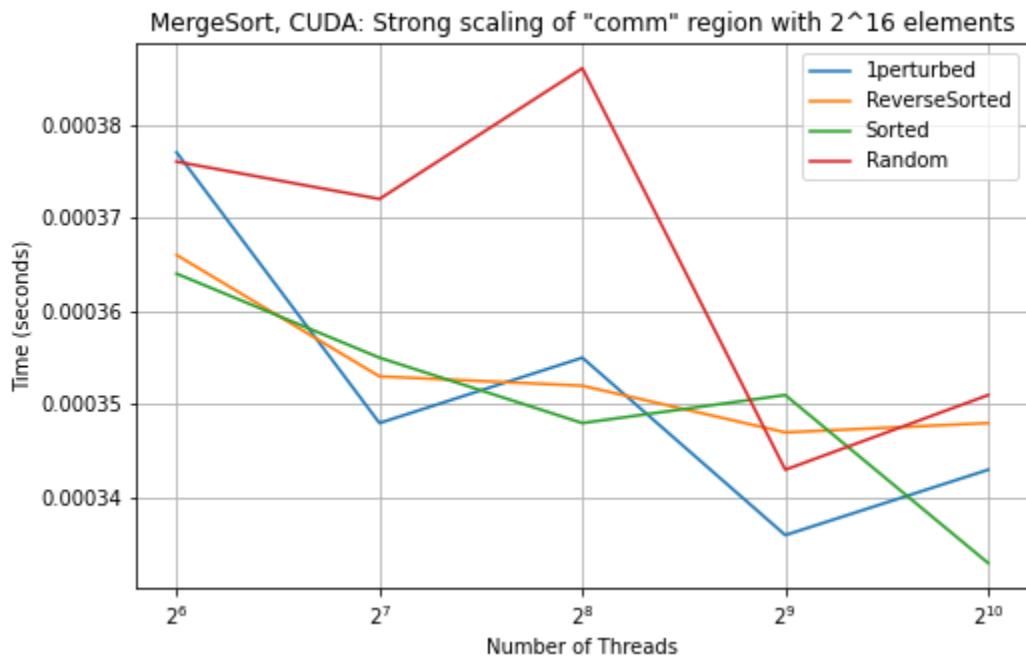


One would expect that as the number of processors increases, the runtime of the mergesort algorithm runtime decreases up until a point of diminishing returns. This is because the communication time will increase, with more workers splitting up the overall task. This holds true, and it is clear that Computation time and communication time both contribute to the overall time as the number of processors increases. It seems that, in

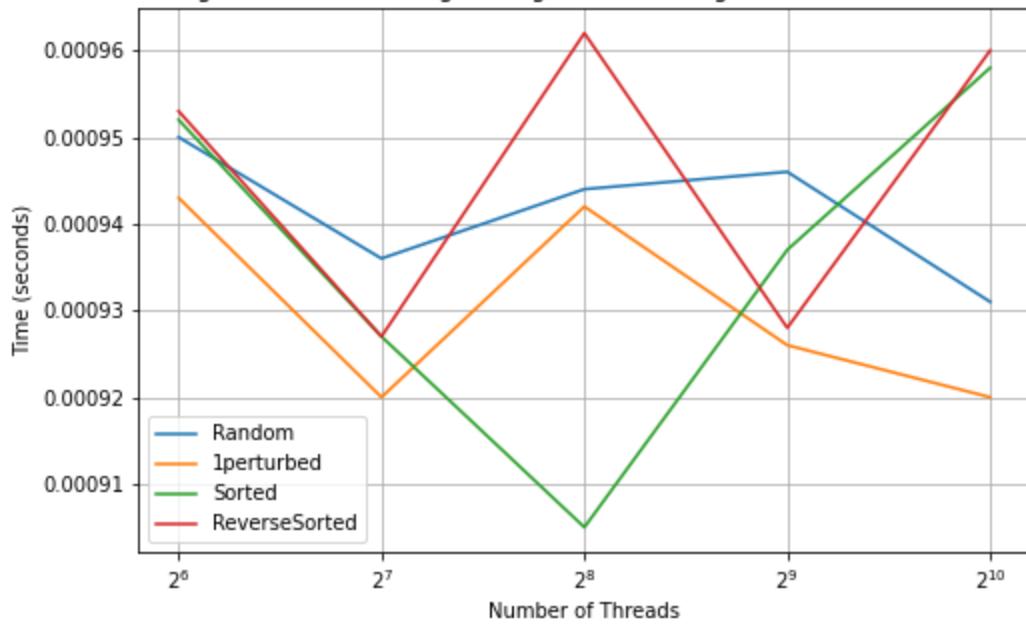
terms of communication, 2^8 processors is the ideal efficiency of this algorithm implementation at the large array input sizes, as after that, communication time drastically increases and makes the overall algorithm slower, regardless of the gains in computation time. Computation time itself has a very clear logarithmic trend of reducing time as processors increase. The computation time has a very large speedup at maximum processors, while communication speedup varies greatly. Overall, the algorithm's main function reaches a relatively small speedup of around 16 at largest input sizes, and less than 1 at smaller input sizes and high processors, which could signal some issues with the implementation of the MPI mergesort being distributed among the workers.

Merge Sort CUDA

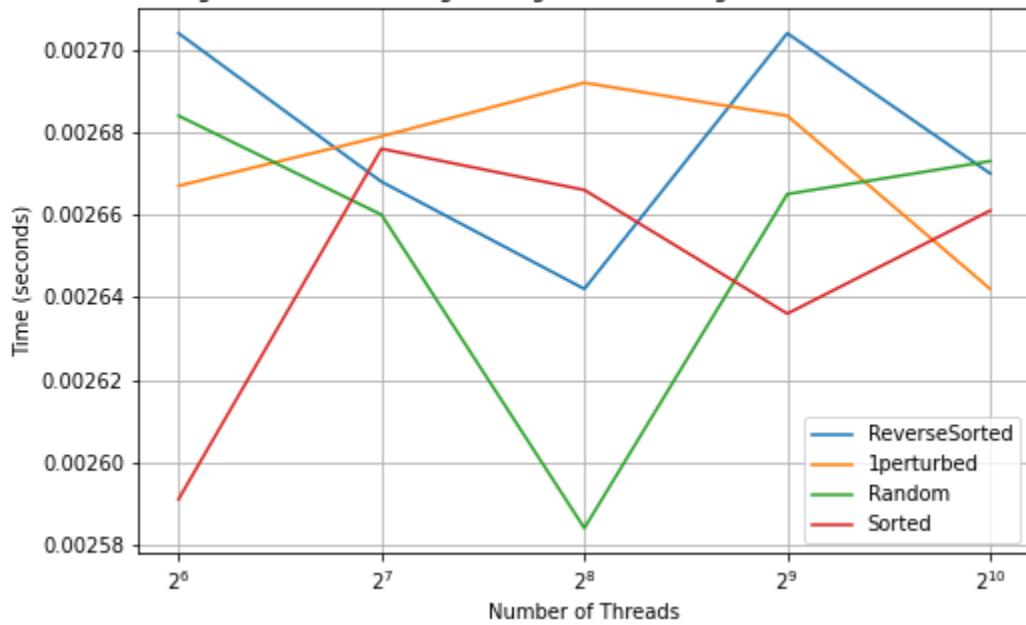
Graphs



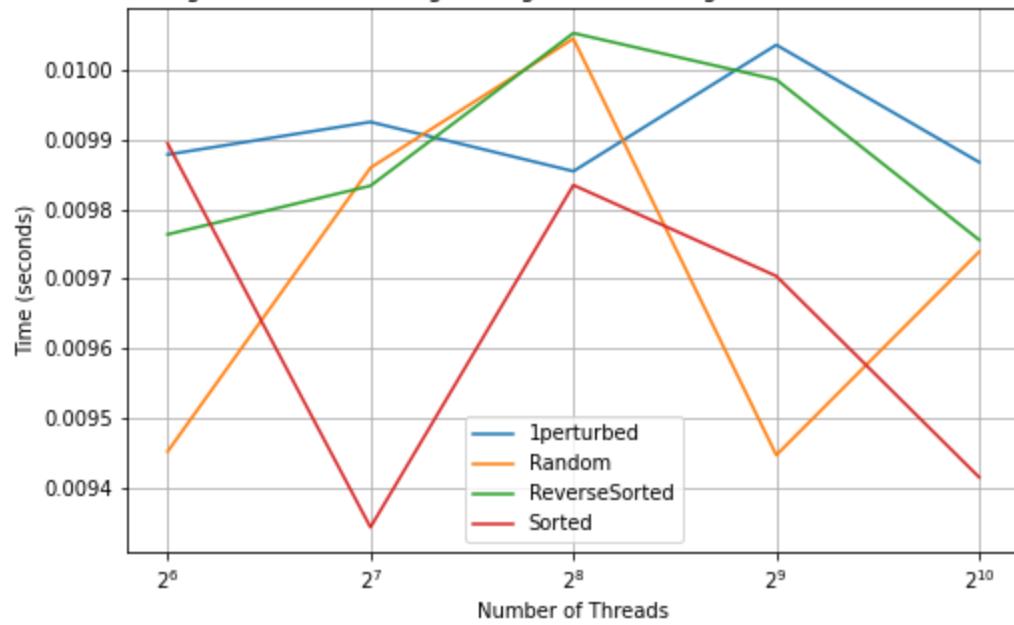
MergeSort, CUDA: Strong scaling of "comm" region with 2^{18} elements



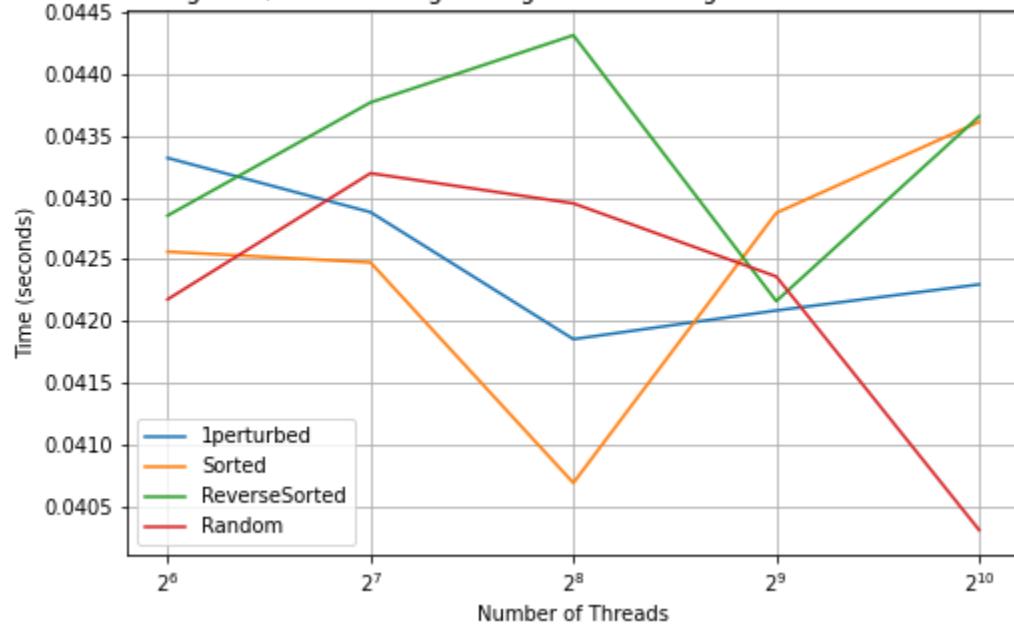
MergeSort, CUDA: Strong scaling of "comm" region with 2^{20} elements



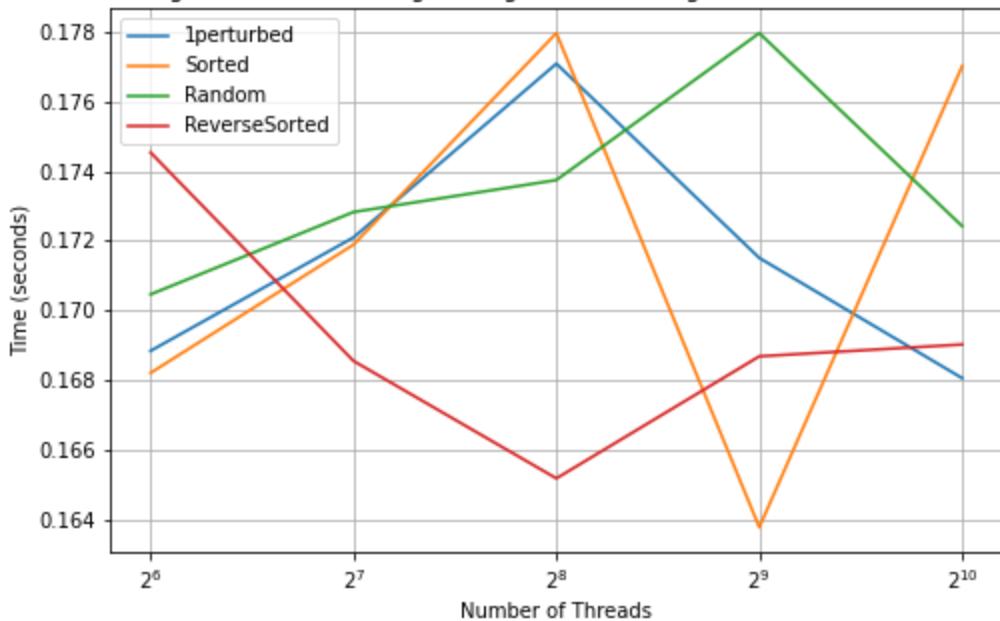
MergeSort, CUDA: Strong scaling of "comm" region with 2^{22} elements



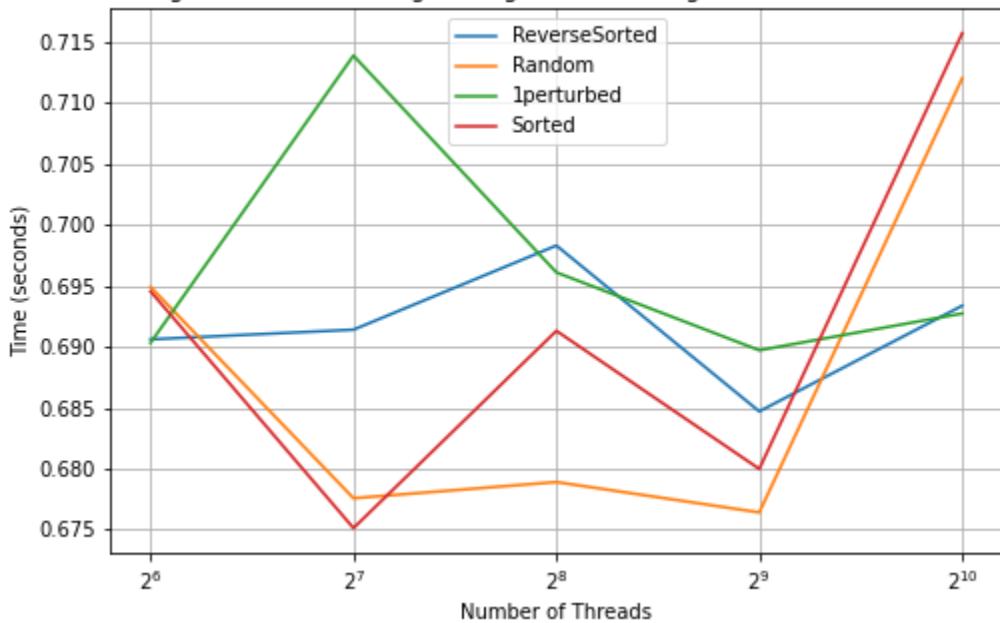
MergeSort, CUDA: Strong scaling of "comm" region with 2^{24} elements



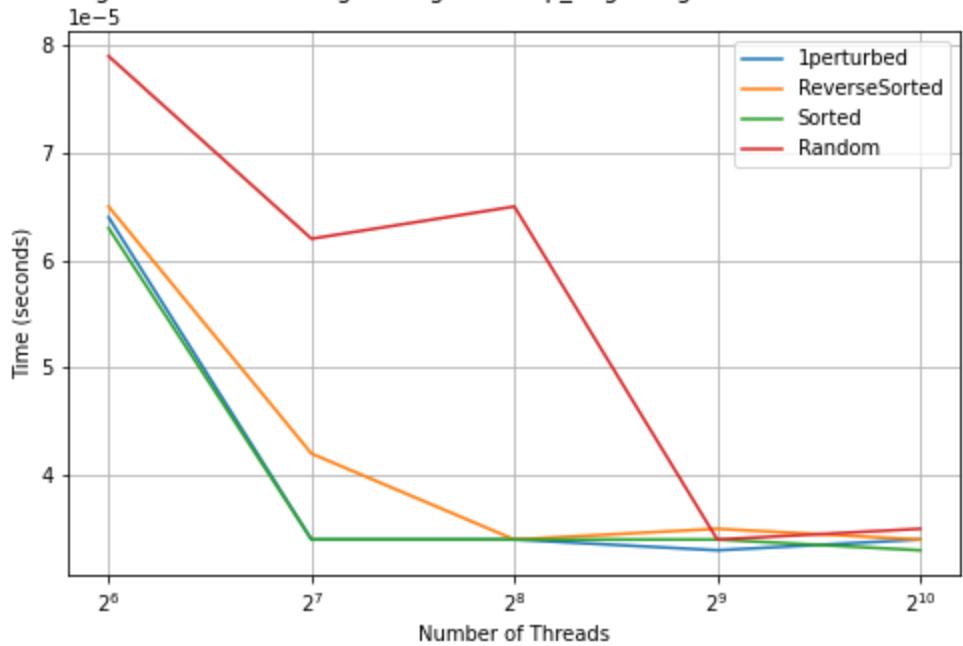
MergeSort, CUDA: Strong scaling of "comm" region with 2^{26} elements



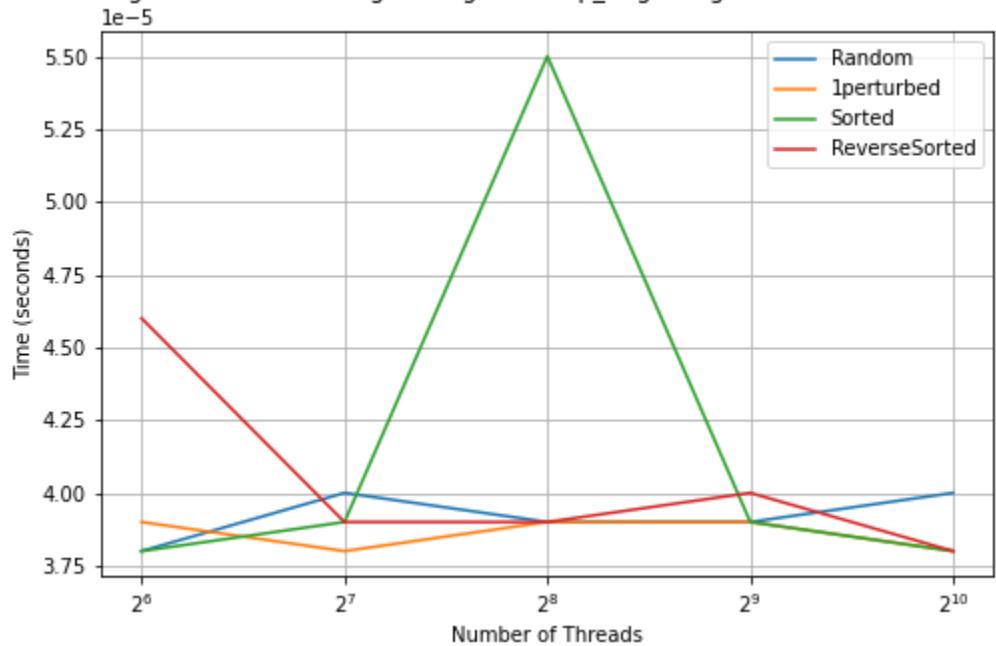
MergeSort, CUDA: Strong scaling of "comm" region with 2^{28} elements

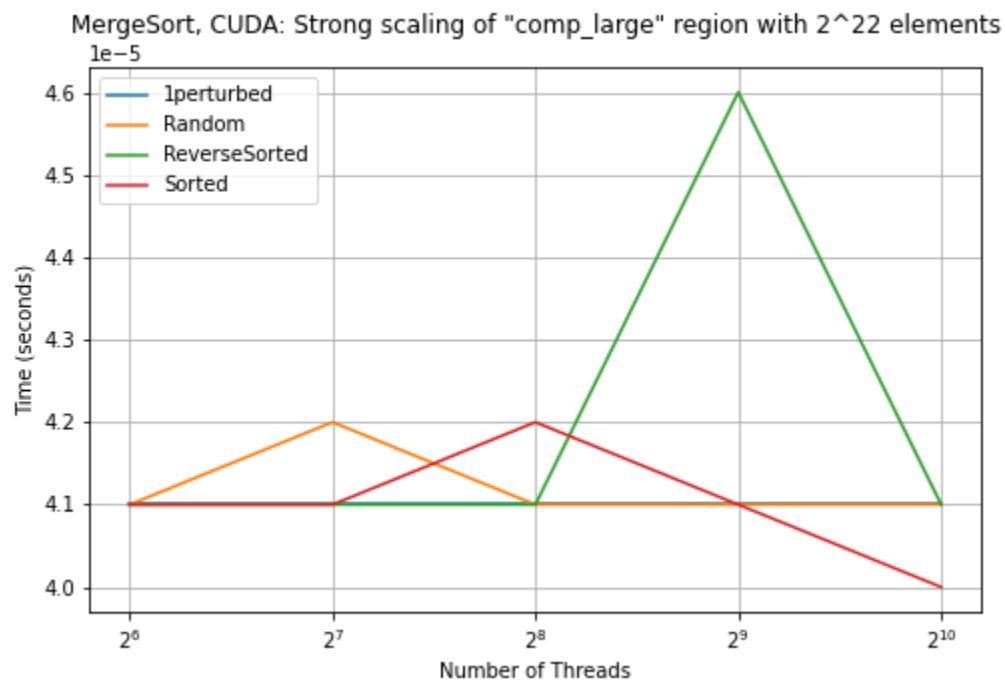
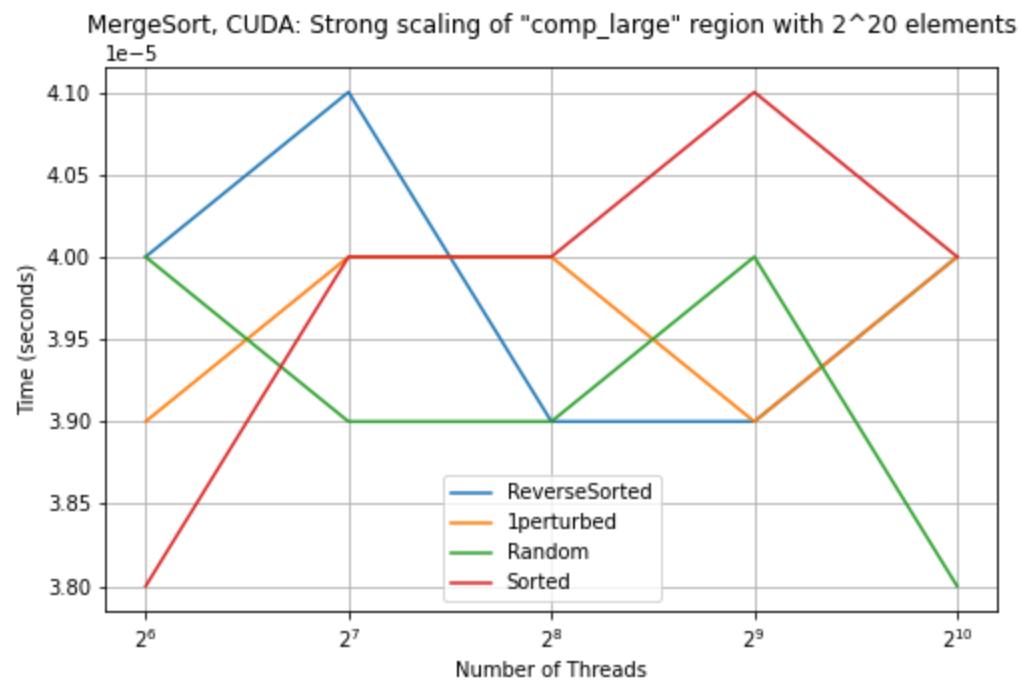


MergeSort, CUDA: Strong scaling of "comp_large" region with 2^{16} elements

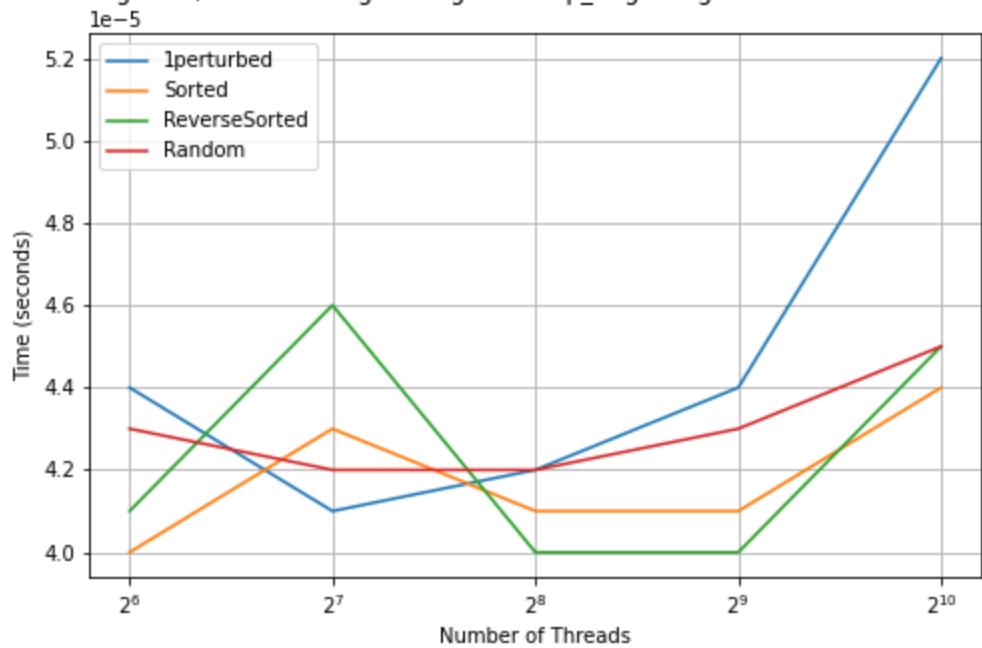


MergeSort, CUDA: Strong scaling of "comp_large" region with 2^{18} elements

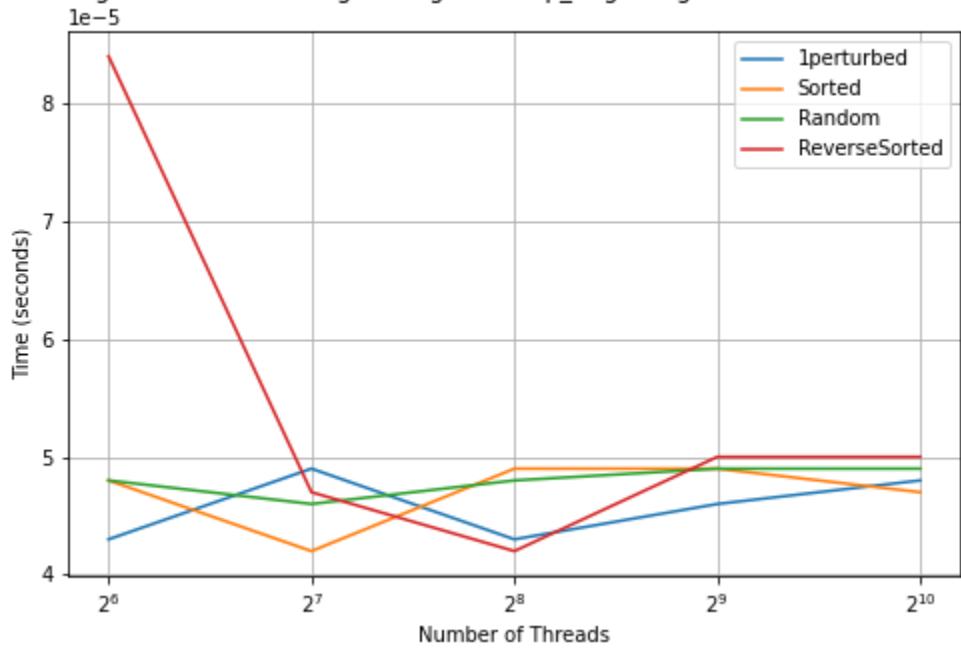




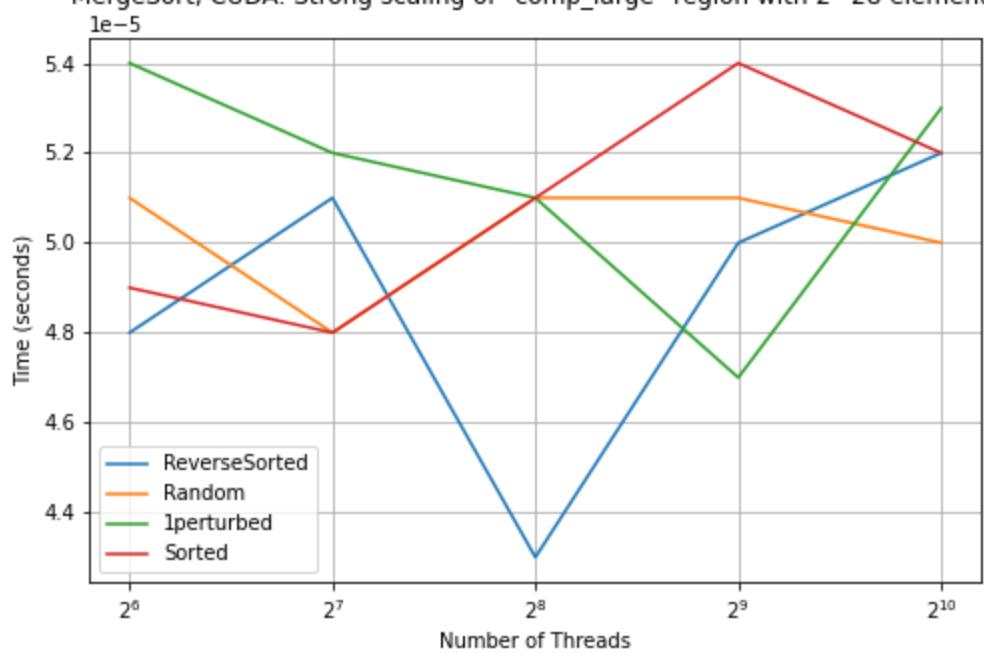
MergeSort, CUDA: Strong scaling of "comp_large" region with 2^{24} elements



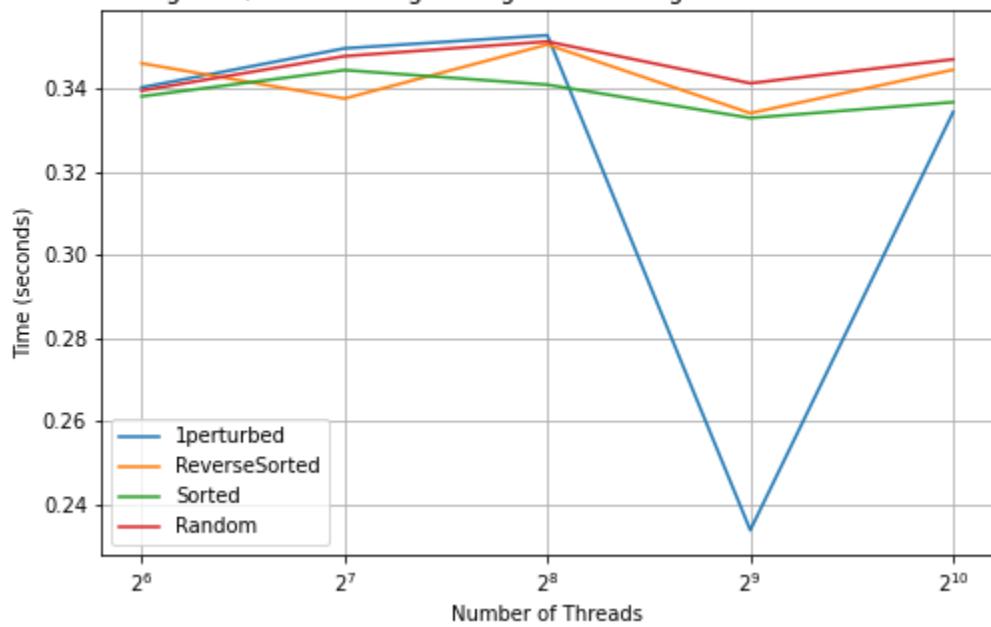
MergeSort, CUDA: Strong scaling of "comp_large" region with 2^{26} elements

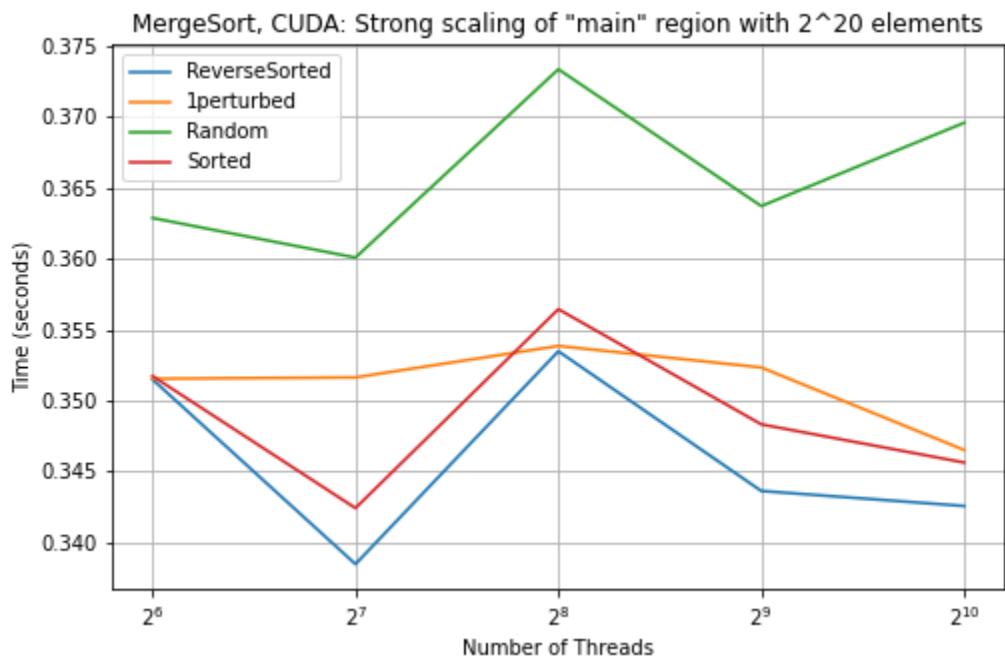
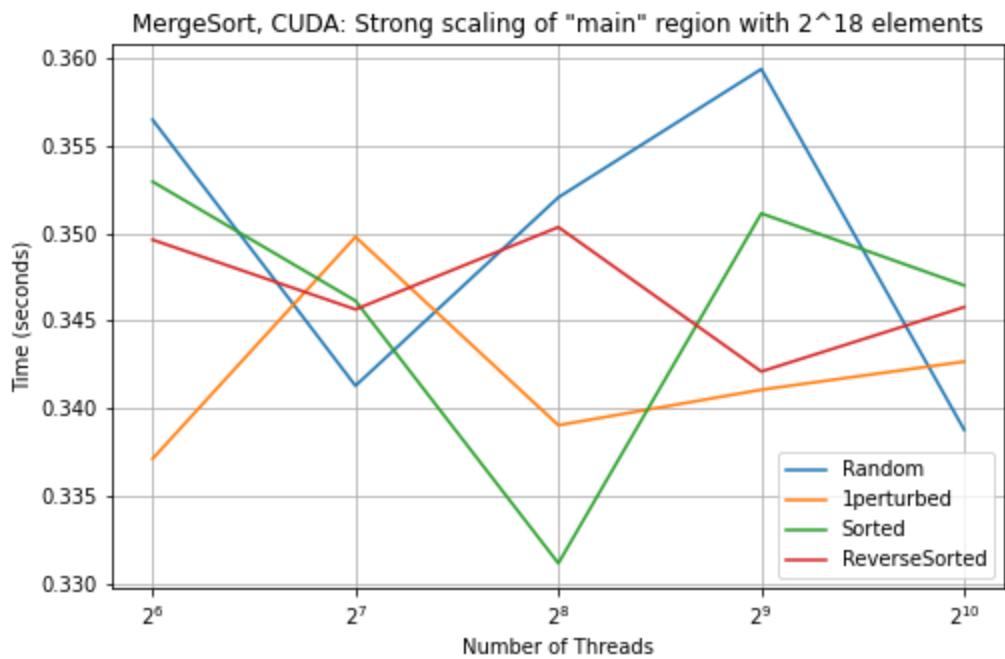


MergeSort, CUDA: Strong scaling of "comp_large" region with 2^{28} elements

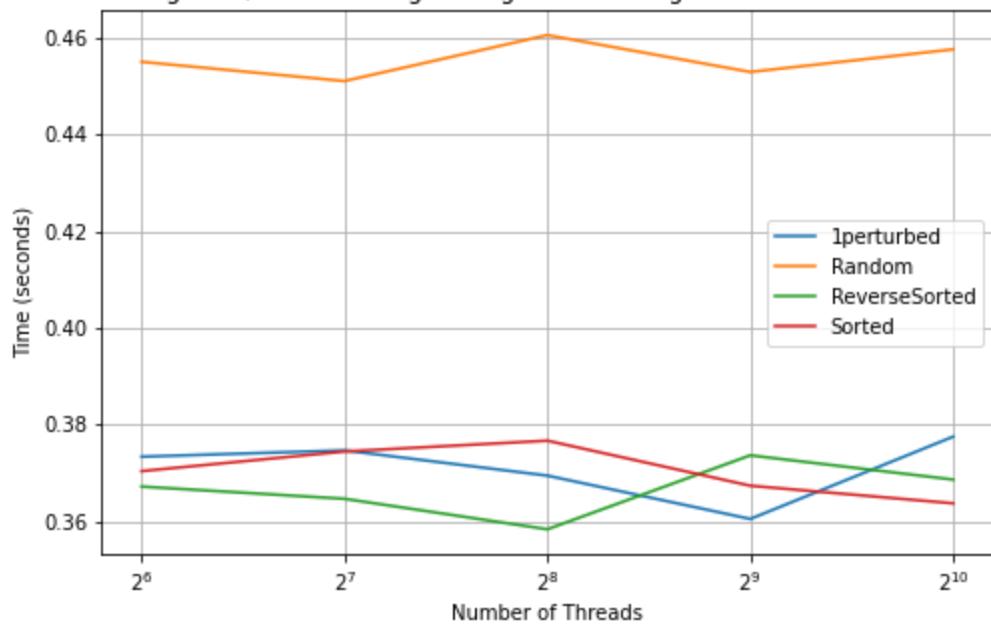


MergeSort, CUDA: Strong scaling of "main" region with 2^{16} elements

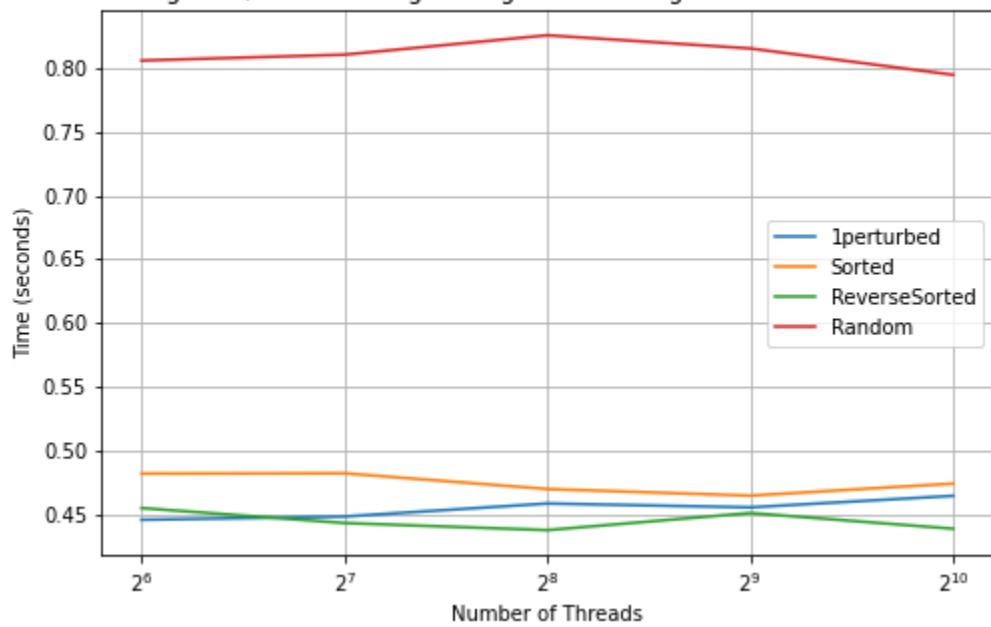




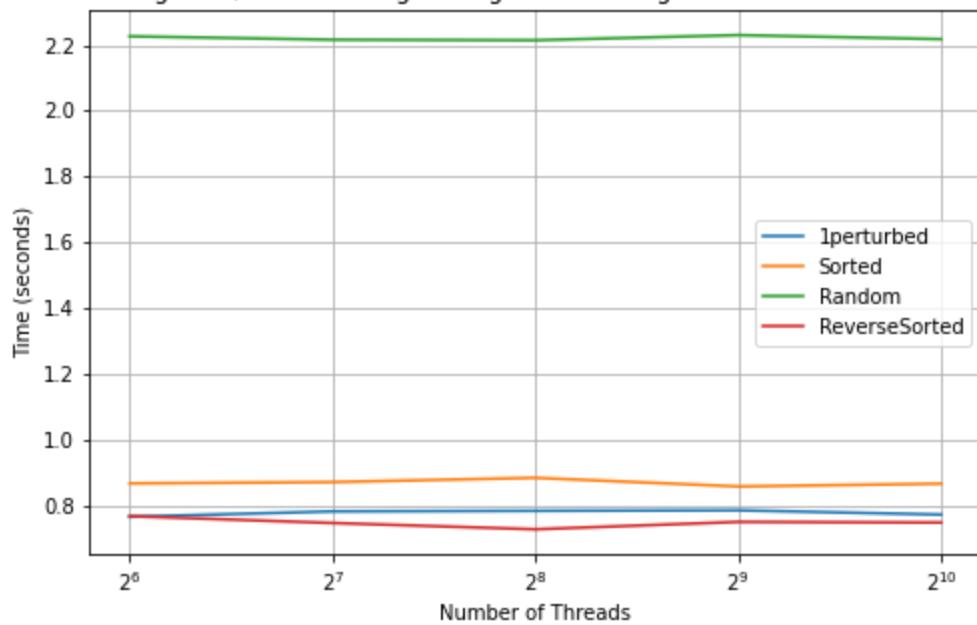
MergeSort, CUDA: Strong scaling of "main" region with 2^{22} elements



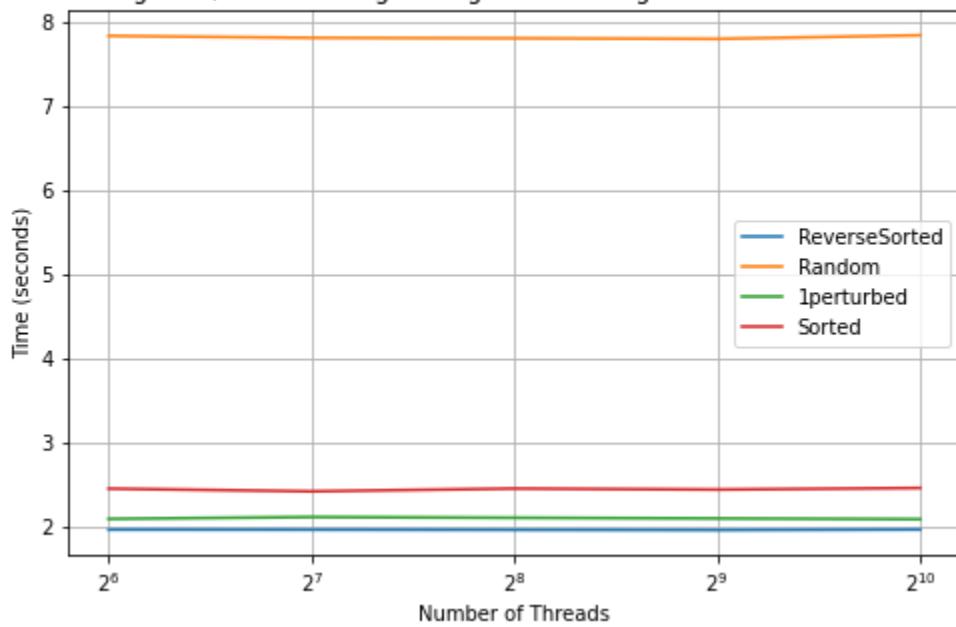
MergeSort, CUDA: Strong scaling of "main" region with 2^{24} elements



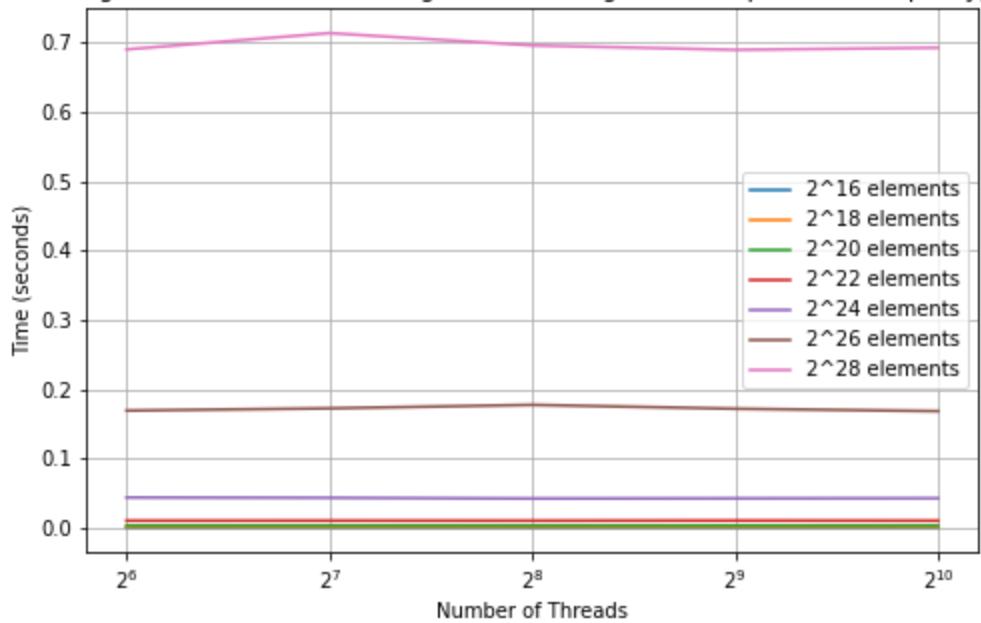
MergeSort, CUDA: Strong scaling of "main" region with 2^{26} elements



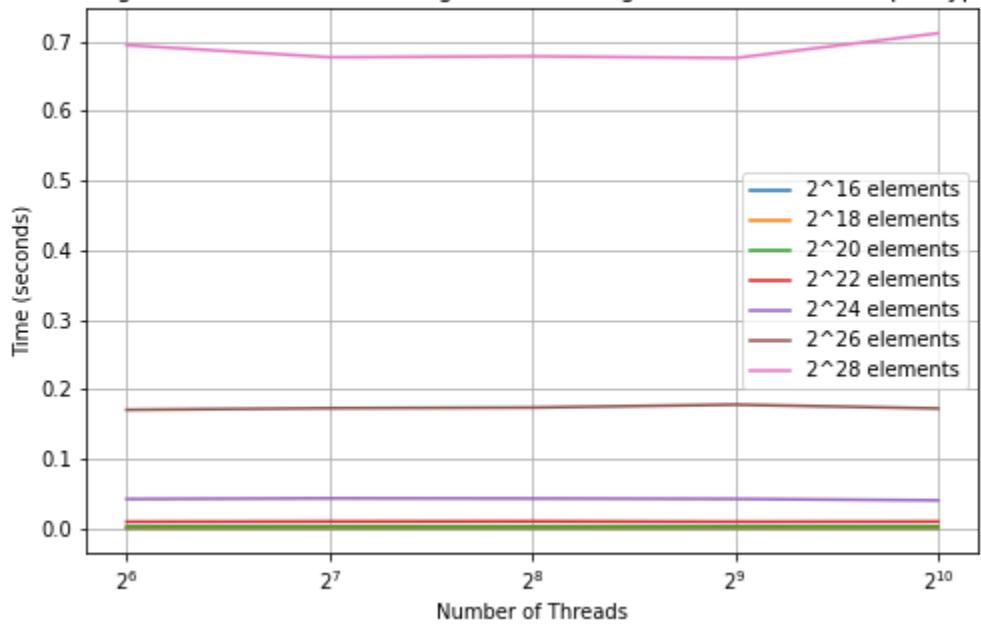
MergeSort, CUDA: Strong scaling of "main" region with 2^{28} elements



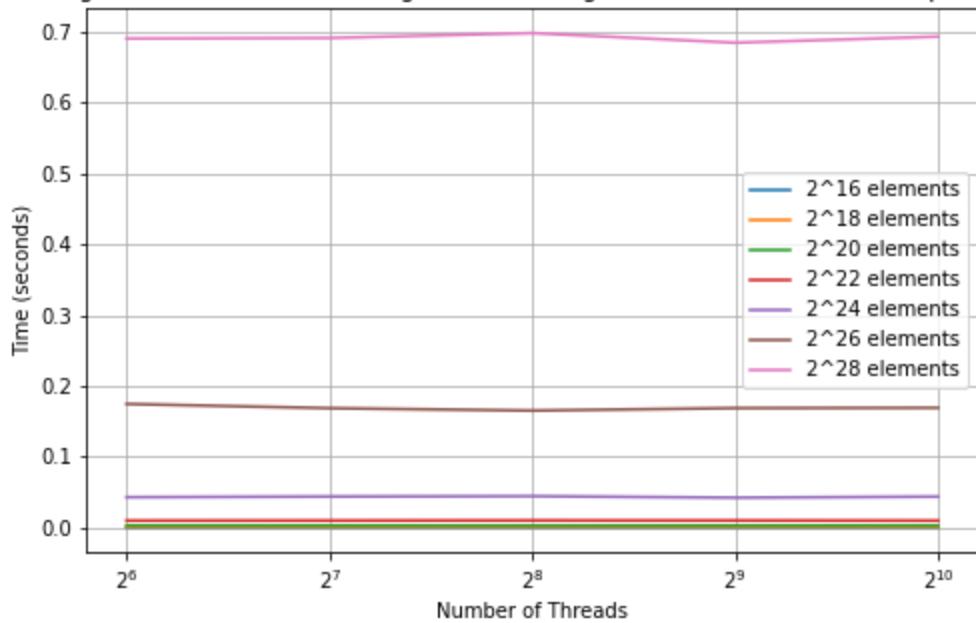
MergeSort, CUDA: Weak scaling of "comm" region with "1perturbed" input type



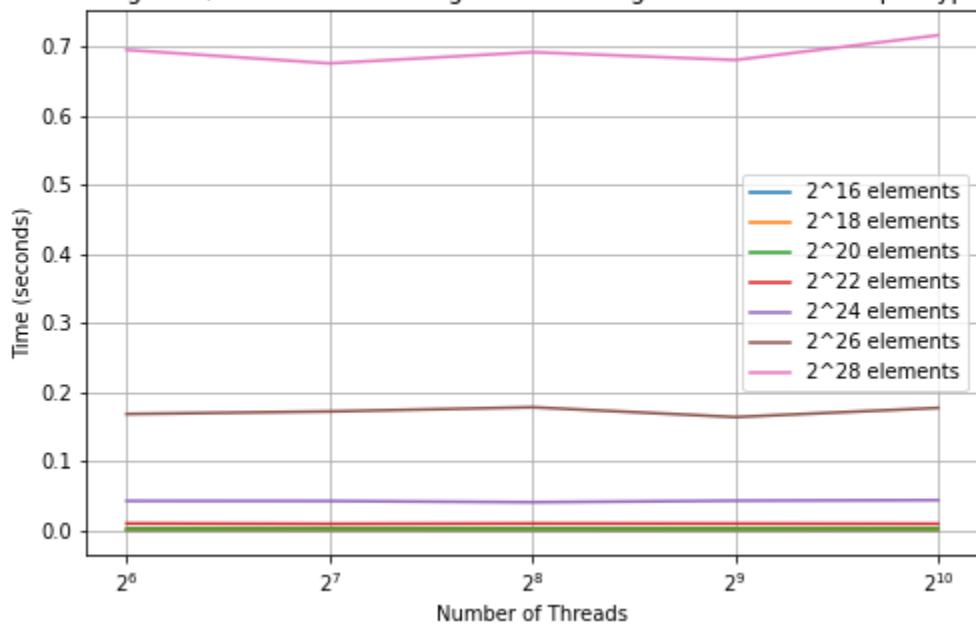
MergeSort, CUDA: Weak scaling of "comm" region with "Random" input type



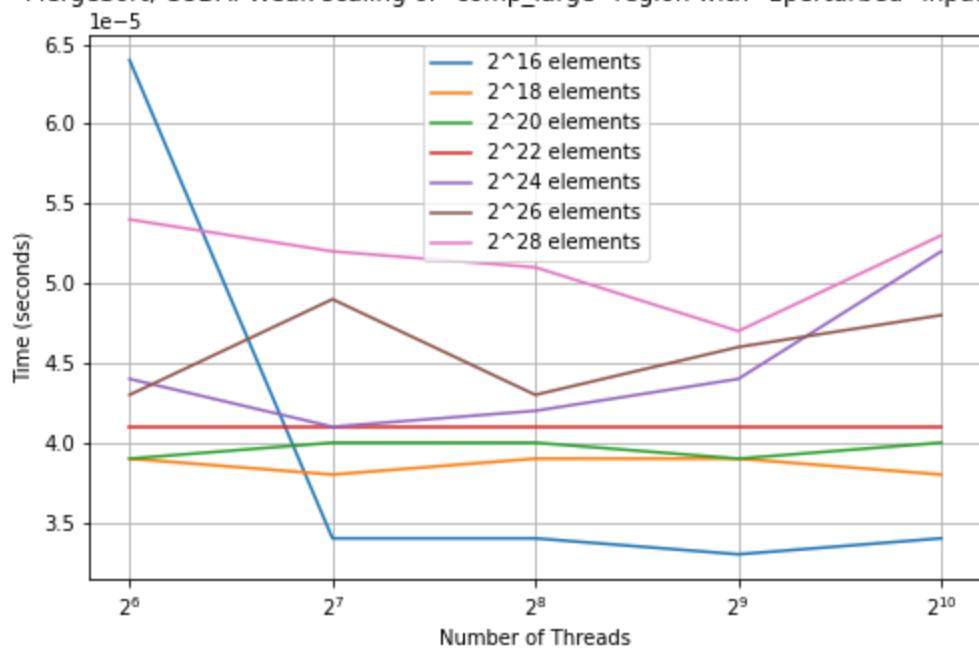
MergeSort, CUDA: Weak scaling of "comm" region with "ReverseSorted" input type



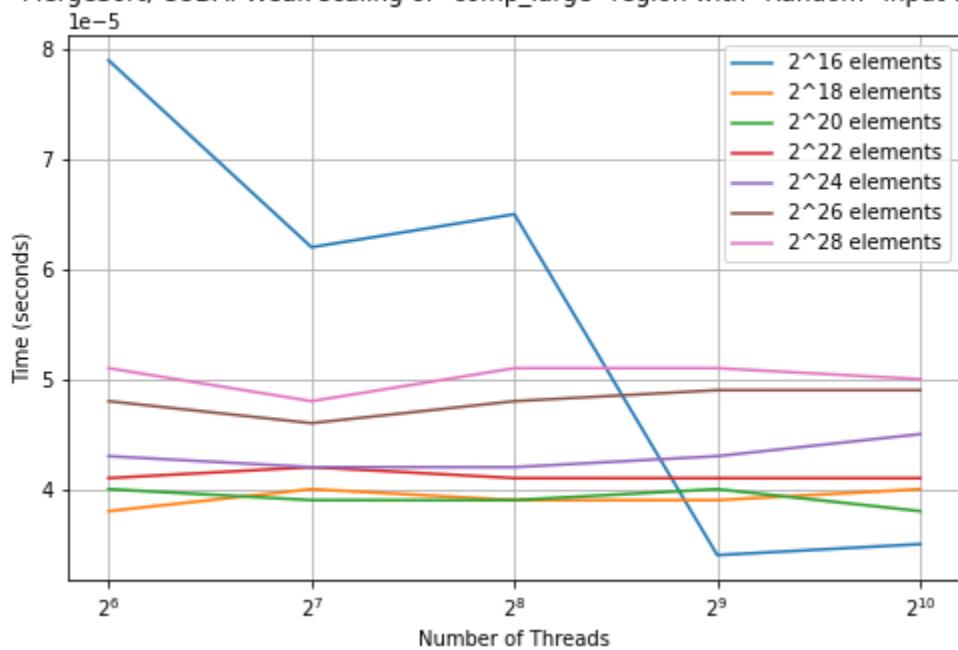
MergeSort, CUDA: Weak scaling of "comm" region with "Sorted" input type



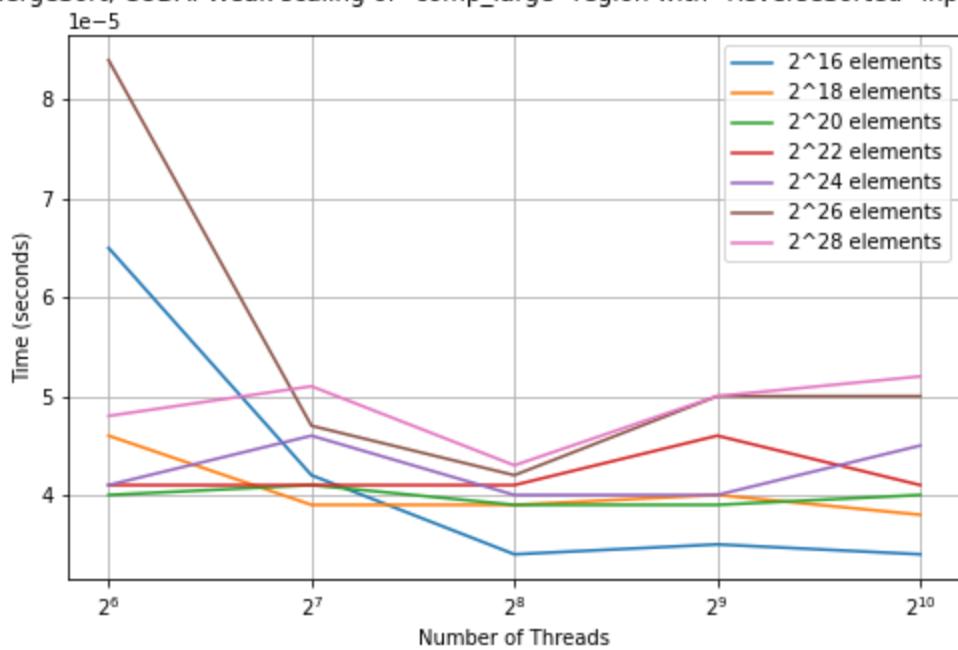
MergeSort, CUDA: Weak scaling of "comp_large" region with "1perturbed" input type



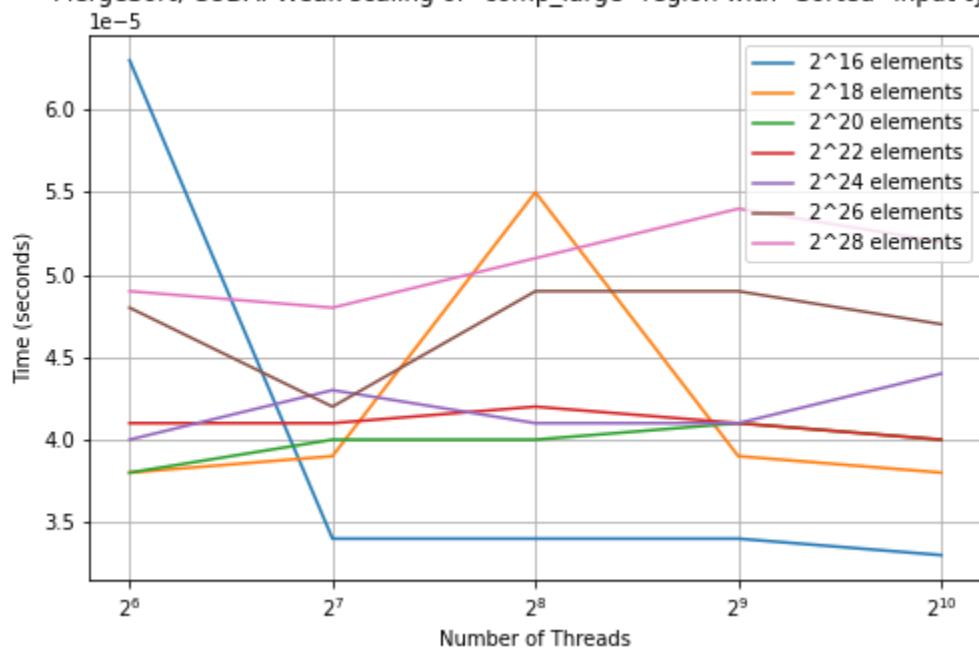
MergeSort, CUDA: Weak scaling of "comp_large" region with "Random" input type



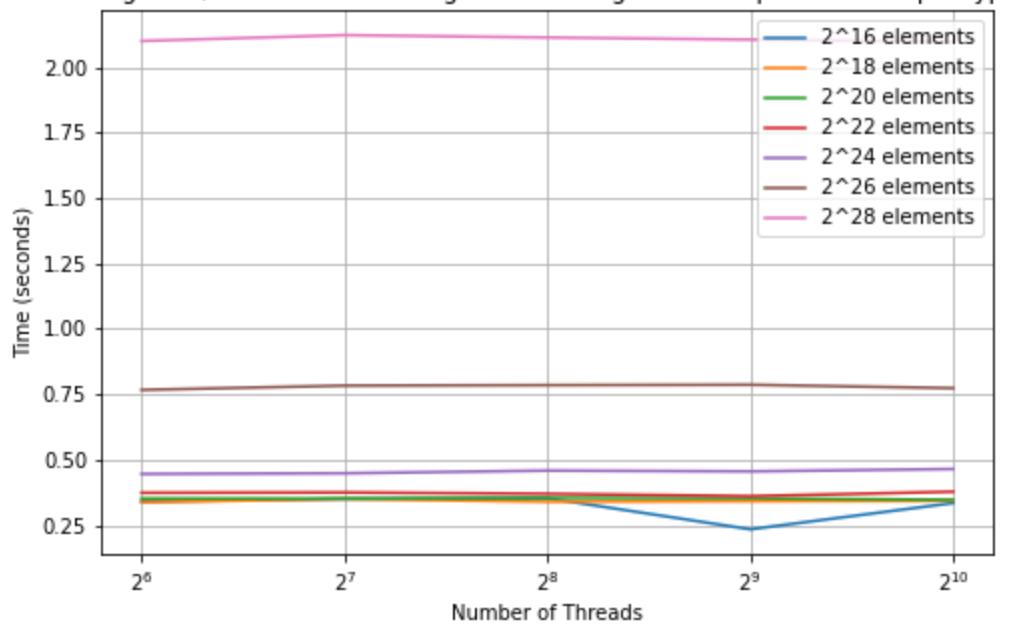
MergeSort, CUDA: Weak scaling of "comp_large" region with "ReverseSorted" input type



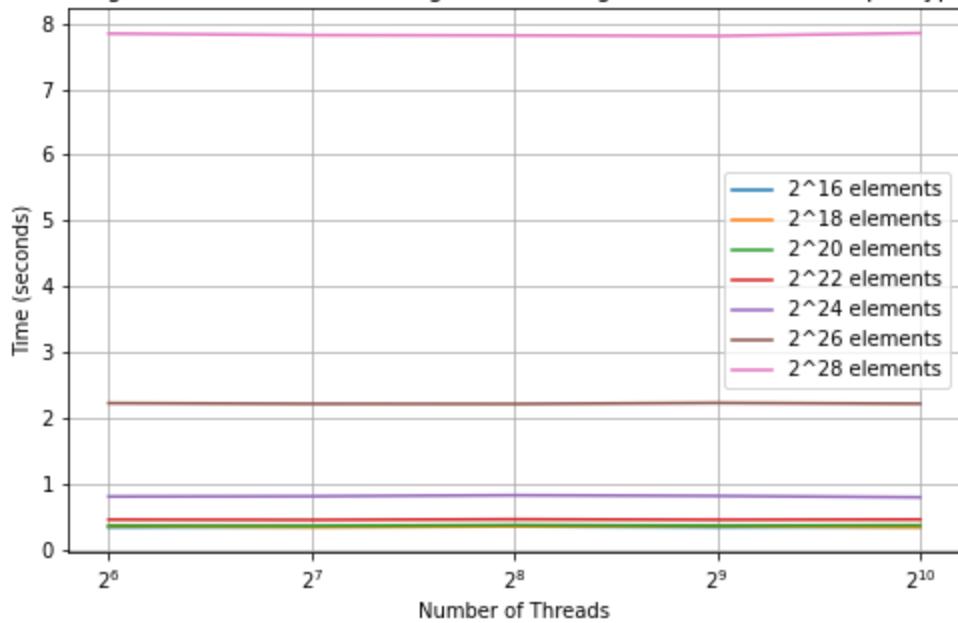
MergeSort, CUDA: Weak scaling of "comp_large" region with "Sorted" input type



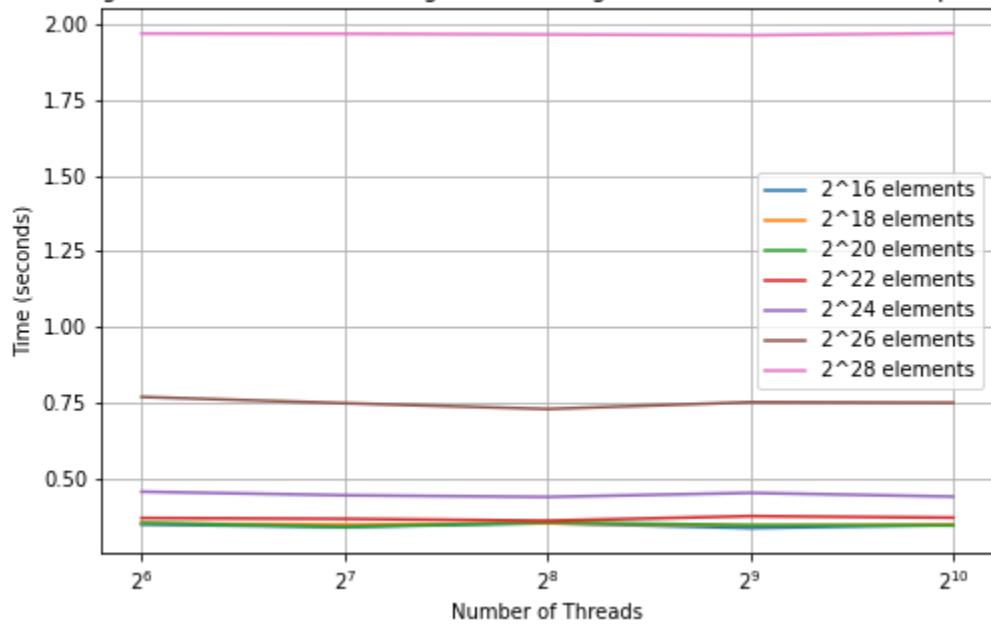
MergeSort, CUDA: Weak scaling of "main" region with "1perturbed" input type



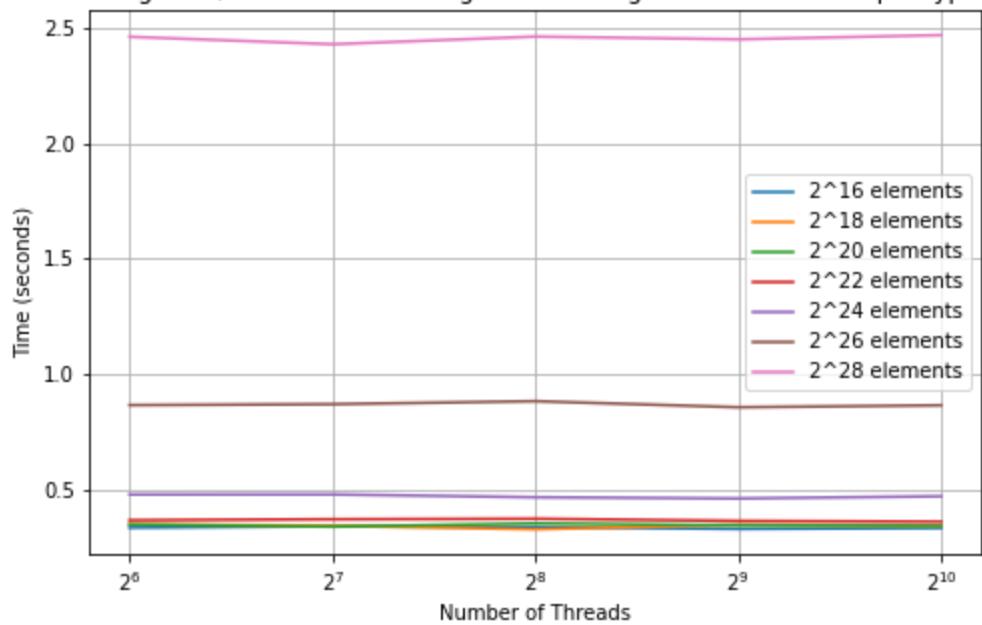
MergeSort, CUDA: Weak scaling of "main" region with "Random" input type

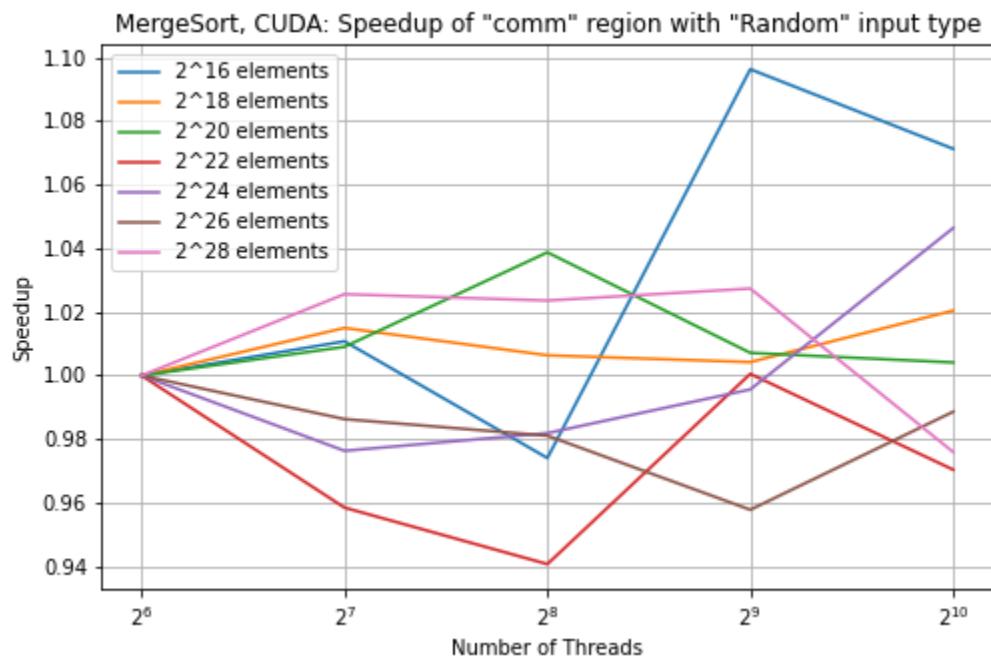
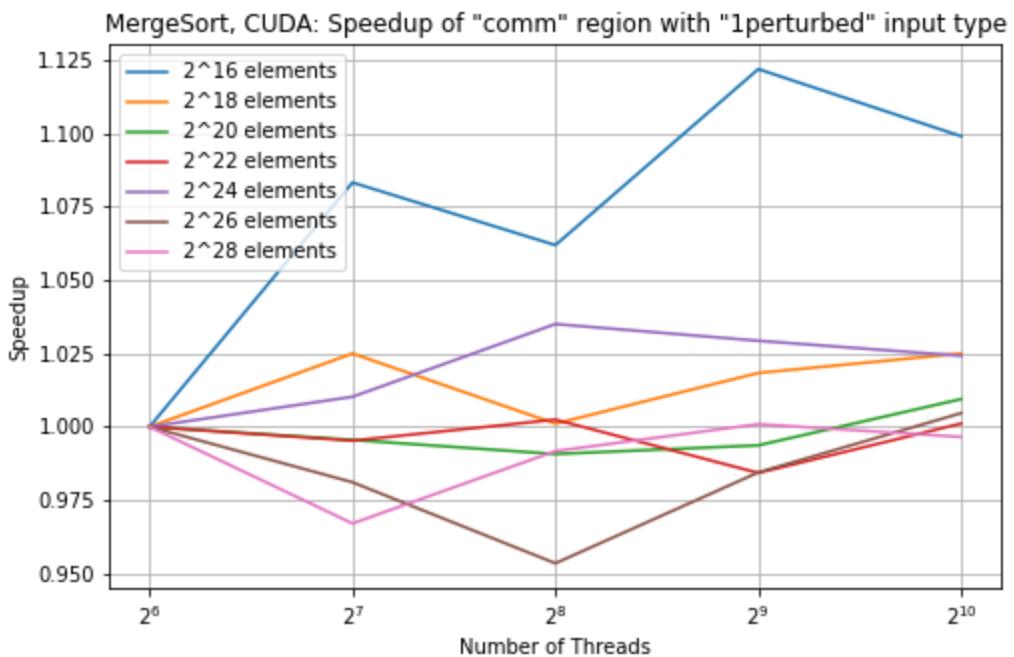


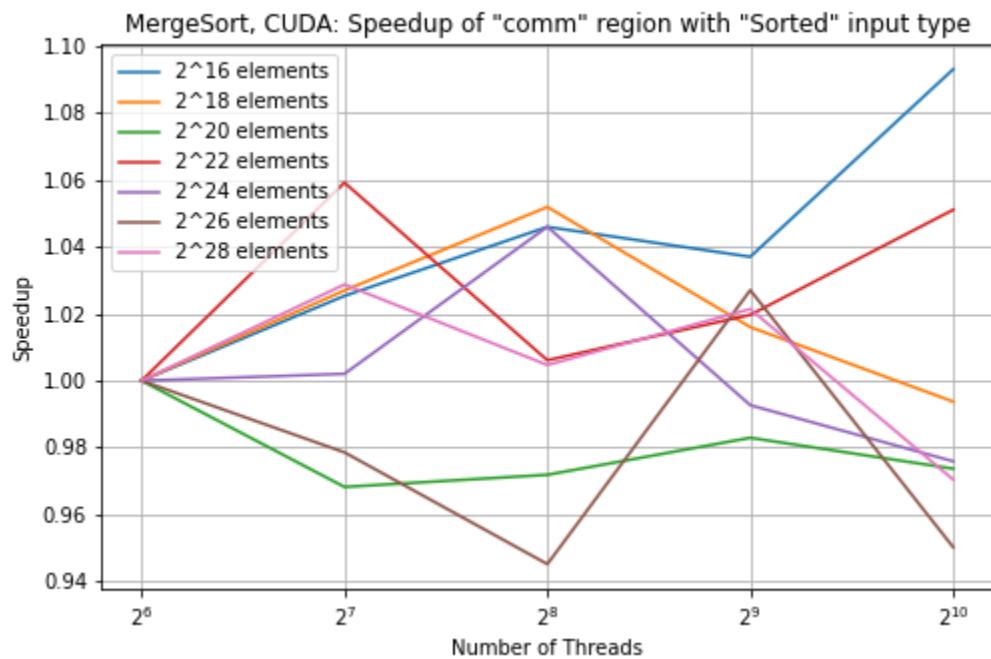
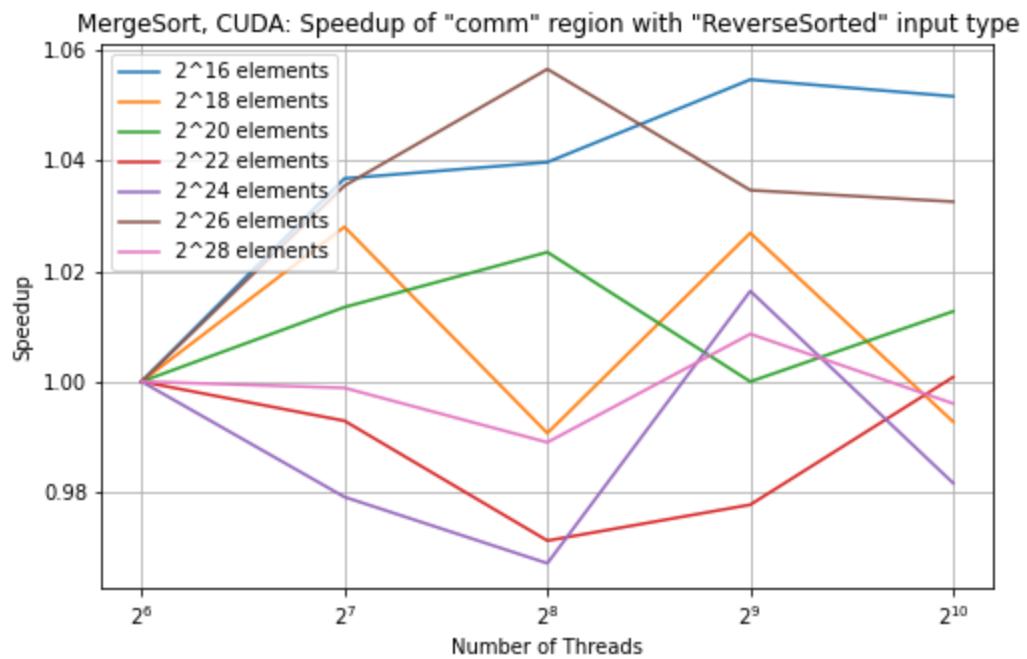
MergeSort, CUDA: Weak scaling of "main" region with "ReverseSorted" input type



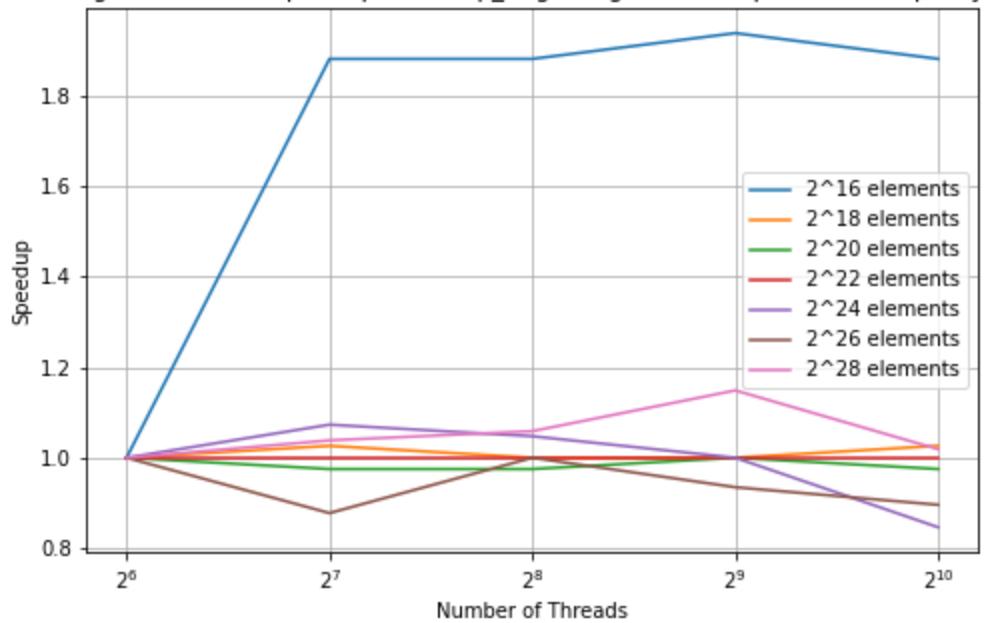
MergeSort, CUDA: Weak scaling of "main" region with "Sorted" input type



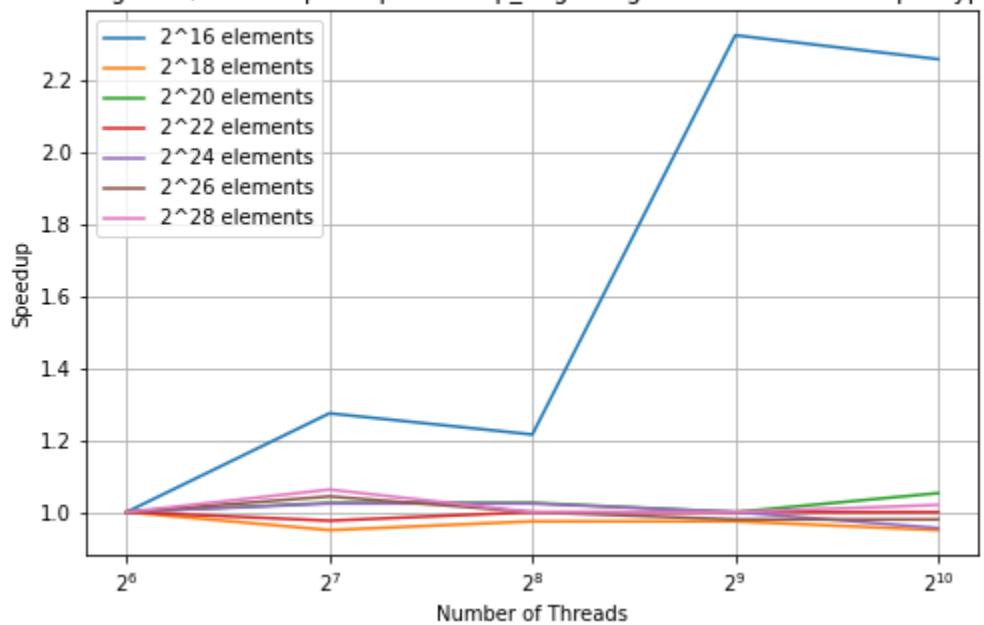




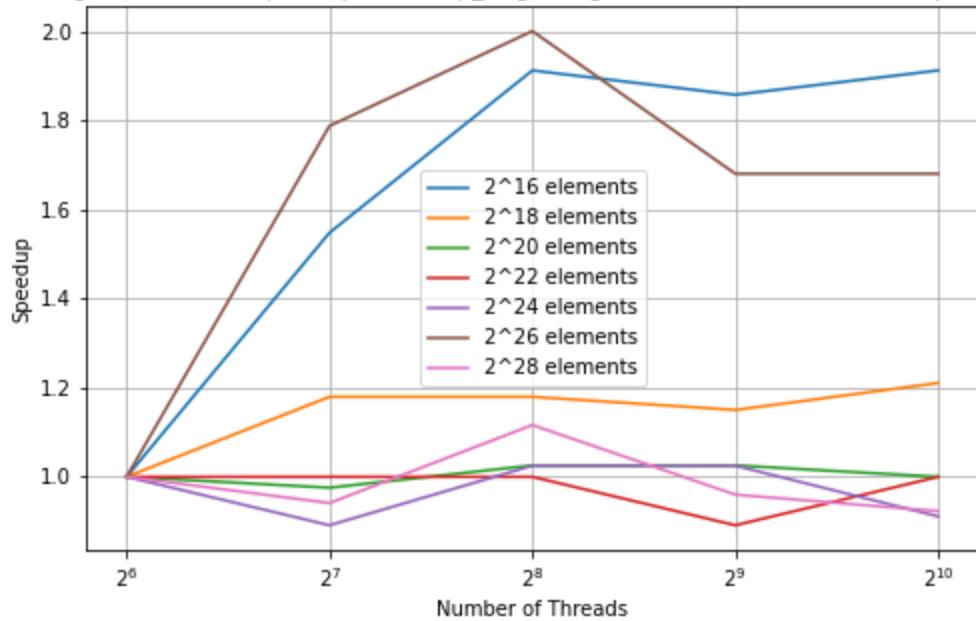
MergeSort, CUDA: Speedup of "comp_large" region with "1perturbed" input type



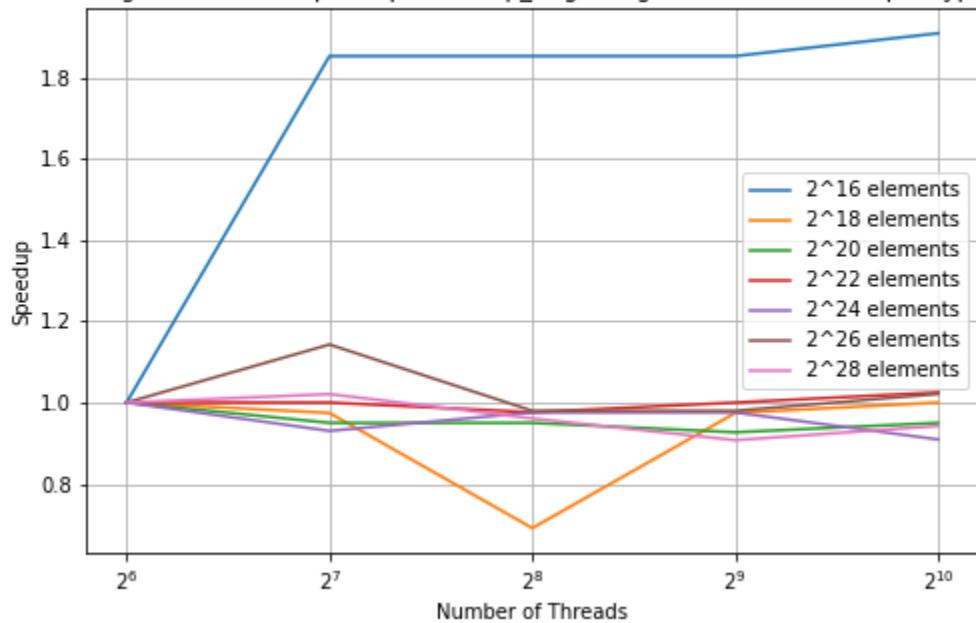
MergeSort, CUDA: Speedup of "comp_large" region with "Random" input type



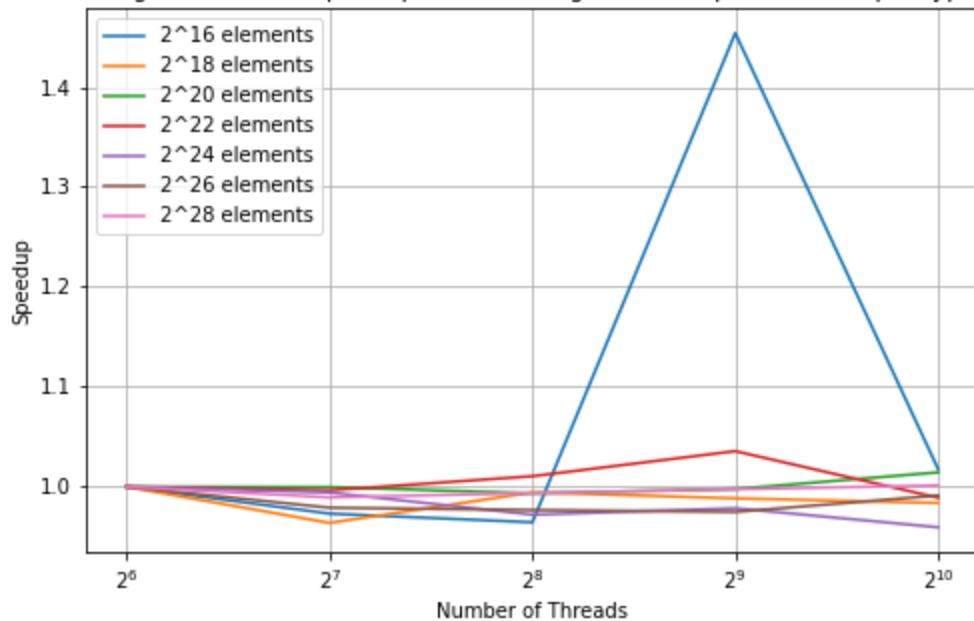
MergeSort, CUDA: Speedup of "comp_large" region with "ReverseSorted" input type



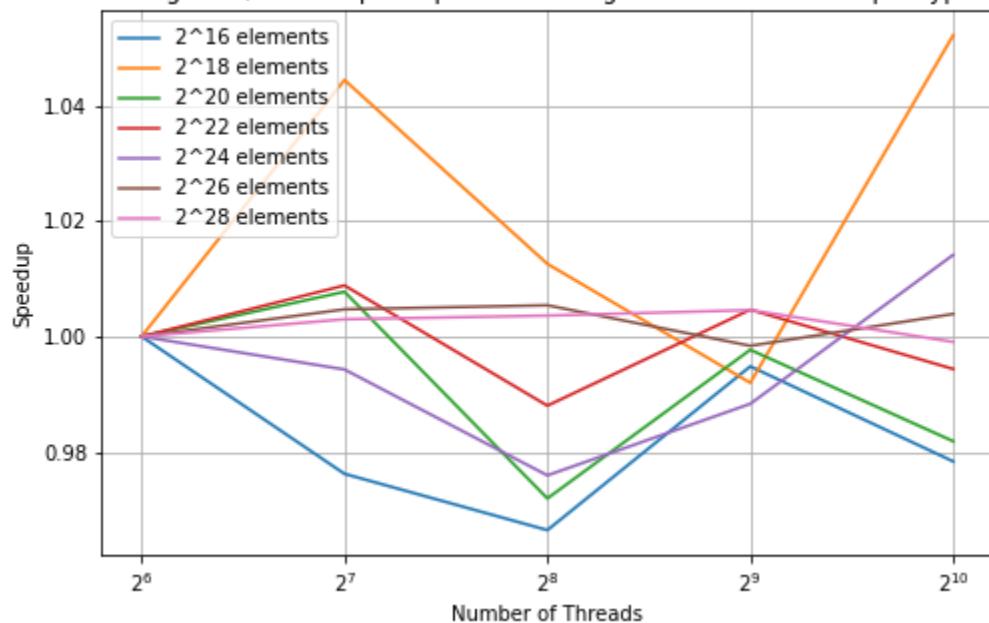
MergeSort, CUDA: Speedup of "comp_large" region with "Sorted" input type



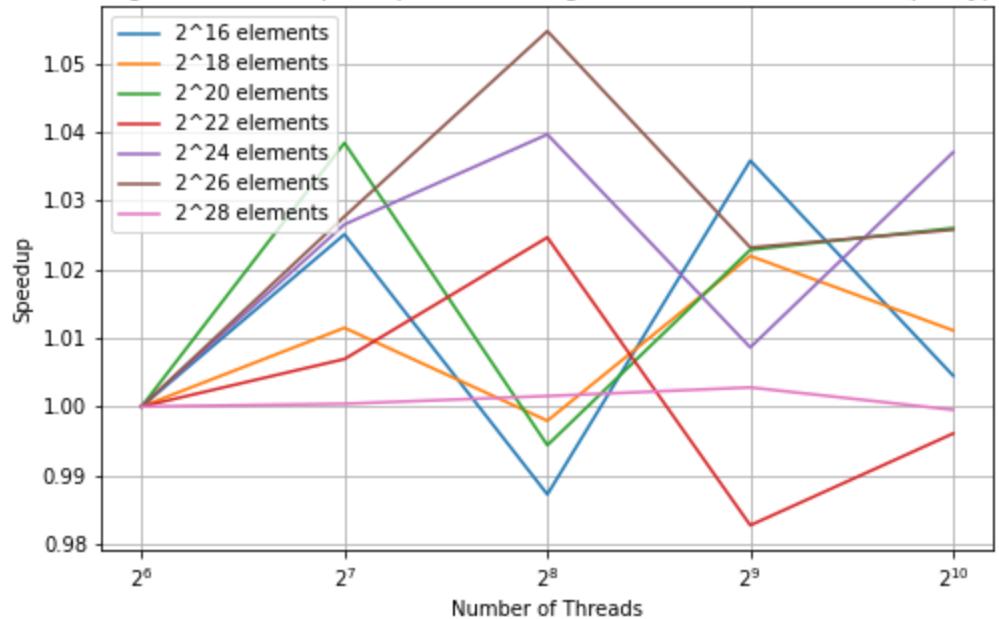
MergeSort, CUDA: Speedup of "main" region with "1perturbed" input type



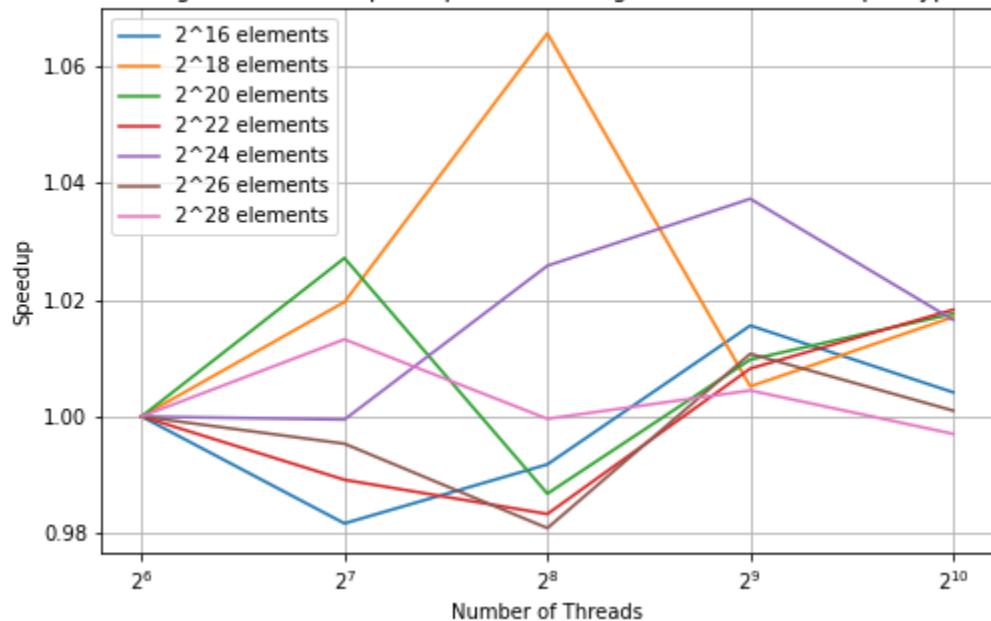
MergeSort, CUDA: Speedup of "main" region with "Random" input type



MergeSort, CUDA: Speedup of "main" region with "ReverseSorted" input type



MergeSort, CUDA: Speedup of "main" region with "Sorted" input type

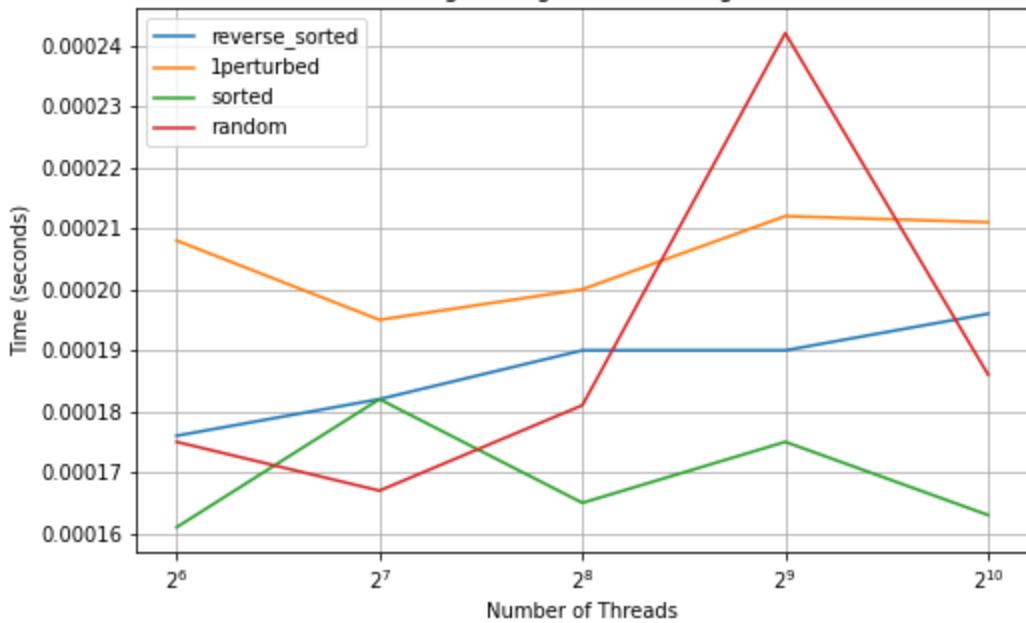


As the number of threads increases, the runtime of the mergesort CUDA algorithm varies greatly. There is not a clear trend of decreasing runtime with increasing threads from these graphs, which may signal an implementation issue with CUDA mergesort in this code. For the smallest array input size, there is a general trend of time reduction in

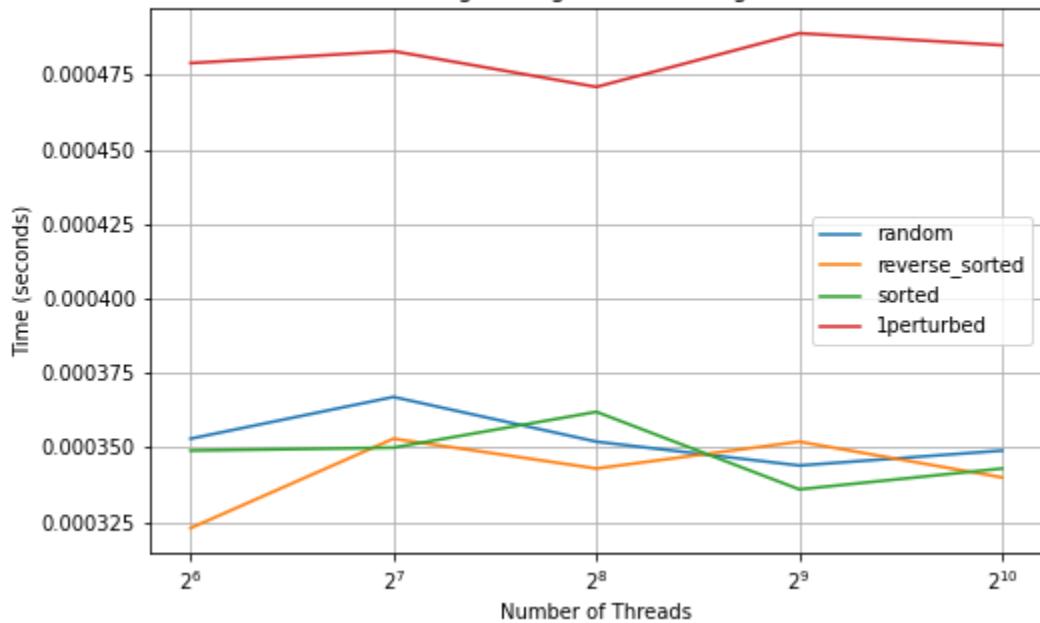
the strong scaling graphs, however, the higher array input sizes fail to show clear trends. There are linear lines in the weak scaling of the algorithm at various input arrays, signaling that regardless of the thread size, the algorithm runtime will stay consistent. This shows that the splitting of the tasks between threads is not happening efficiently, or the sequential portions of the algorithm are still having the greatest impact on overall performance. For some array input sizes, there is minimal speedup of around 1, which again might indicate that the mergesort CUDA implementation is not effectively splitting up the tasks.

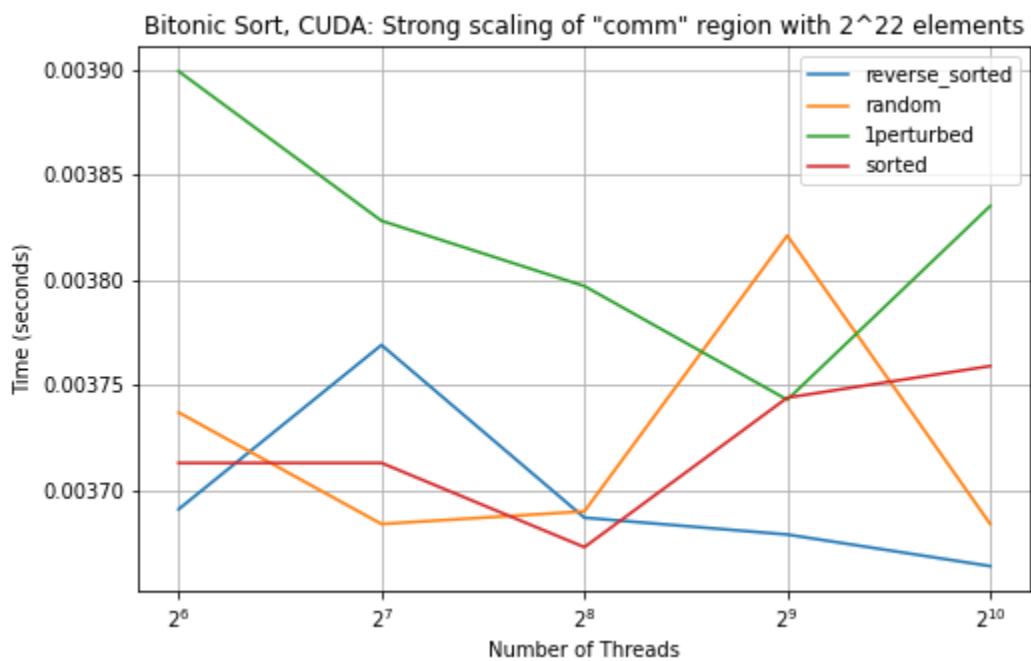
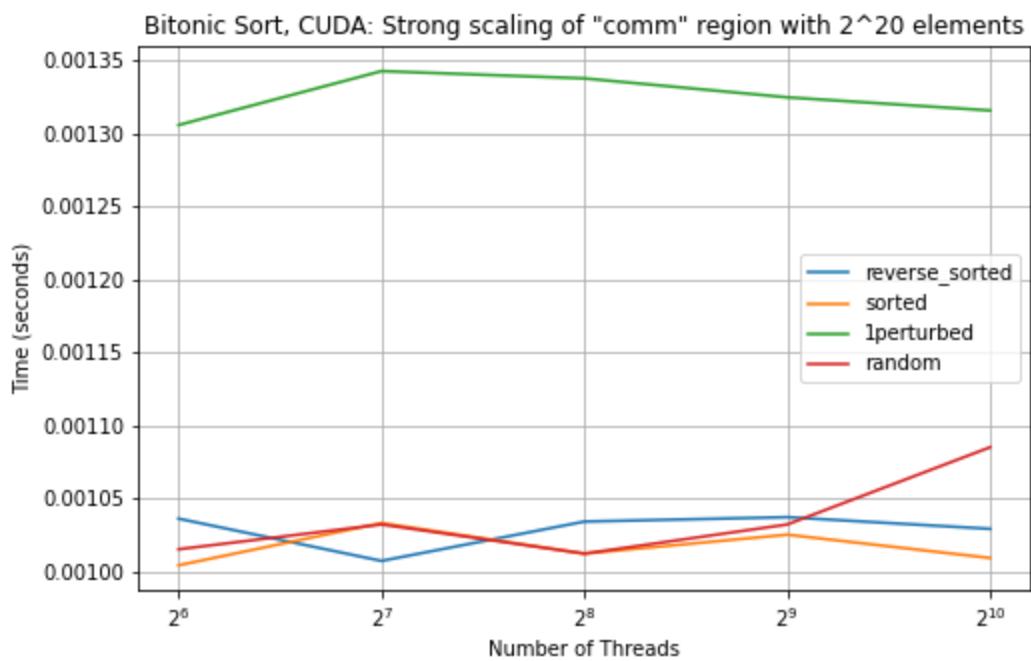
Bitonic Sort CUDA

Bitonic Sort, CUDA: Strong scaling of "comm" region with 2^{16} elements

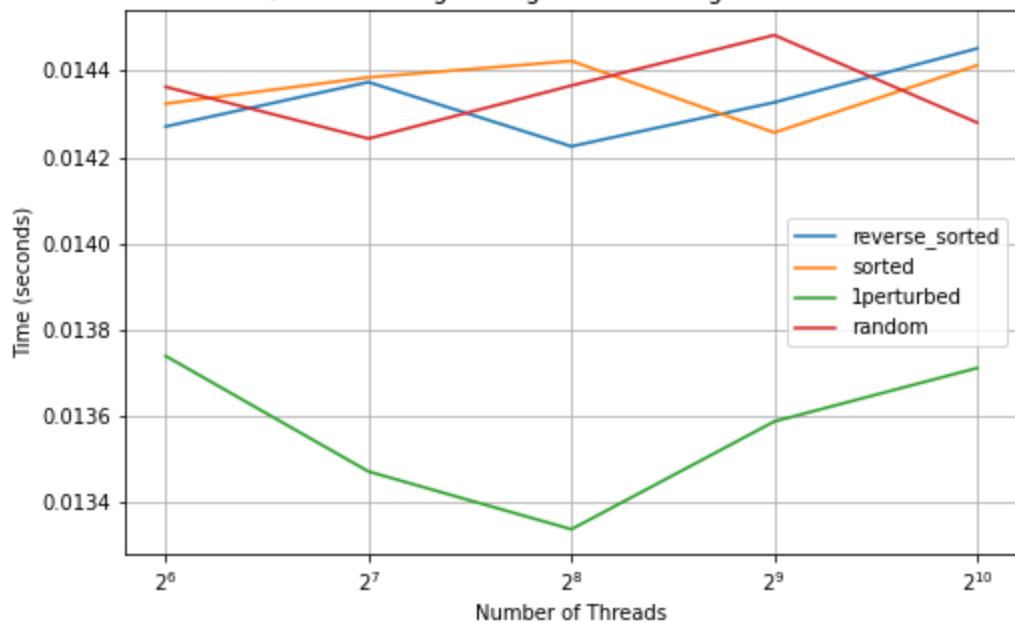


Bitonic Sort, CUDA: Strong scaling of "comm" region with 2^{18} elements

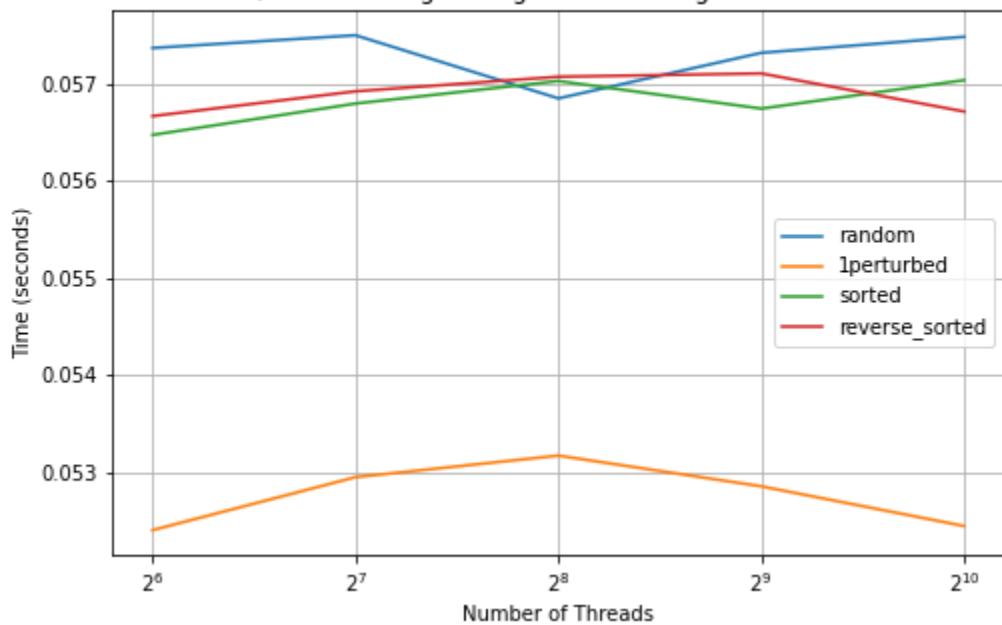


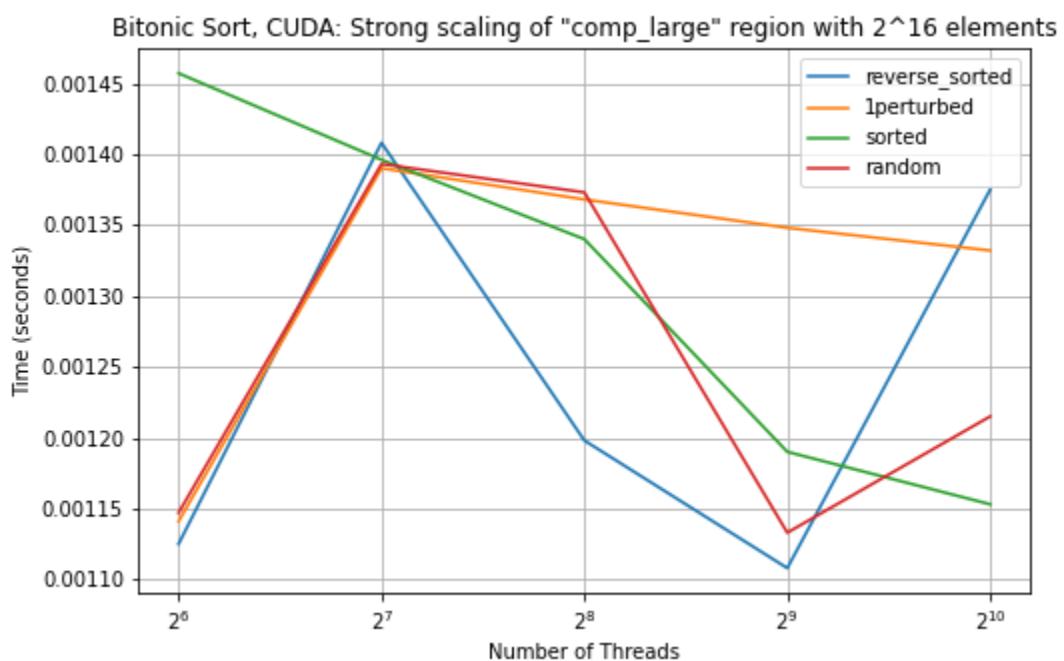
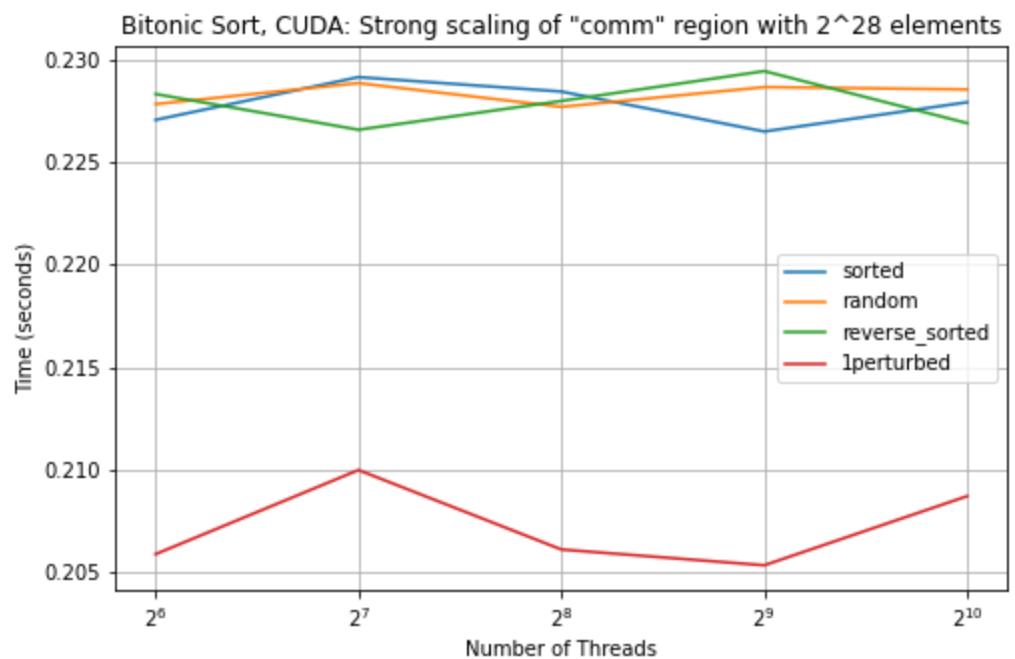


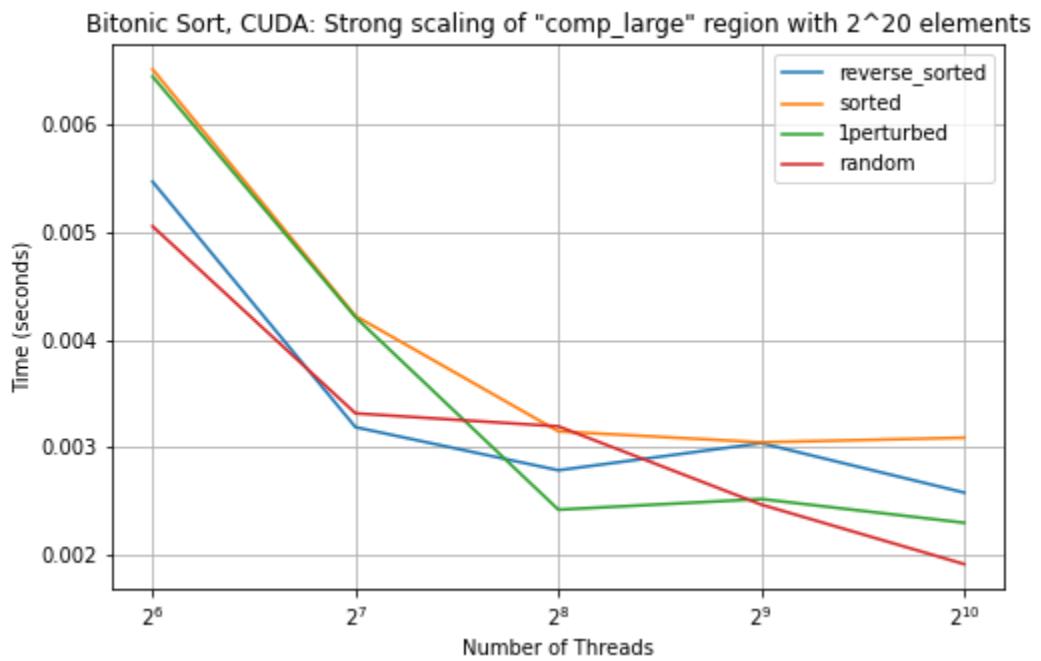
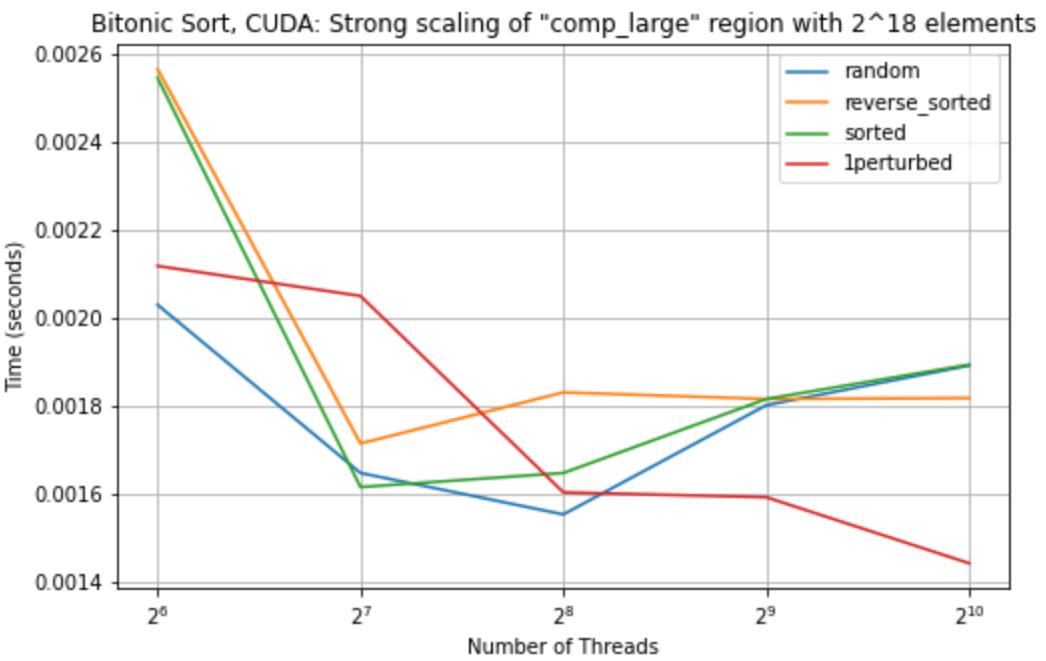
Bitonic Sort, CUDA: Strong scaling of "comm" region with 2^{24} elements

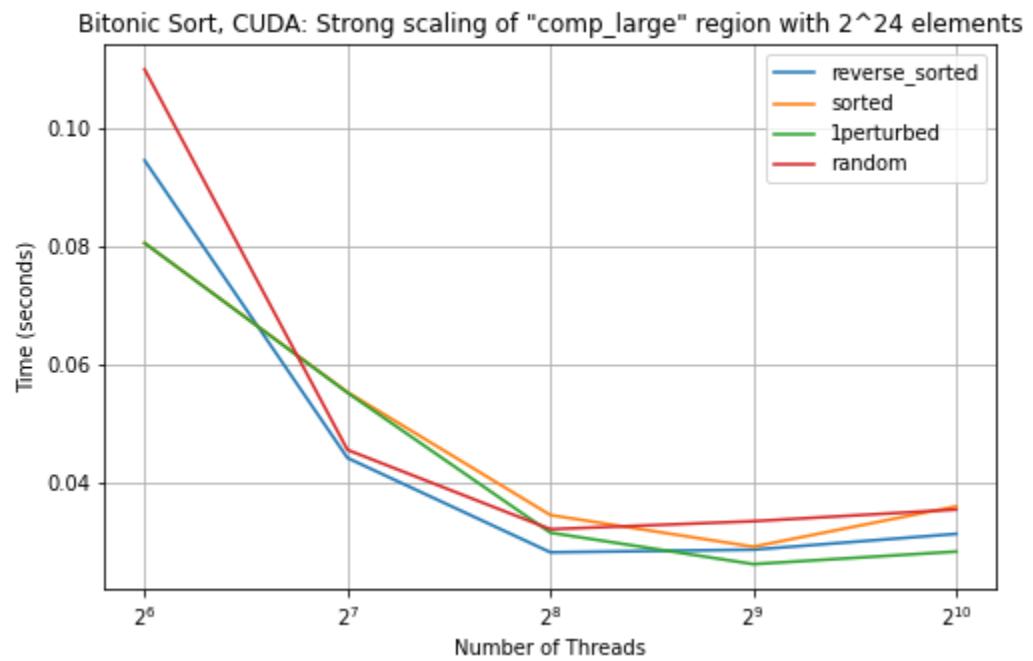
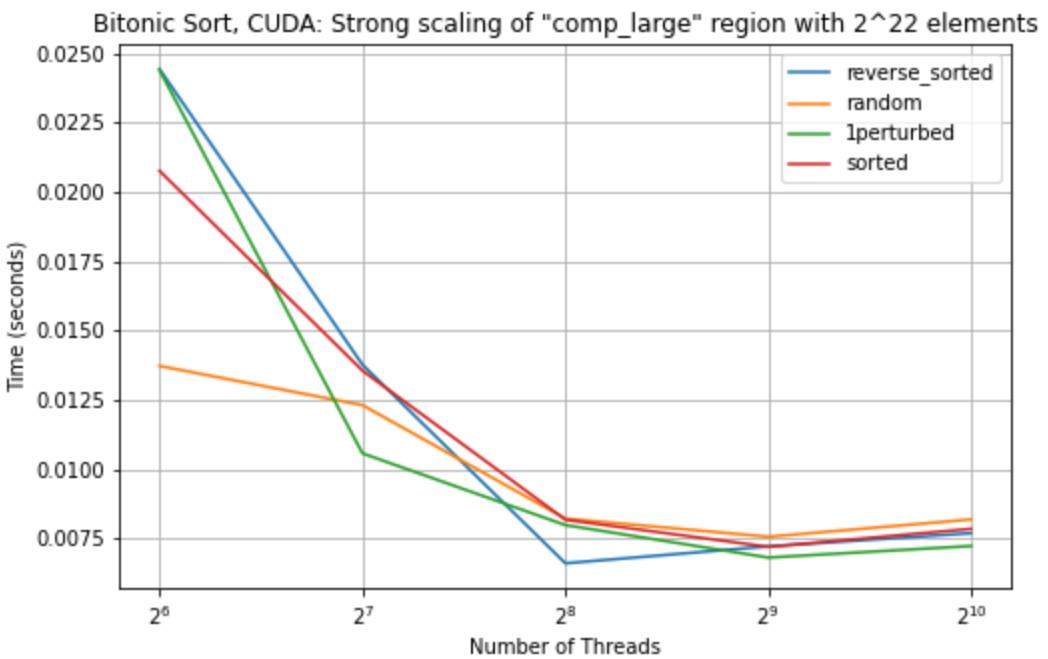


Bitonic Sort, CUDA: Strong scaling of "comm" region with 2^{26} elements

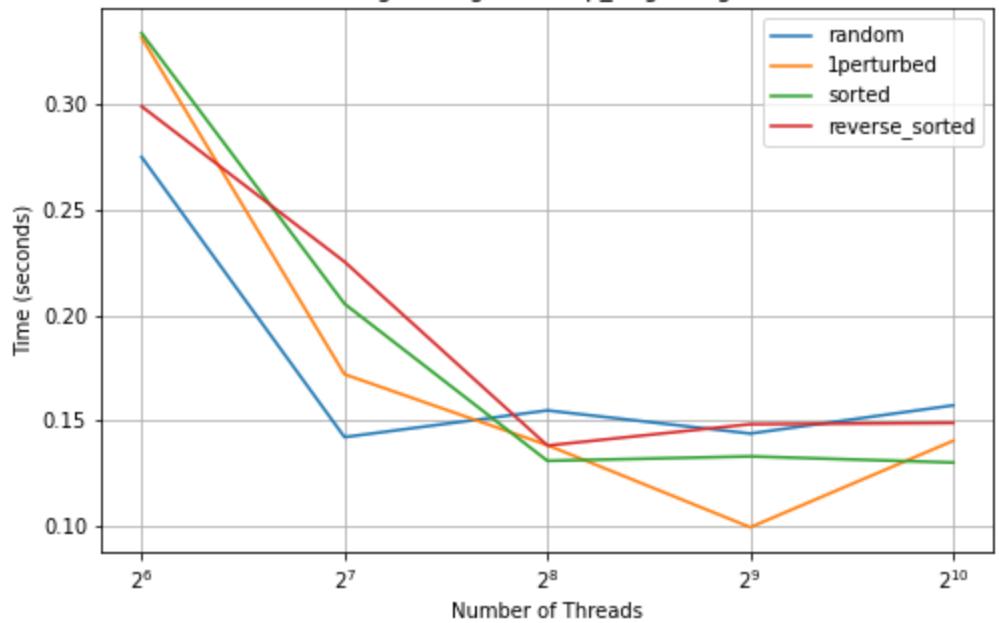




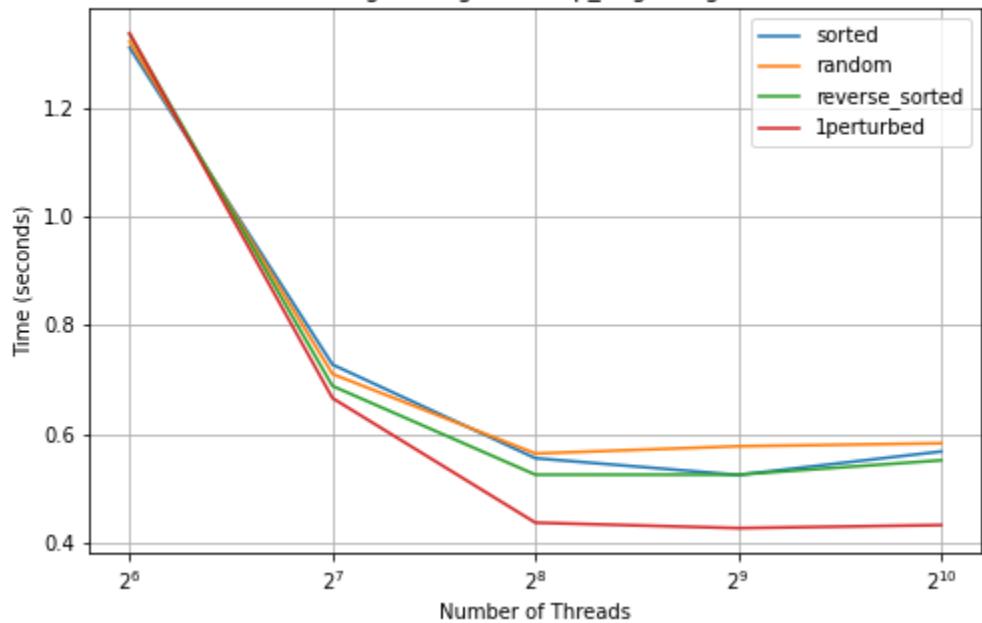




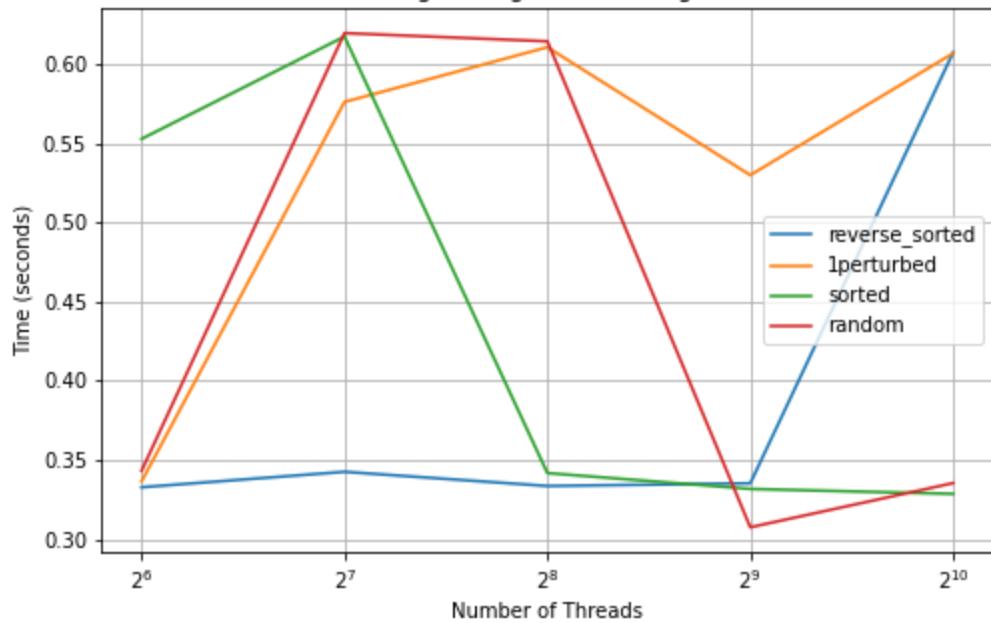
Bitonic Sort, CUDA: Strong scaling of "comp_large" region with 2^{26} elements



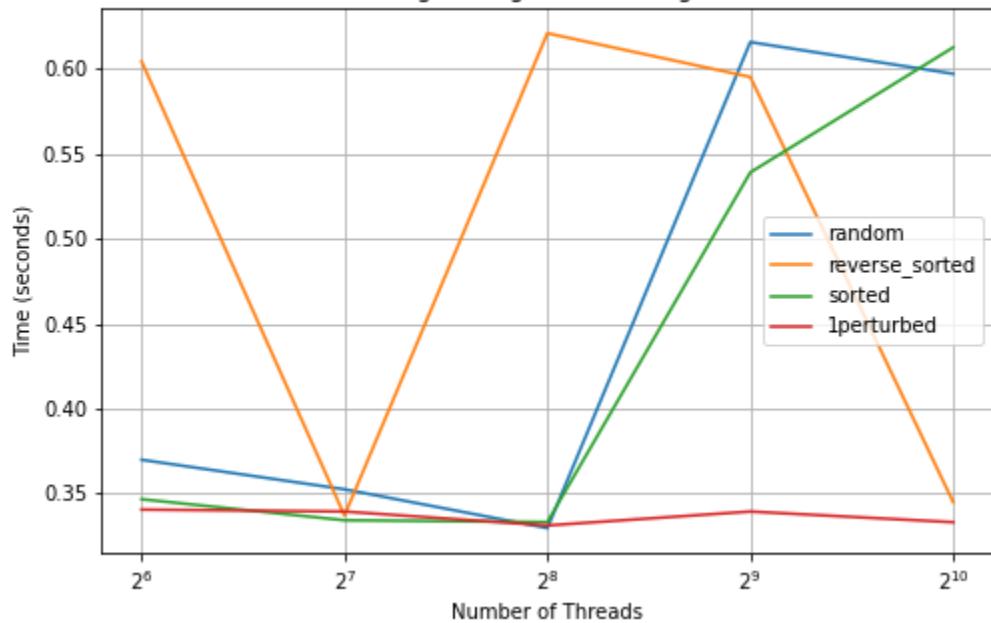
Bitonic Sort, CUDA: Strong scaling of "comp_large" region with 2^{28} elements



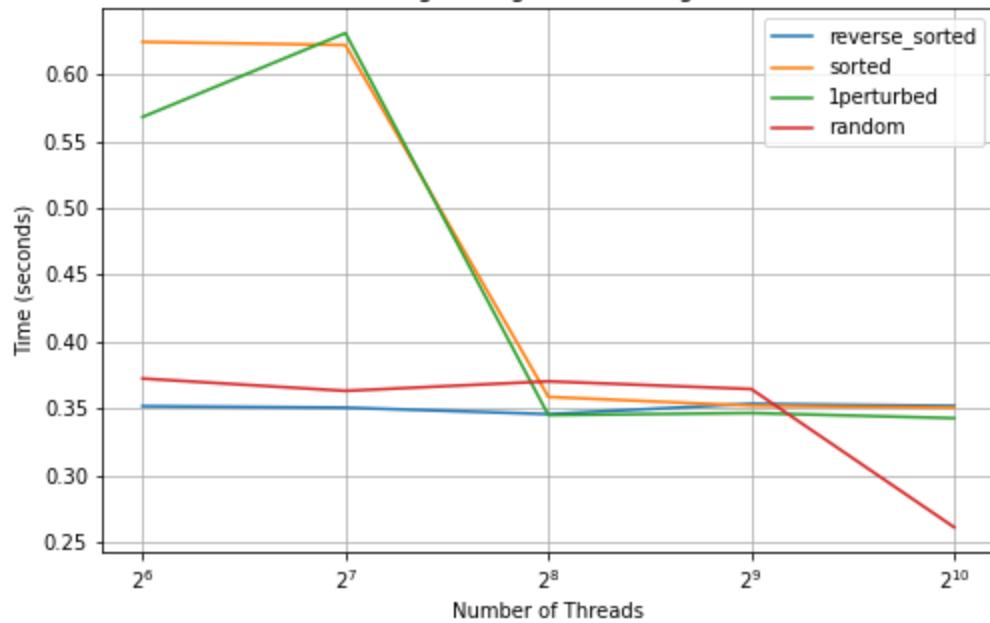
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{16} elements



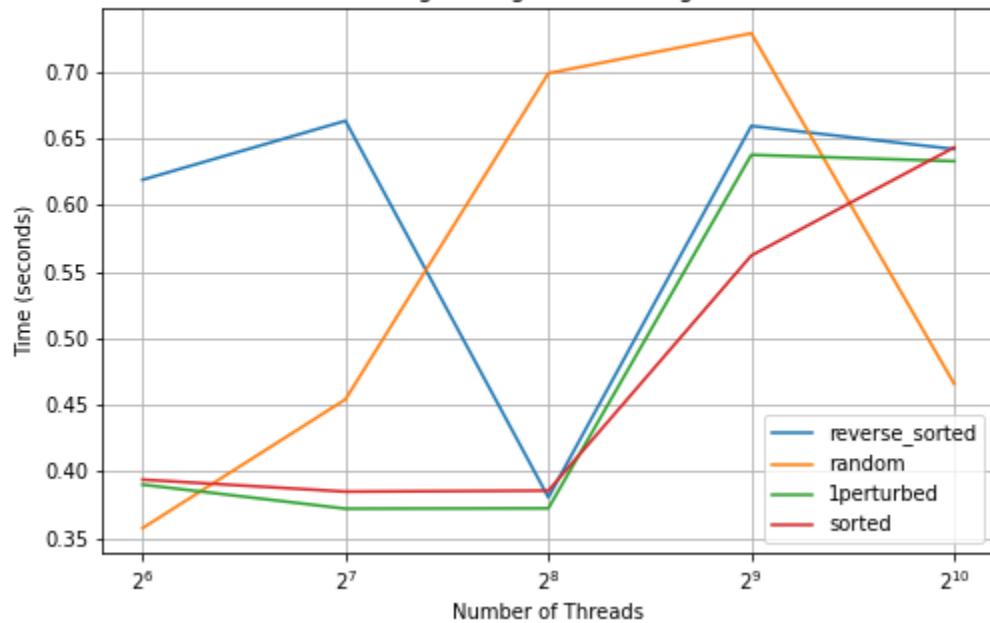
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{18} elements



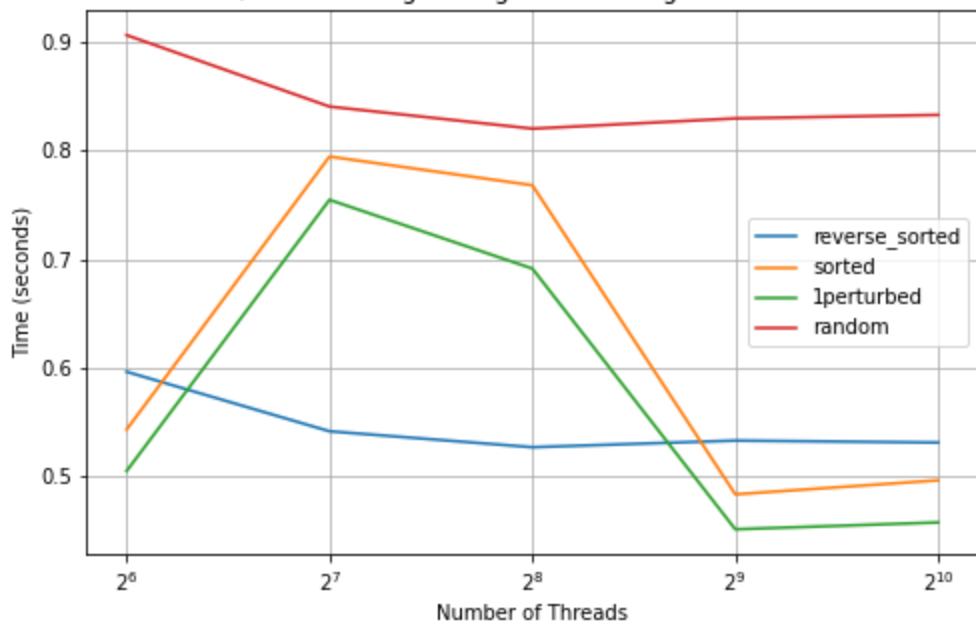
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{20} elements



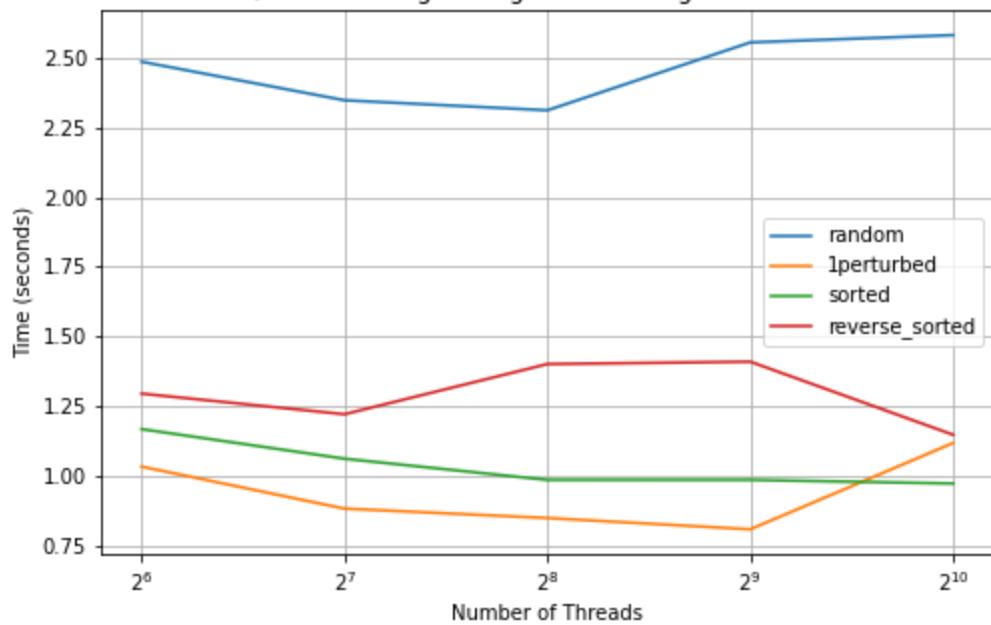
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{22} elements



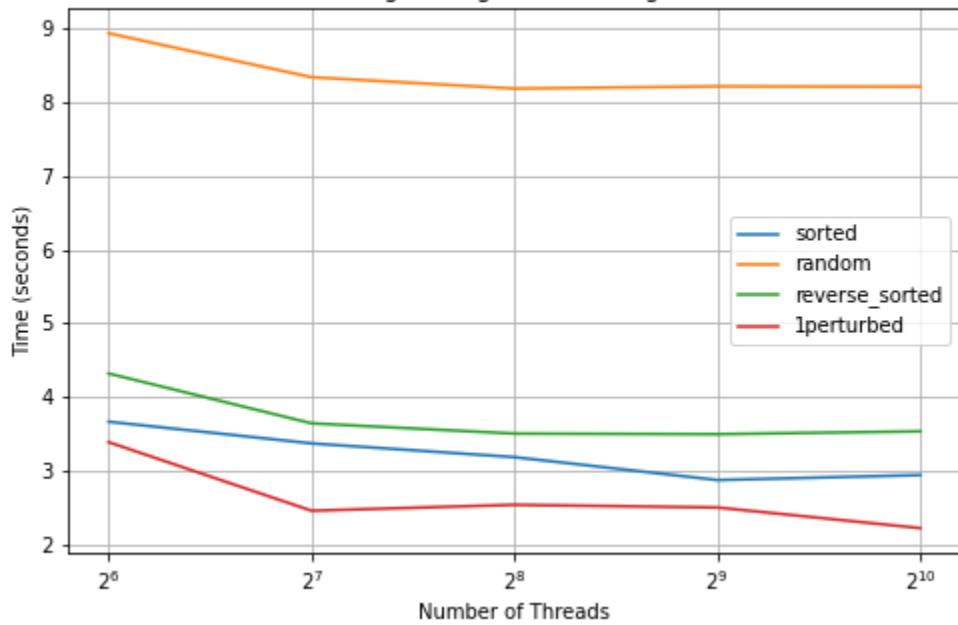
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{24} elements



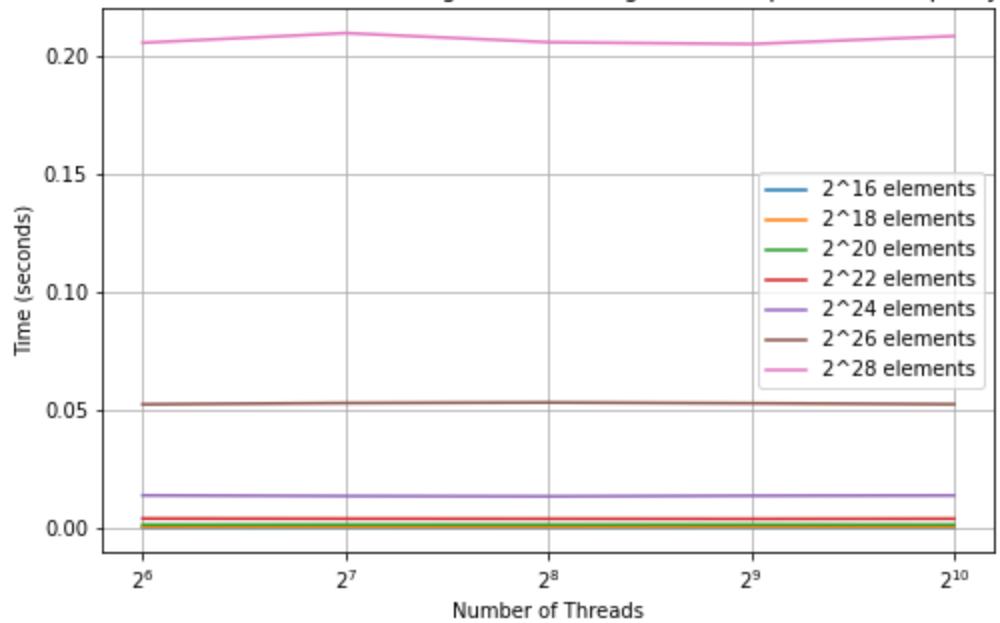
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{26} elements



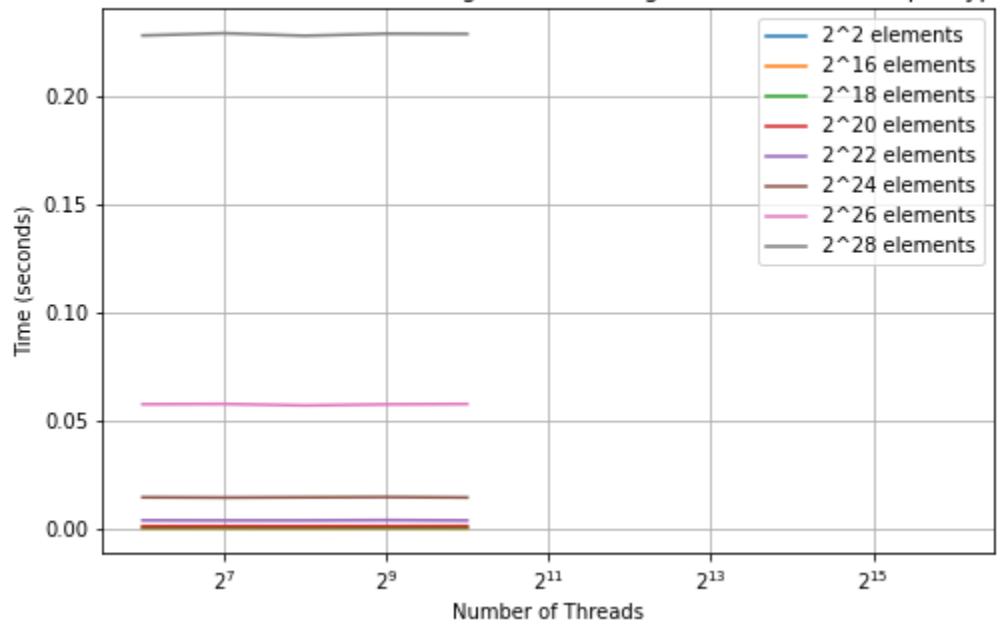
Bitonic Sort, CUDA: Strong scaling of "main" region with 2^{28} elements



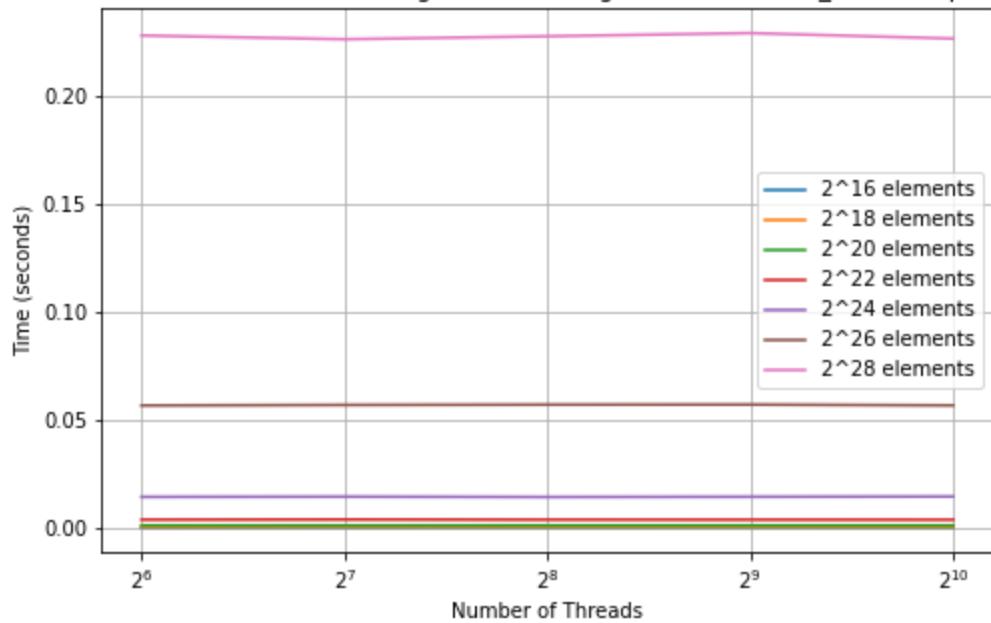
BitonicSort, CUDA: Weak scaling of "comm" region with "1perturbed" input type



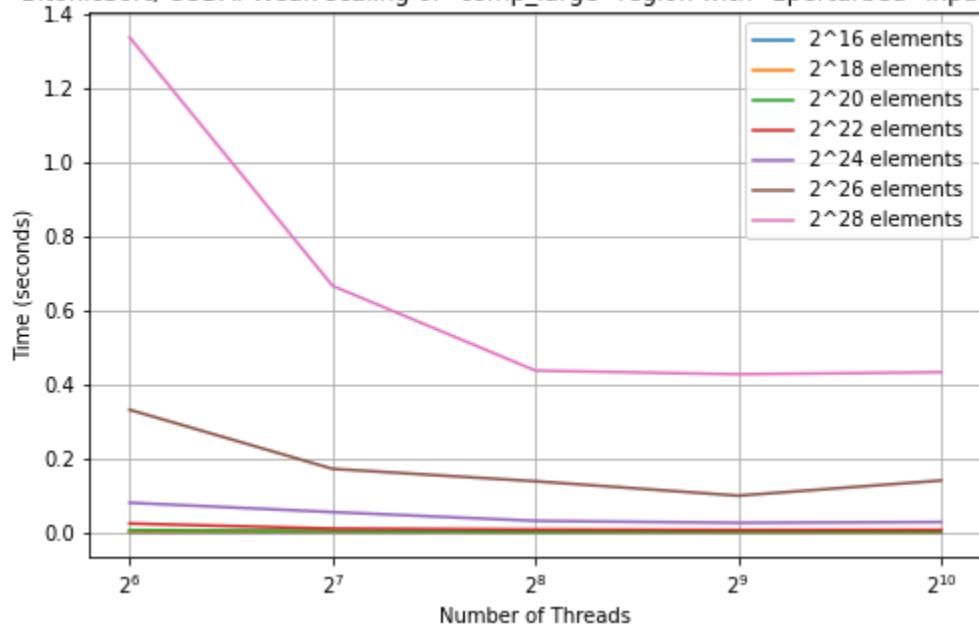
BitonicSort, CUDA: Weak scaling of "comm" region with "random" input type



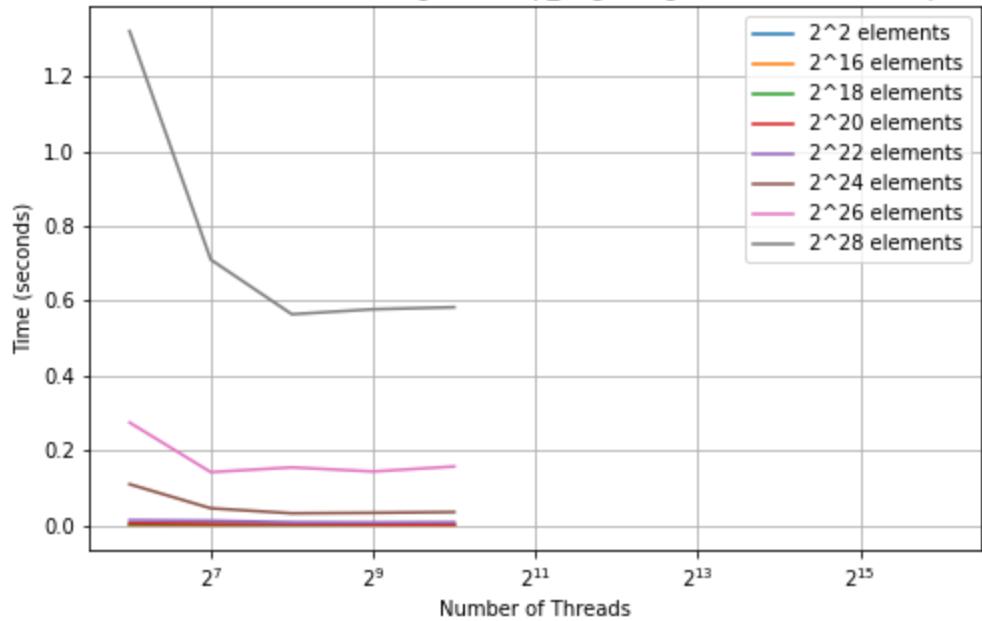
BitonicSort, CUDA: Weak scaling of "comm" region with "reverse_sorted" input type



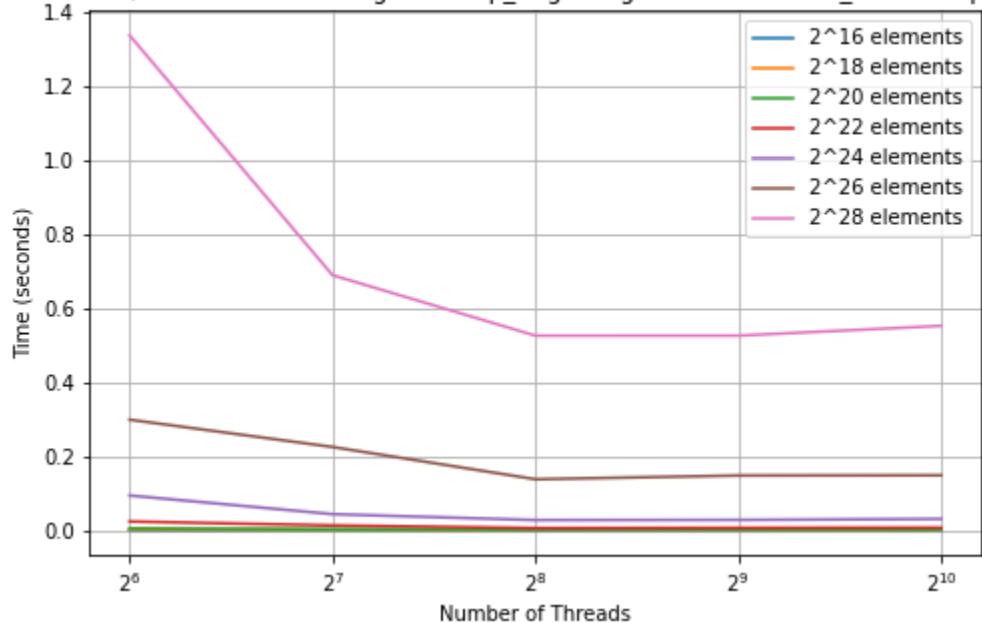
BitonicSort, CUDA: Weak scaling of "comp_large" region with "1perturbed" input type



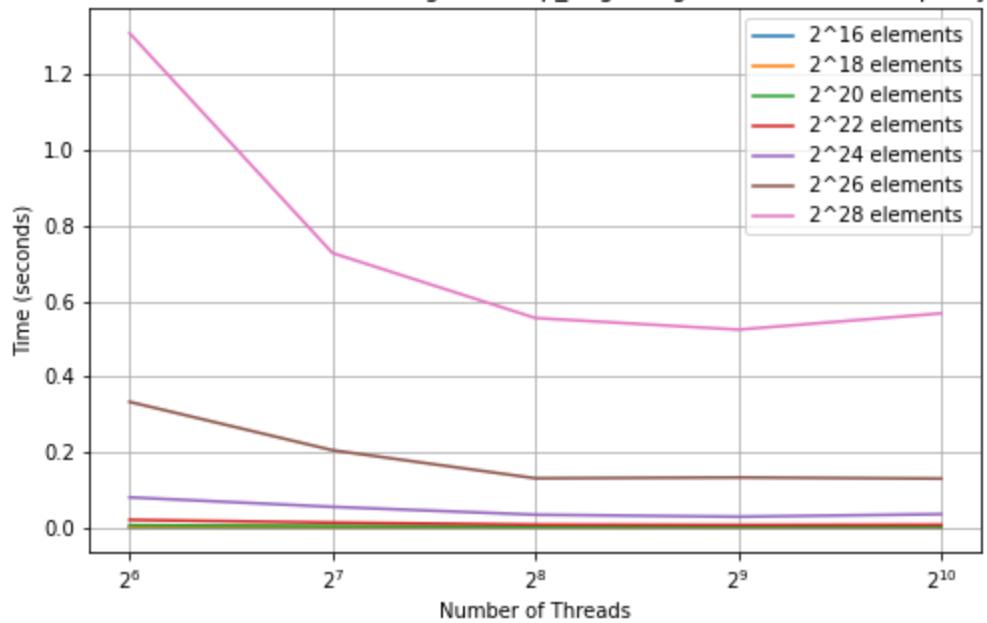
BitonicSort, CUDA: Weak scaling of "comp_large" region with "random" input type



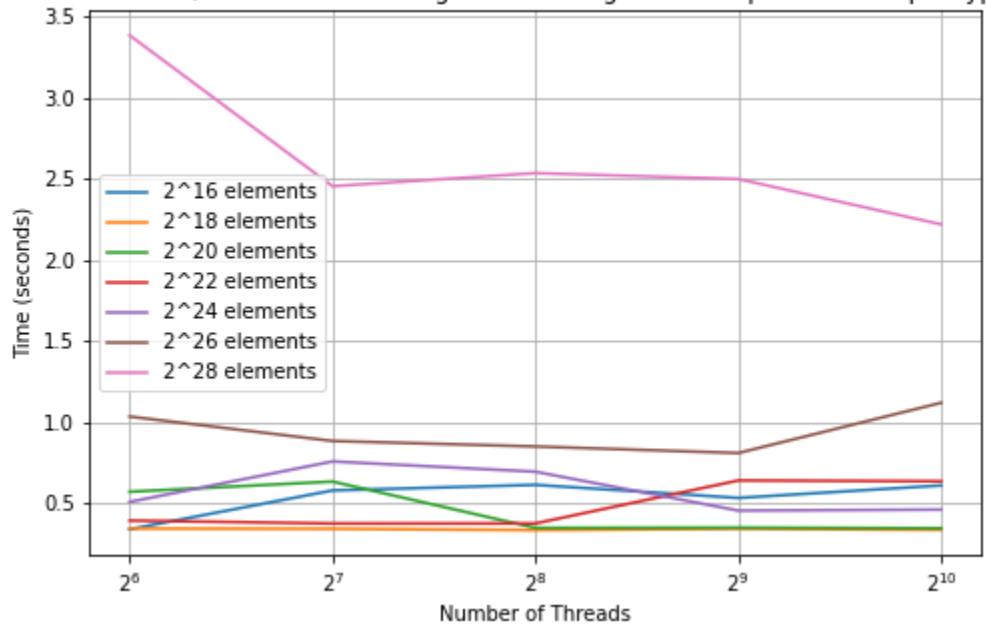
BitonicSort, CUDA: Weak scaling of "comp_large" region with "reverse_sorted" input type



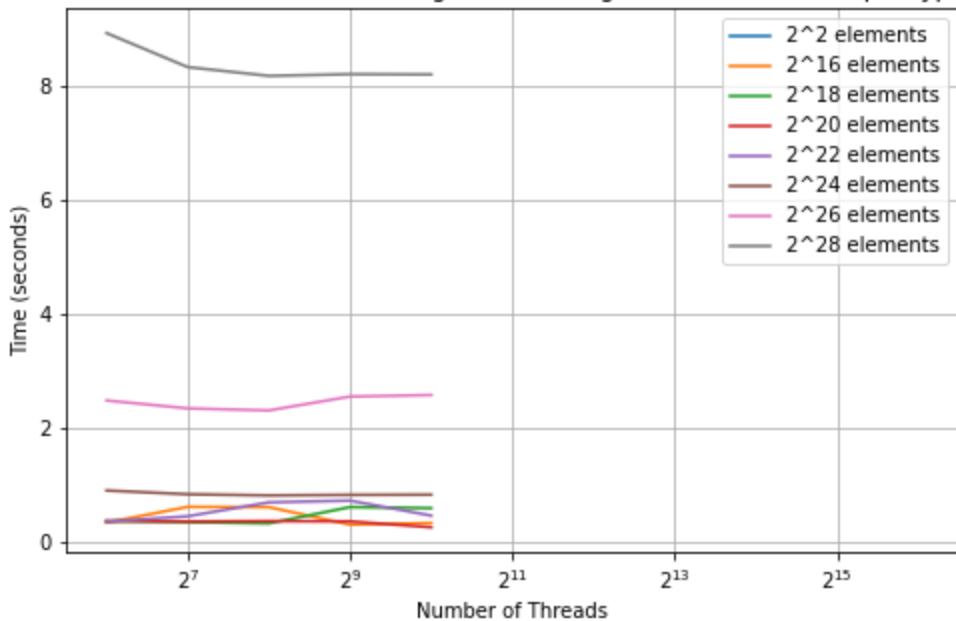
BitonicSort, CUDA: Weak scaling of "comp_large" region with "sorted" input type



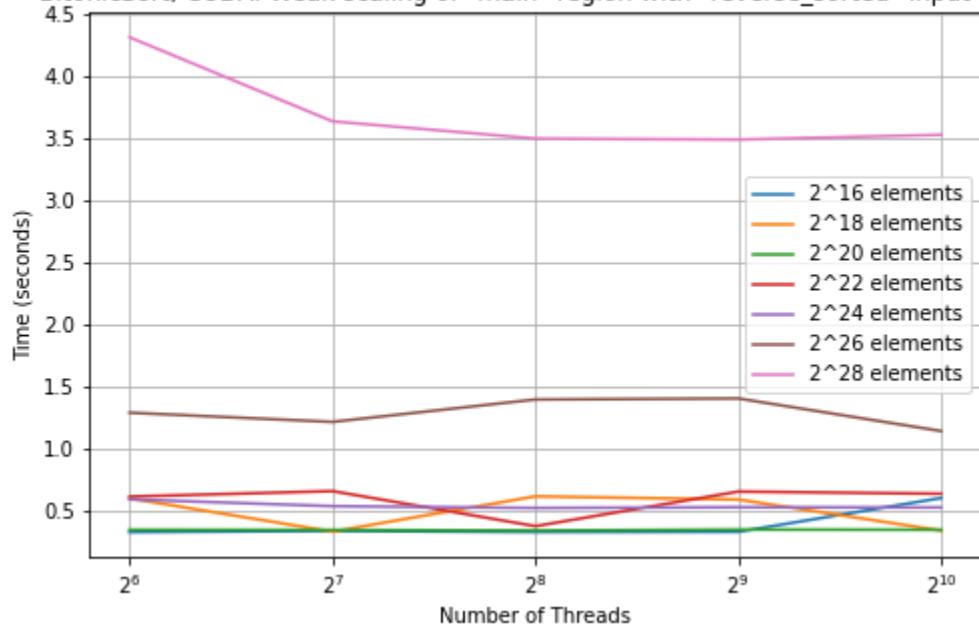
BitonicSort, CUDA: Weak scaling of "main" region with "1perturbed" input type



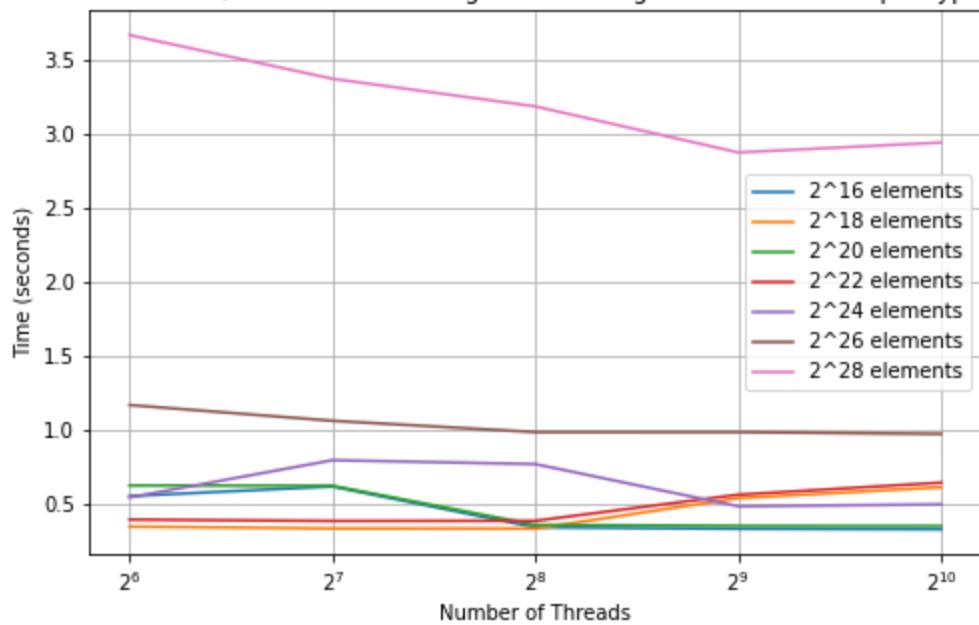
BitonicSort, CUDA: Weak scaling of "main" region with "random" input type



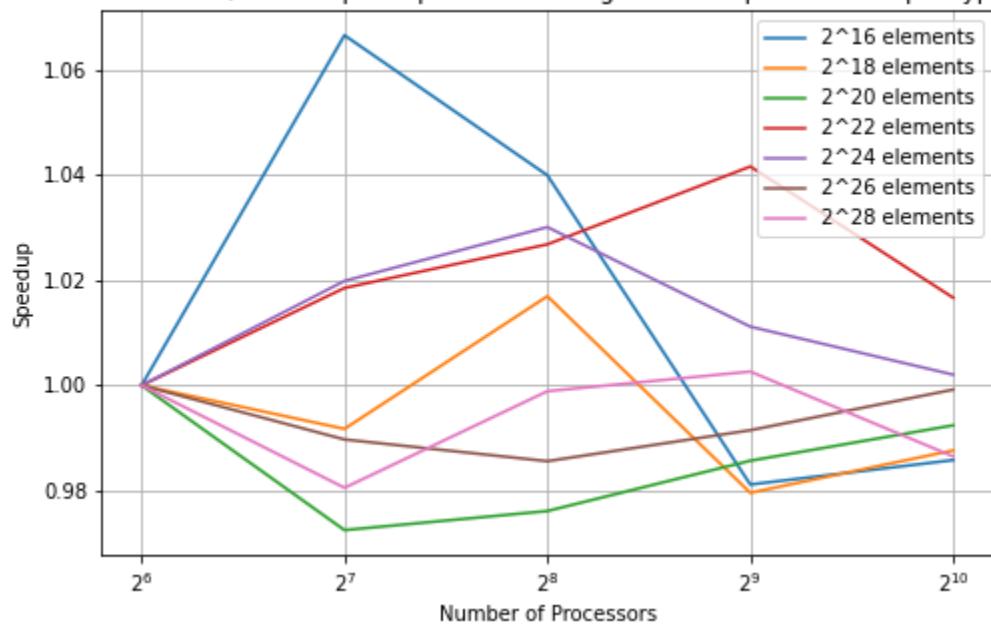
BitonicSort, CUDA: Weak scaling of "main" region with "reverse_sorted" input type



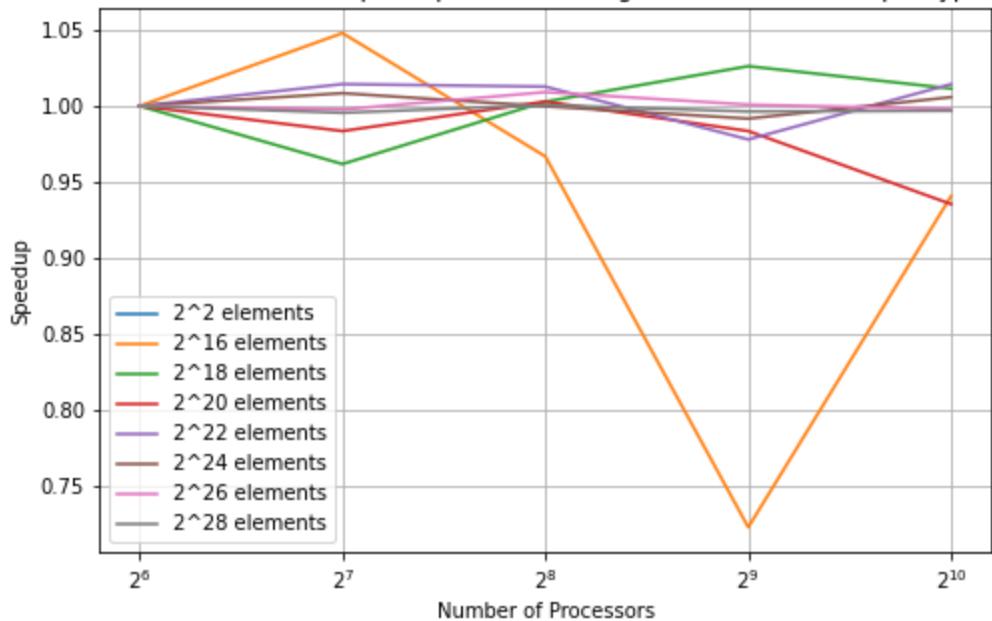
BitonicSort, CUDA: Weak scaling of "main" region with "sorted" input type



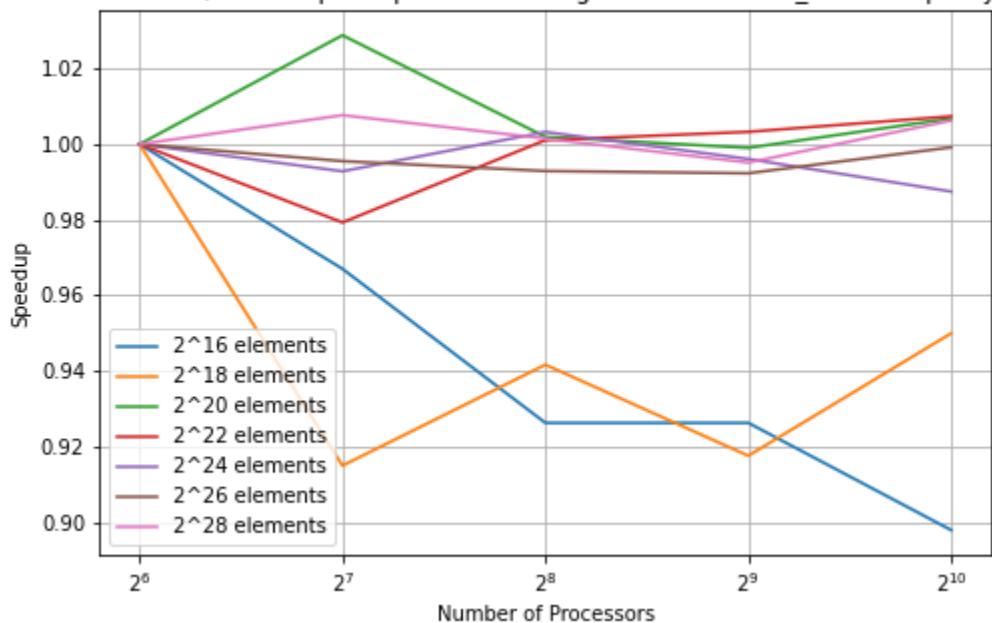
Bitonic Sort, CUDA: Speedup of "comm" region with "1perturbed" input type

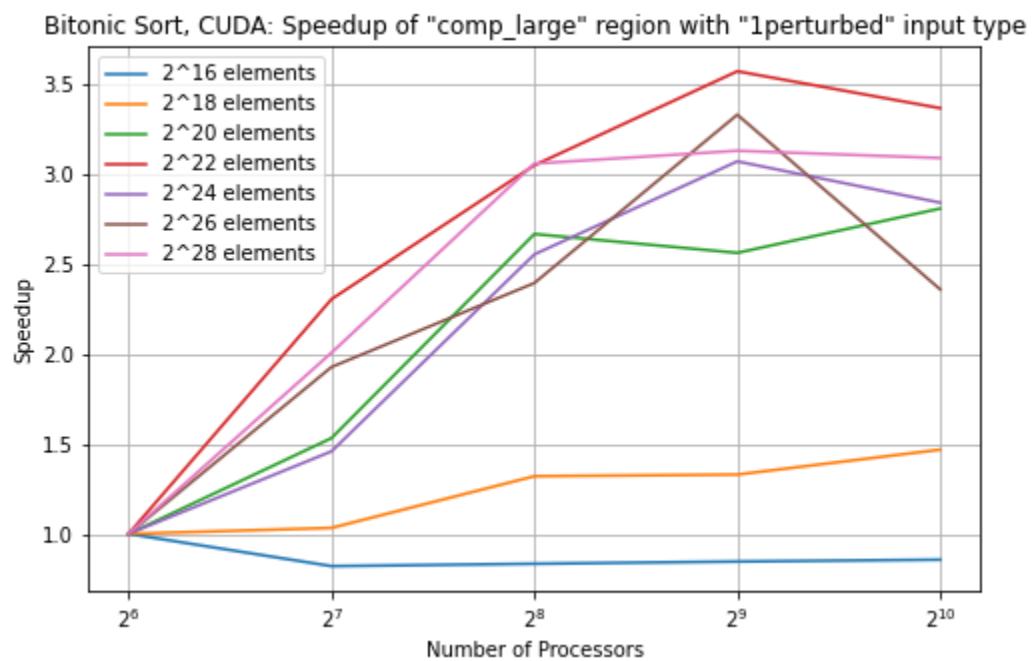
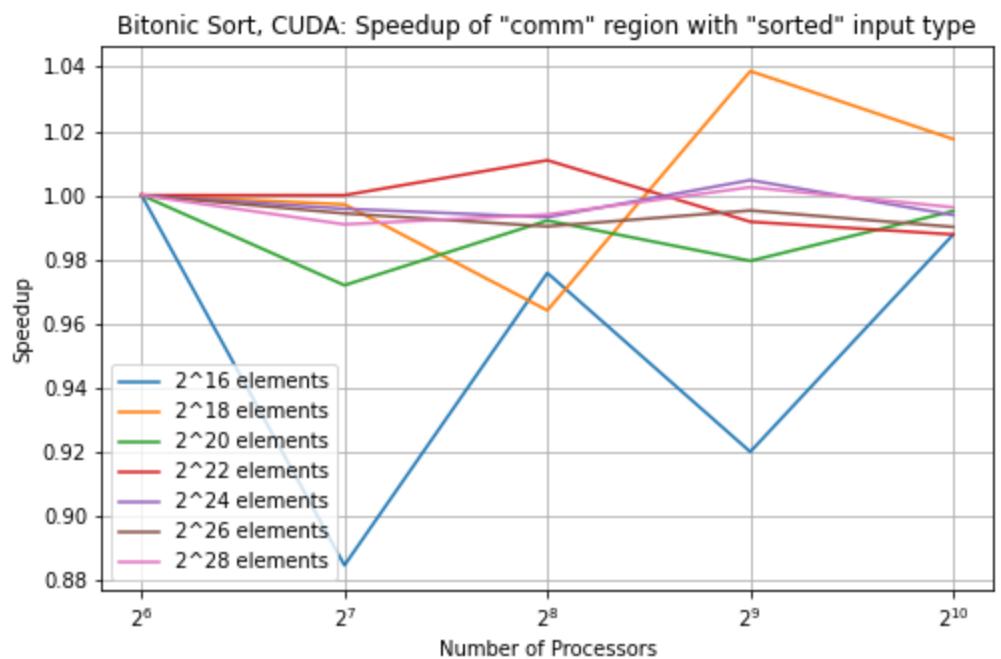


Bitonic Sort, CUDA: Speedup of "comm" region with "random" input type

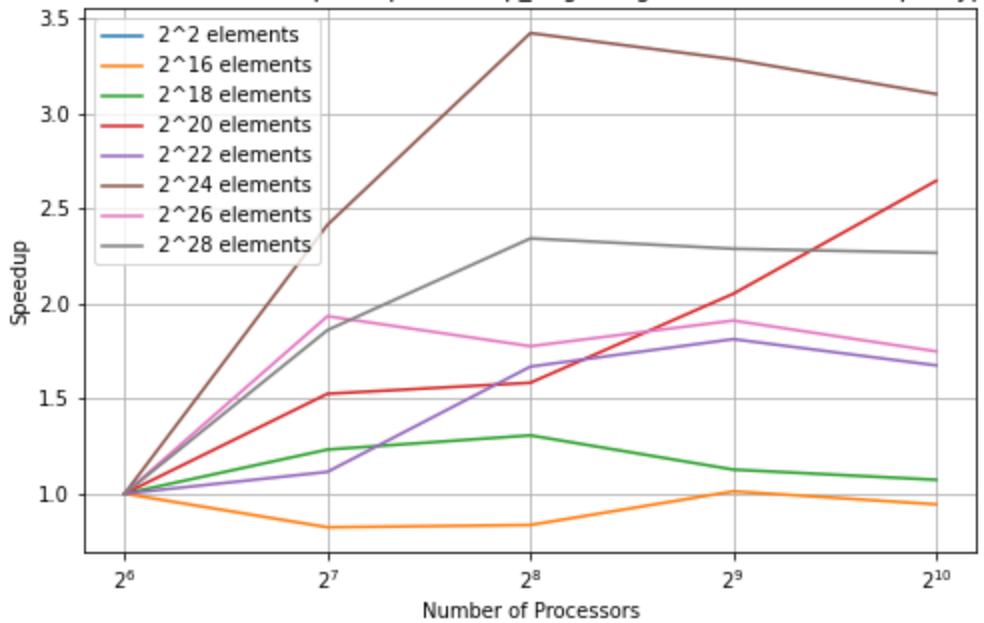


Bitonic Sort, CUDA: Speedup of "comm" region with "reverse_sorted" input type

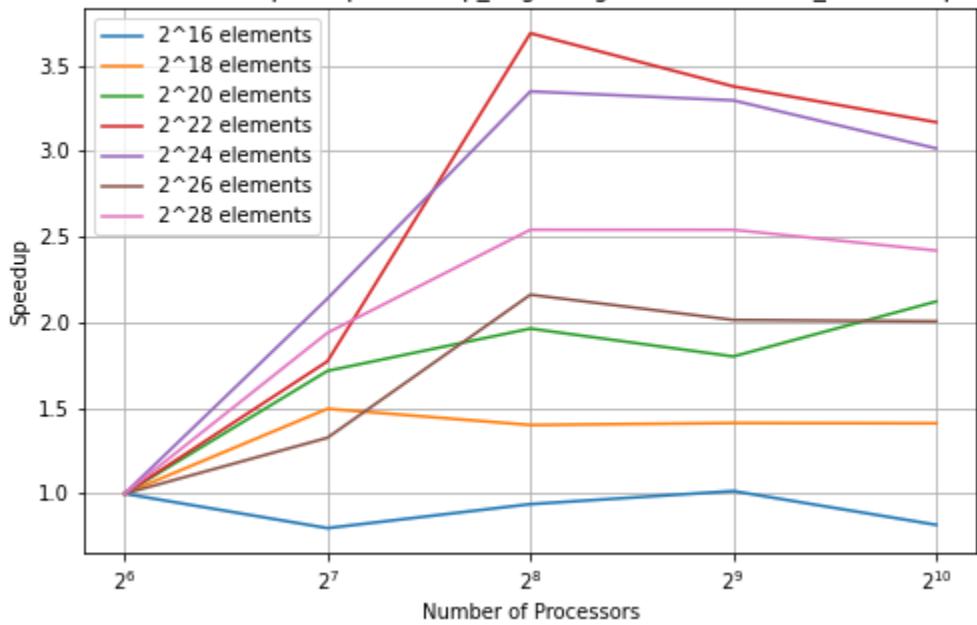




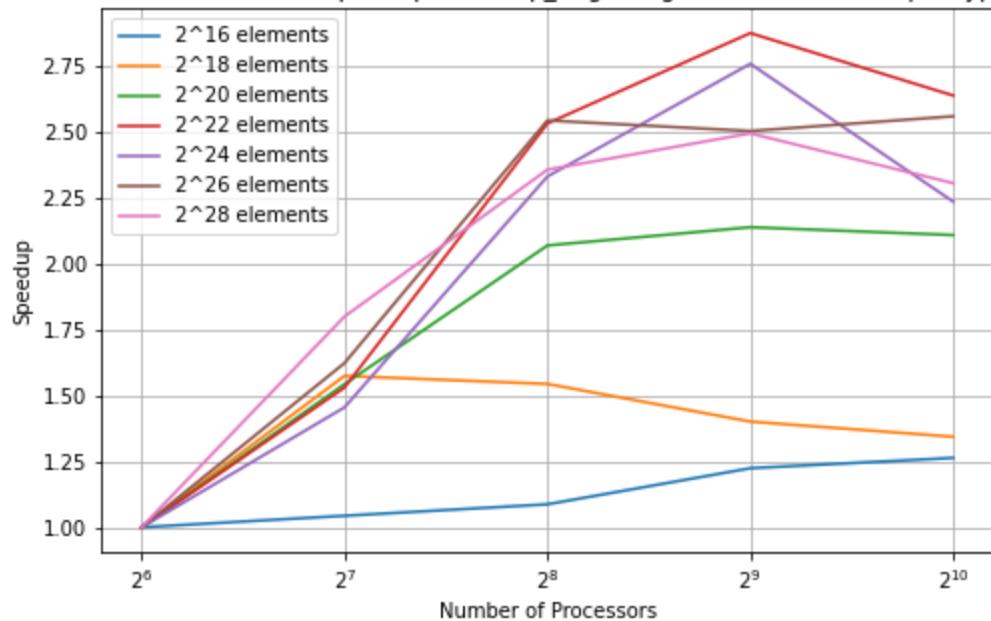
Bitonic Sort, CUDA: Speedup of "comp_large" region with "random" input type



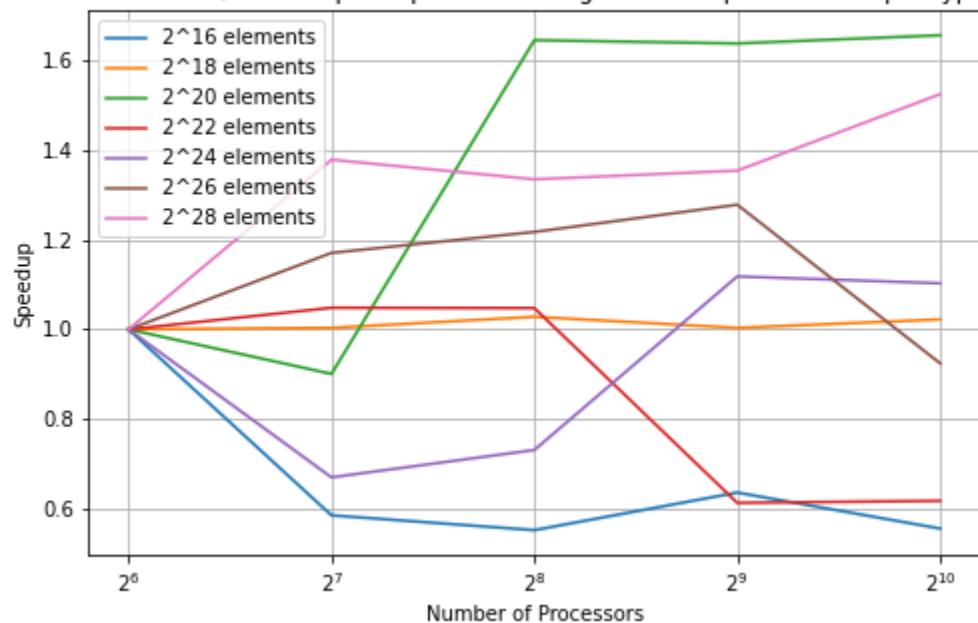
Bitonic Sort, CUDA: Speedup of "comp_large" region with "reverse_sorted" input type



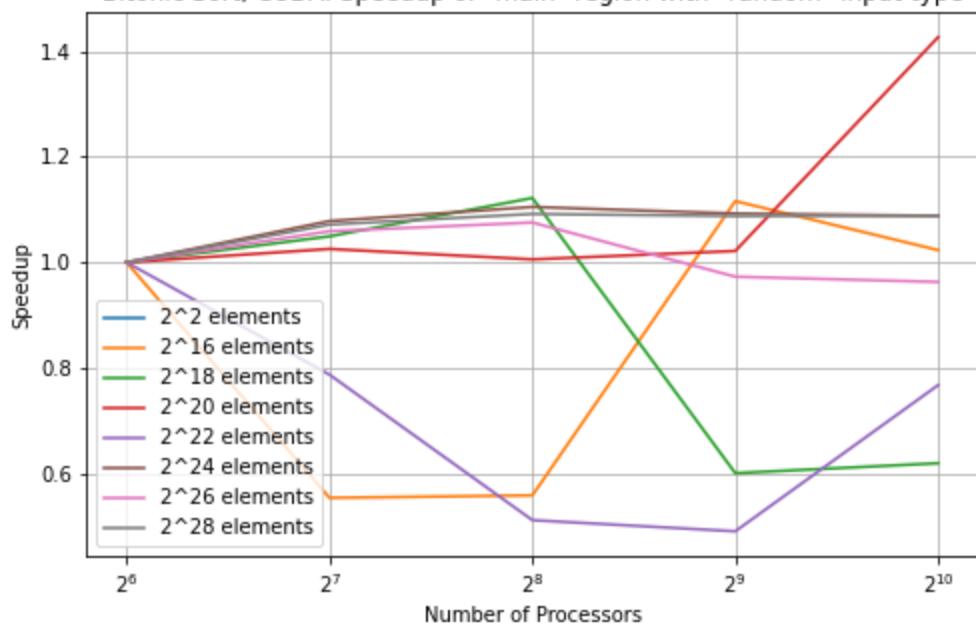
Bitonic Sort, CUDA: Speedup of "comp_large" region with "sorted" input type



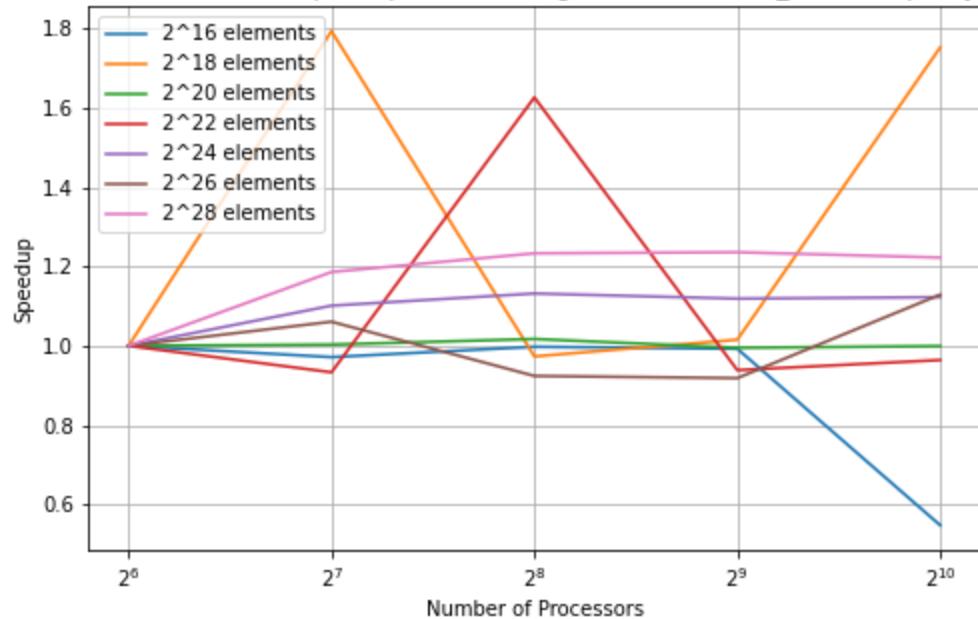
Bitonic Sort, CUDA: Speedup of "main" region with "1perturbed" input type



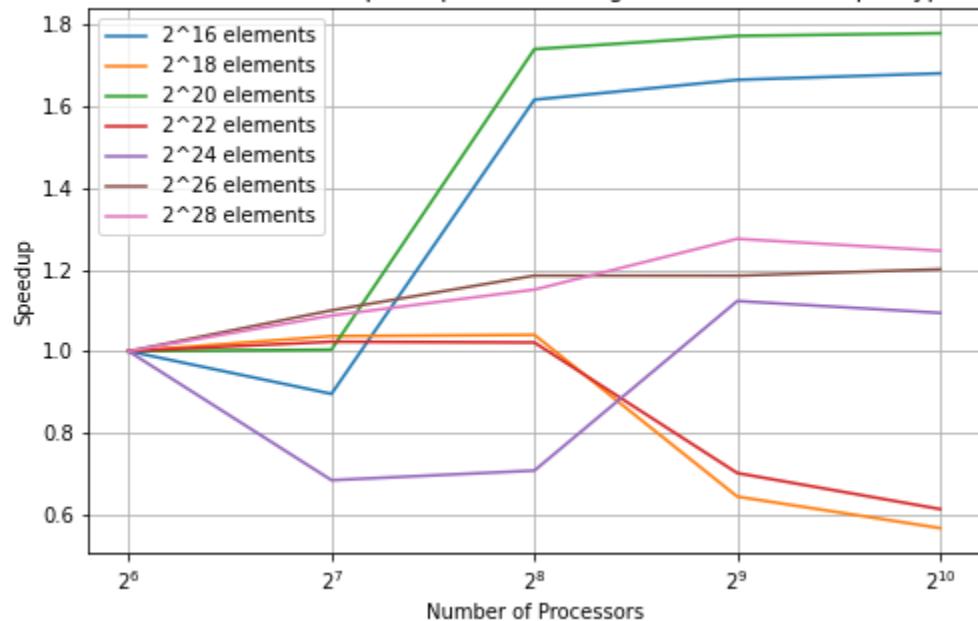
Bitonic Sort, CUDA: Speedup of "main" region with "random" input type



Bitonic Sort, CUDA: Speedup of "main" region with "reverse_sorted" input type



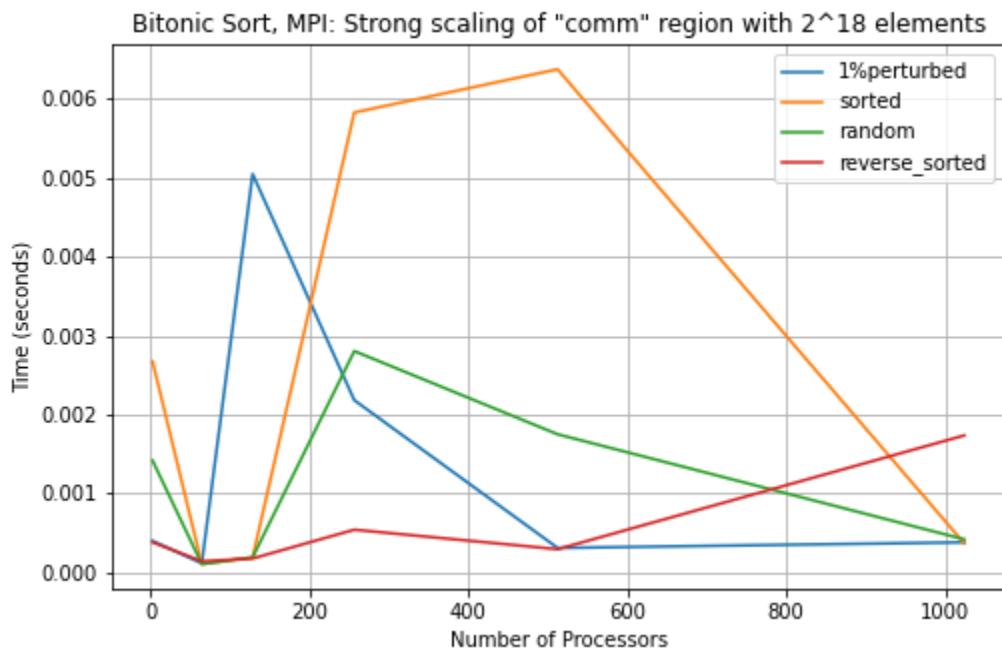
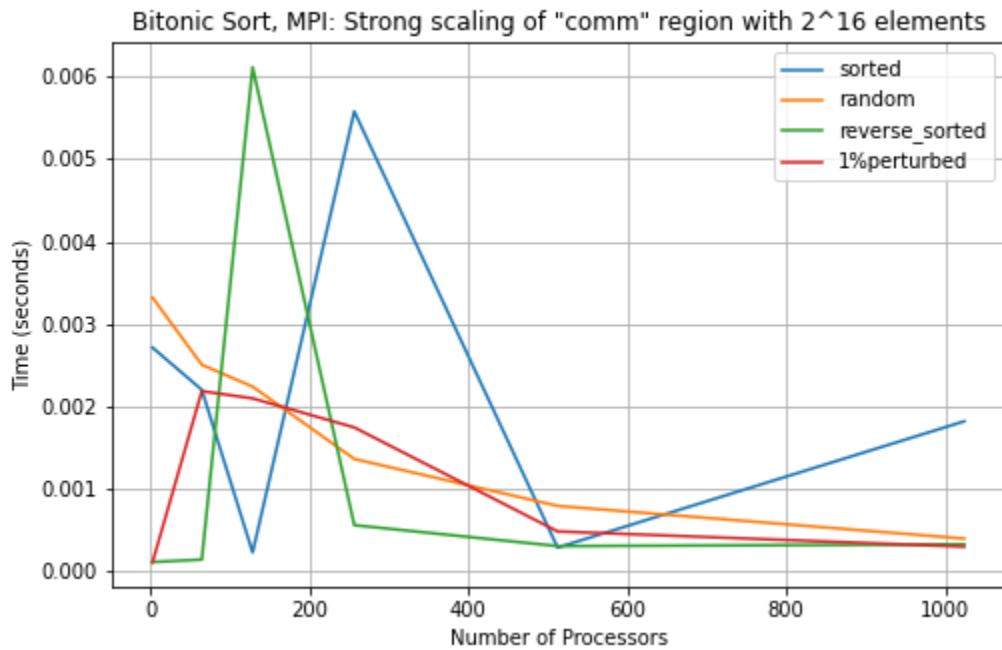
Bitonic Sort, CUDA: Speedup of "main" region with "sorted" input type

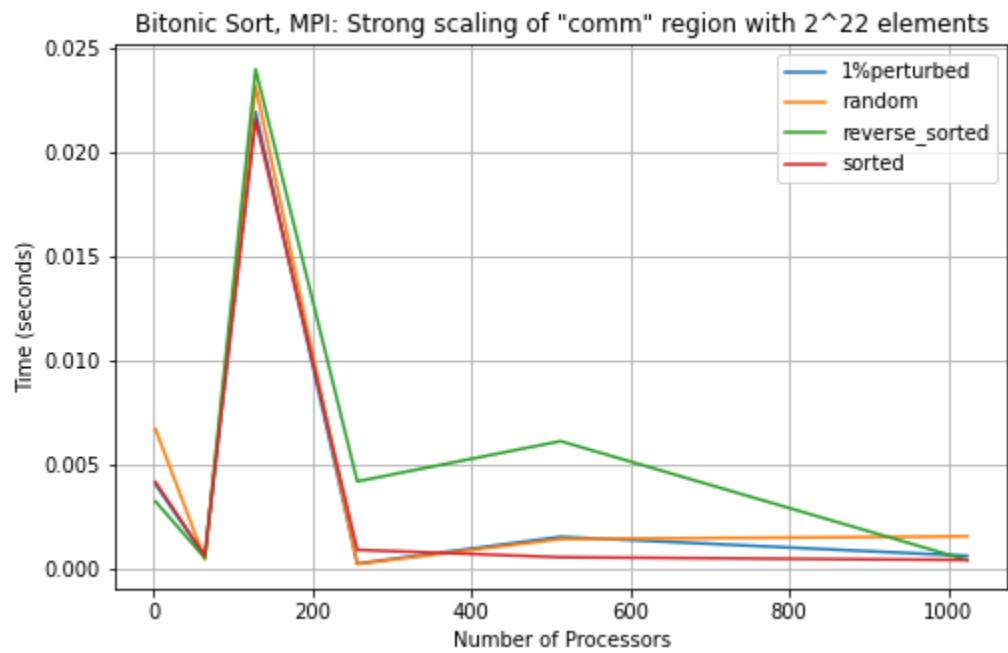
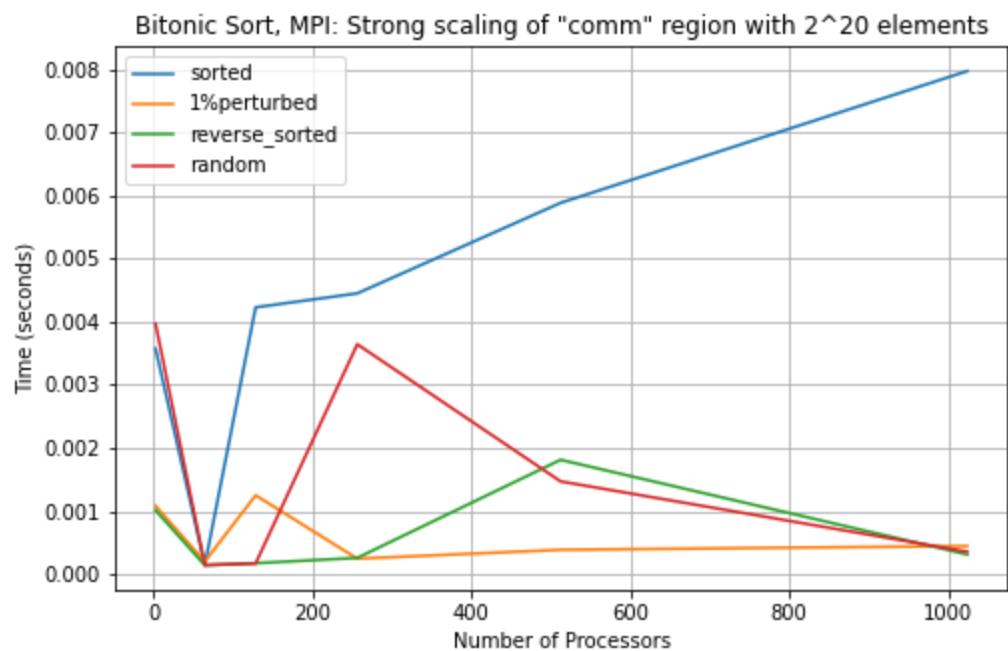


Throughout all of the implementations, we can expect an increase in the computational and communication time as the size of the array increase. We can also expect bitonic sort parallelism to reach a point of diminishing returns as we increase the number of threads. Additionally, as we change the different array populations such as random, sorted, reverse sorted, and 1% perturbed, we can expect slight differences in our

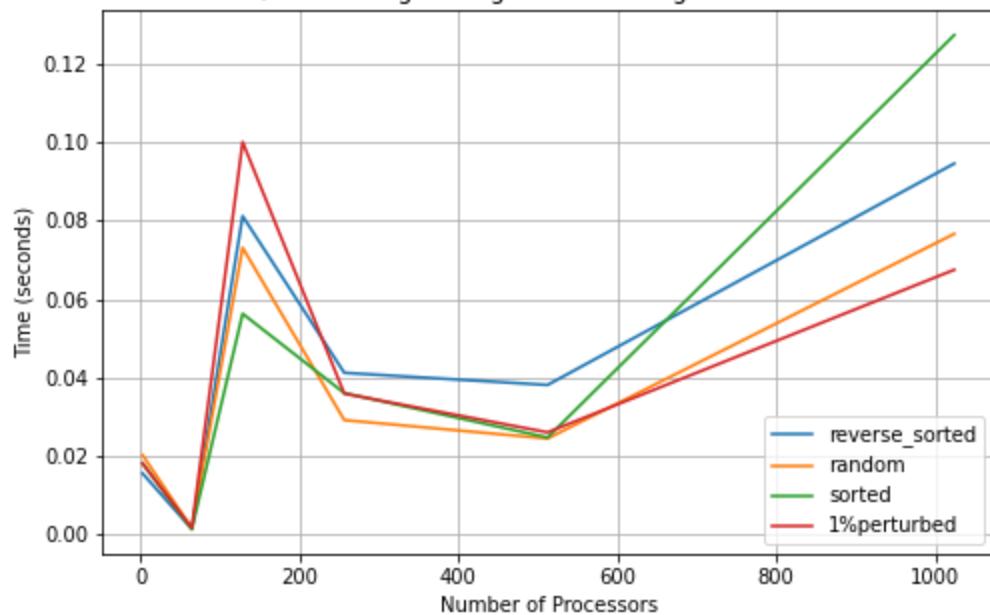
runtimes. For instance, a “sorted” array should have significantly less runtime than an unsorted array.

Bitonic Sort MPI

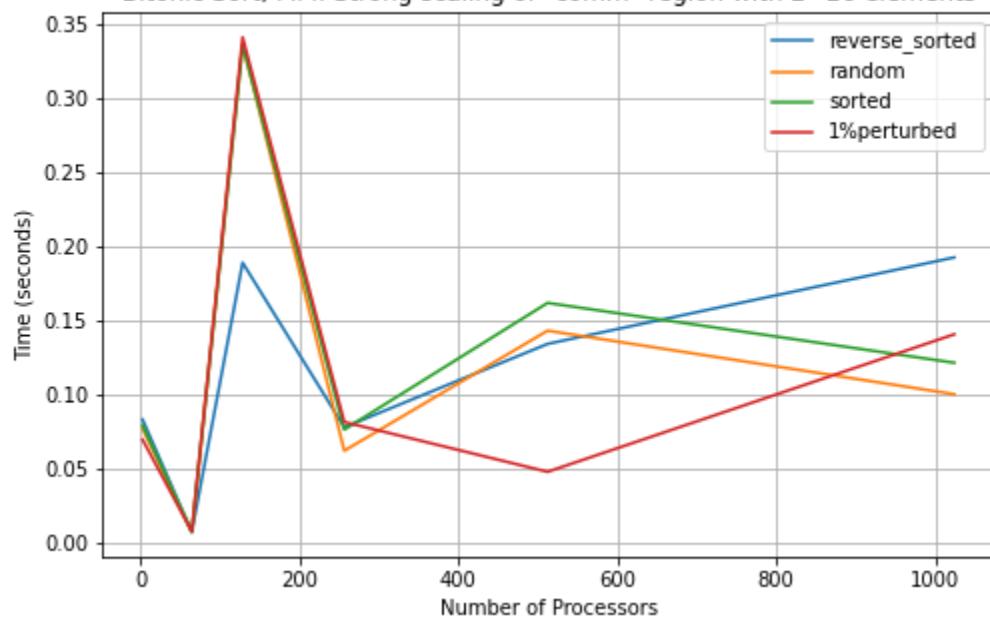


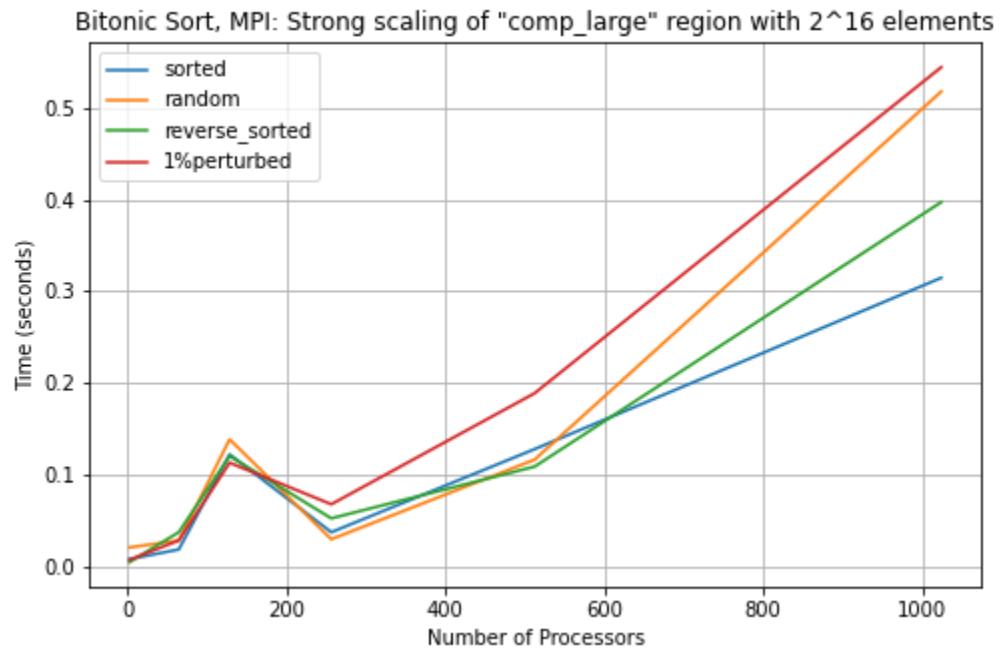
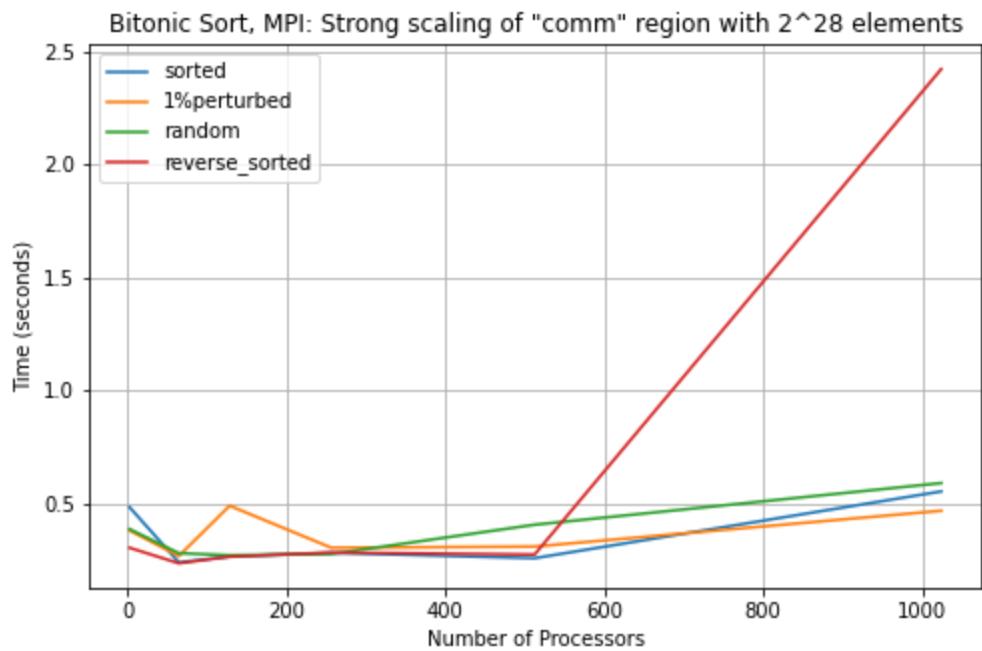


Bitonic Sort, MPI: Strong scaling of "comm" region with 2^{24} elements

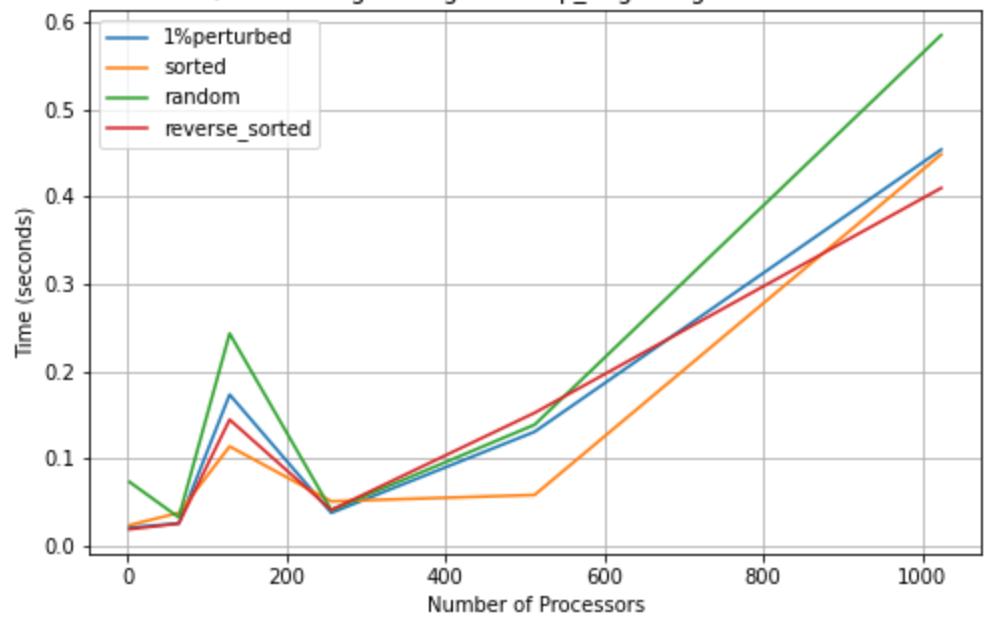


Bitonic Sort, MPI: Strong scaling of "comm" region with 2^{26} elements

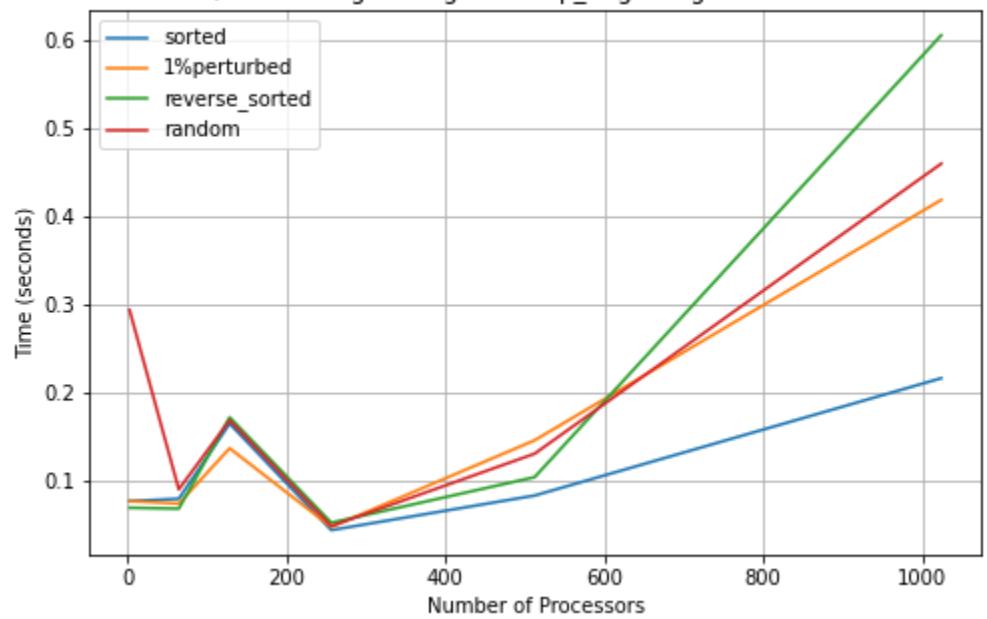




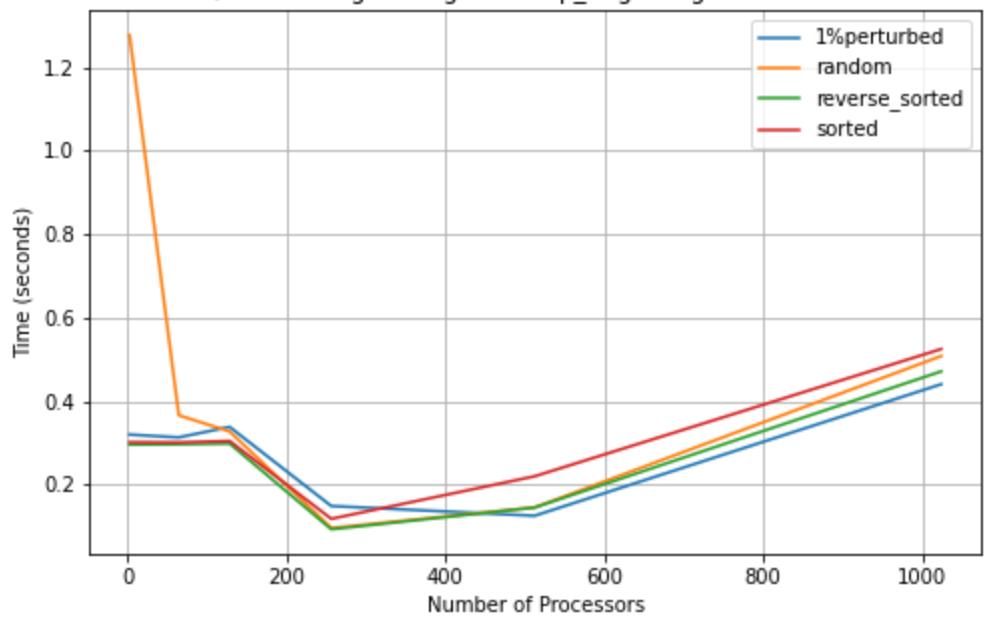
Bitonic Sort, MPI: Strong scaling of "comp_large" region with 2^{18} elements



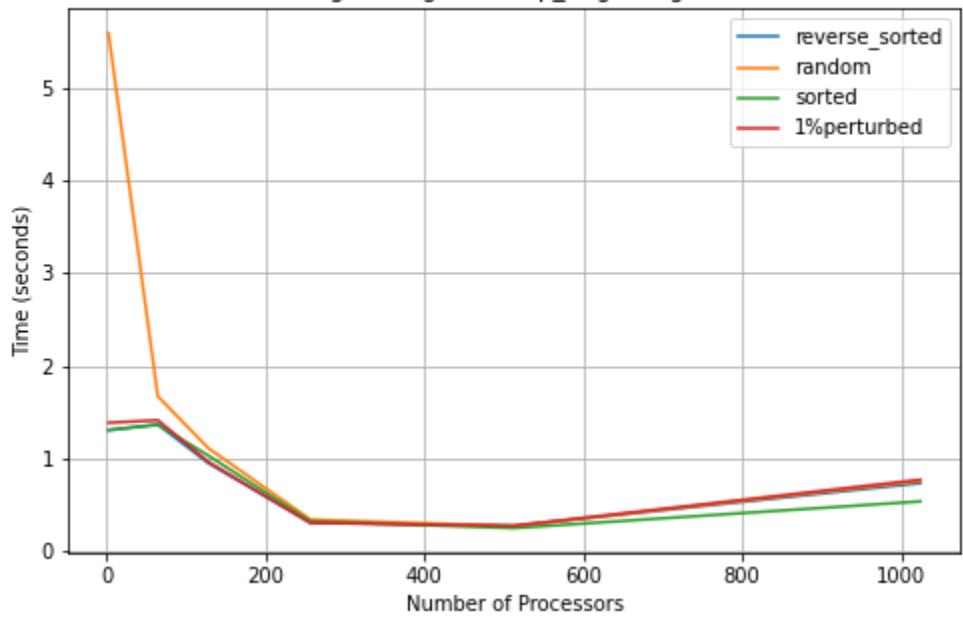
Bitonic Sort, MPI: Strong scaling of "comp_large" region with 2^{20} elements



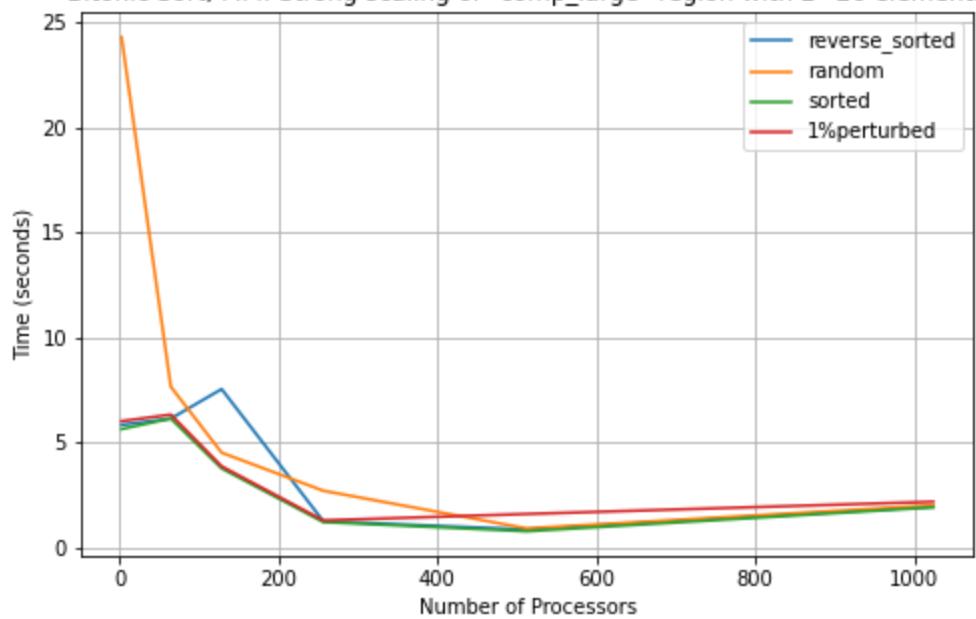
Bitonic Sort, MPI: Strong scaling of "comp_large" region with 2^{22} elements



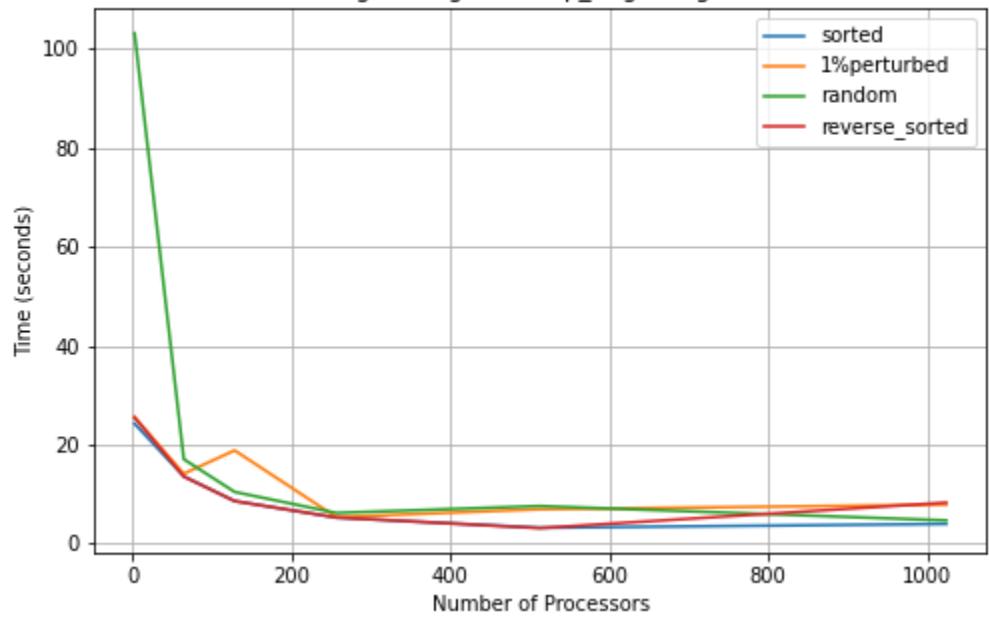
Bitonic Sort, MPI: Strong scaling of "comp_large" region with 2^{24} elements



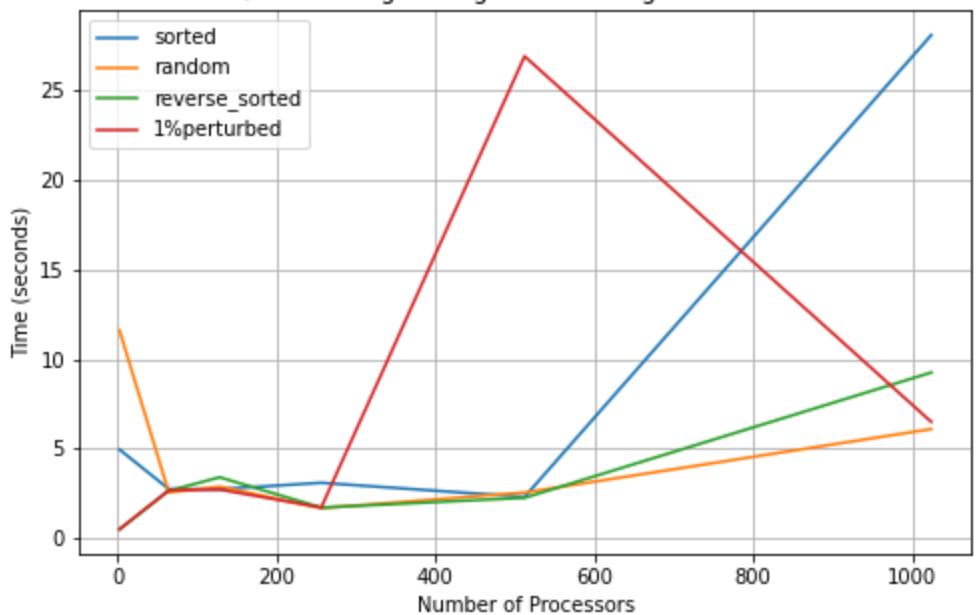
Bitonic Sort, MPI: Strong scaling of "comp_large" region with 2^{26} elements



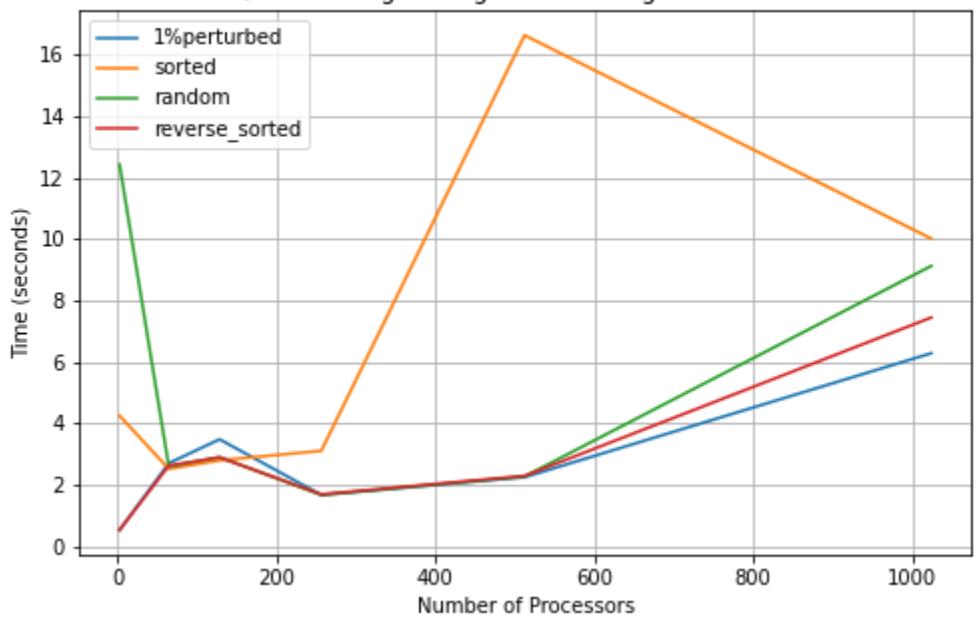
Bitonic Sort, MPI: Strong scaling of "comp_large" region with 2^{28} elements

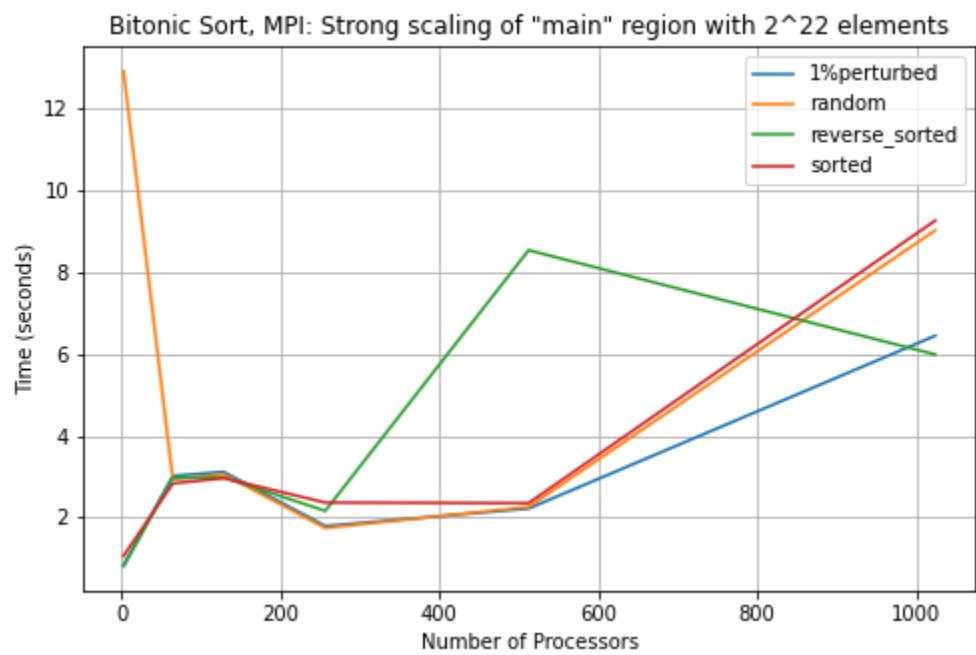
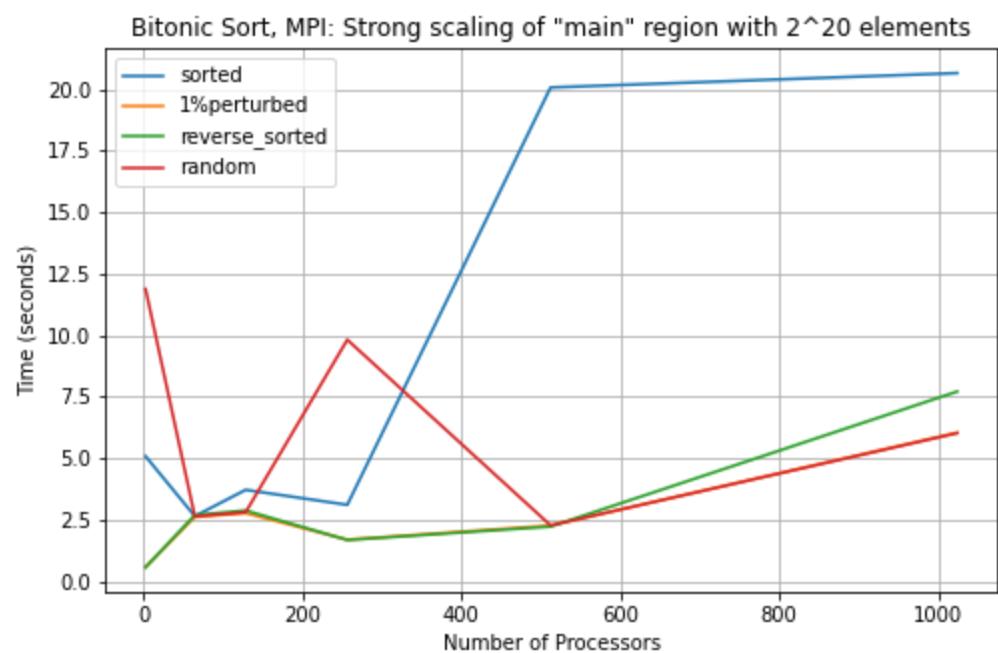


Bitonic Sort, MPI: Strong scaling of "main" region with 2^{16} elements

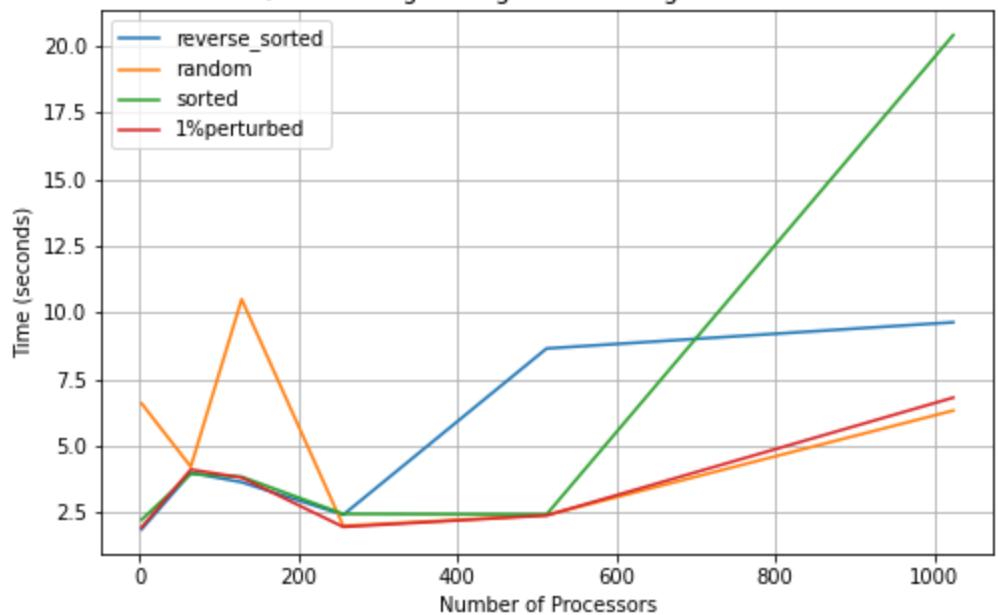


Bitonic Sort, MPI: Strong scaling of "main" region with 2^{18} elements

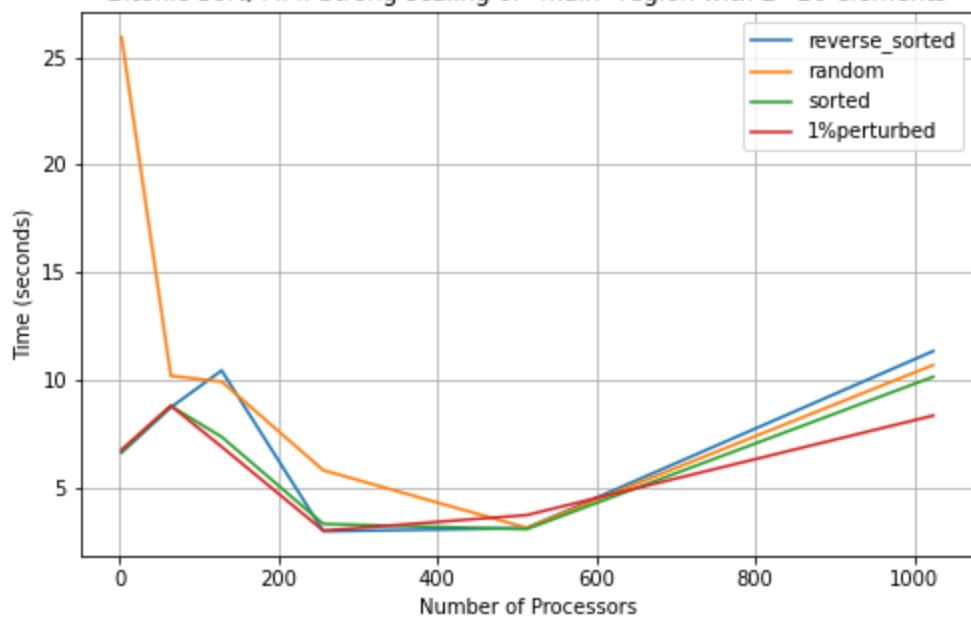




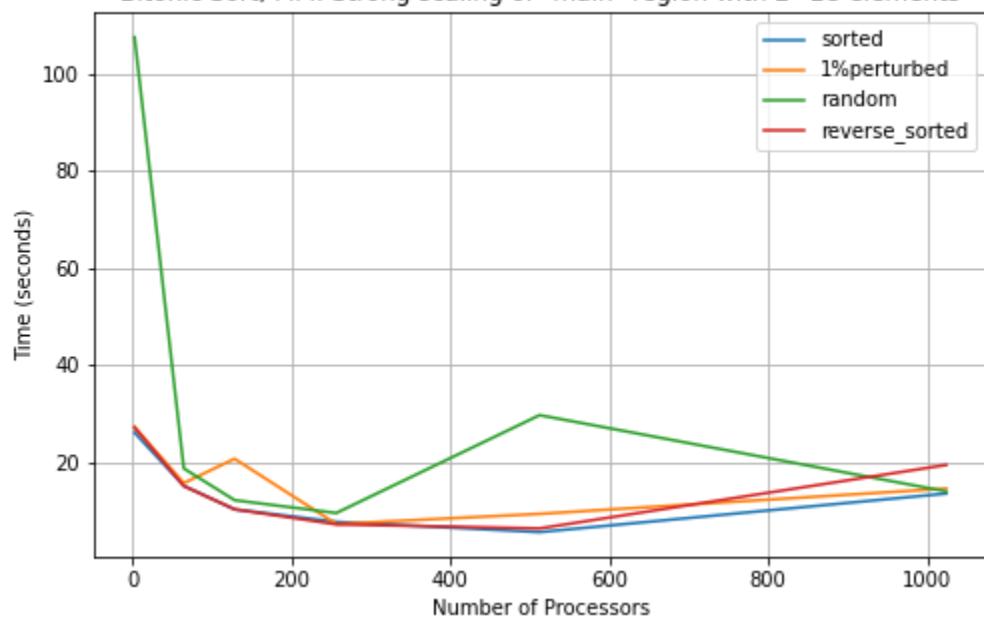
Bitonic Sort, MPI: Strong scaling of "main" region with 2^{24} elements



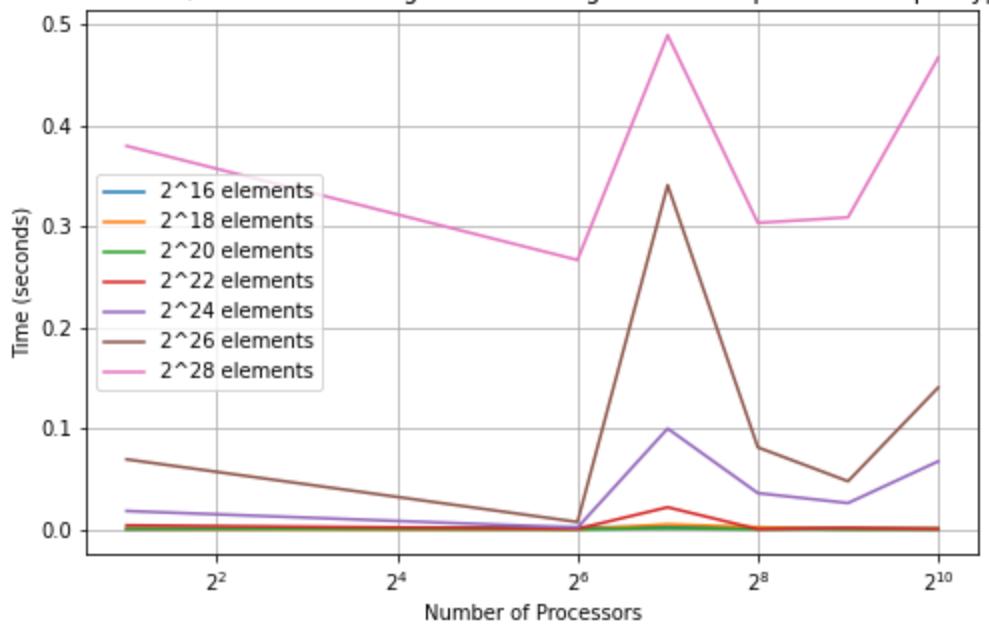
Bitonic Sort, MPI: Strong scaling of "main" region with 2^{26} elements



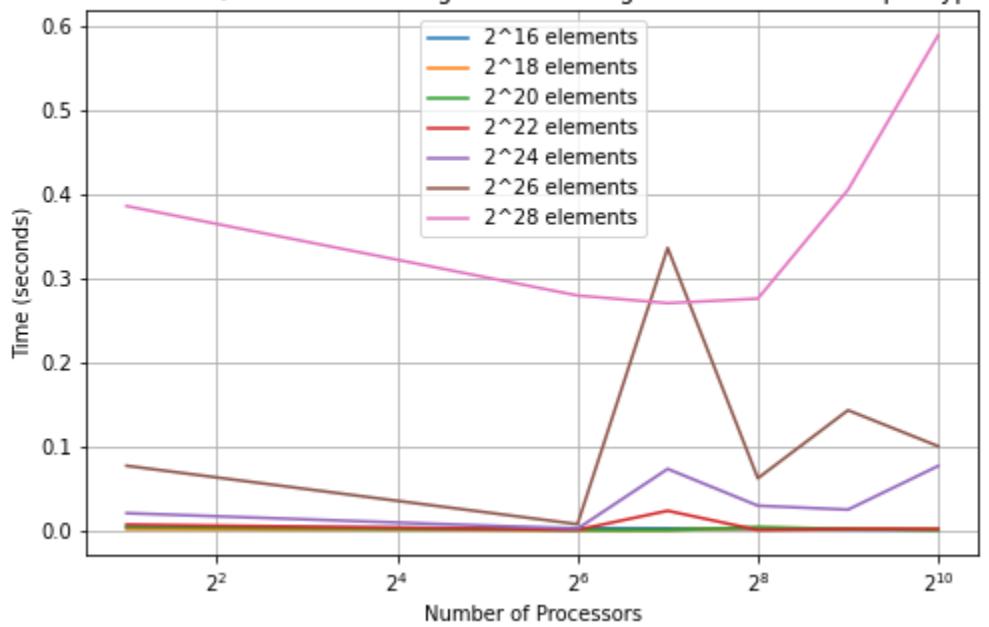
Bitonic Sort, MPI: Strong scaling of "main" region with 2^{28} elements



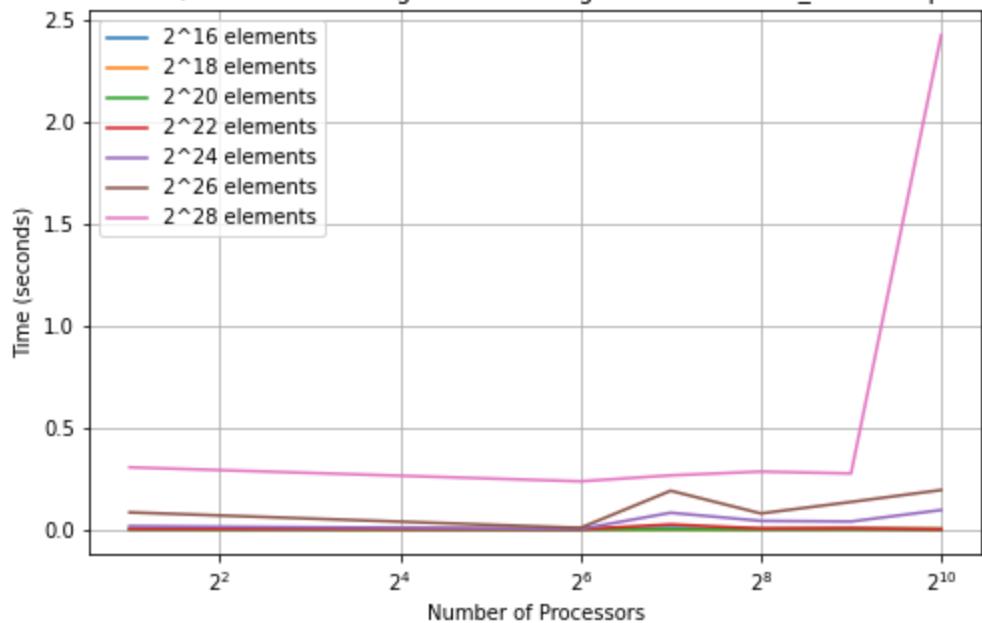
Bitonic Sort, MPI: Weak scaling of "comm" region with "1%perturbed" input type



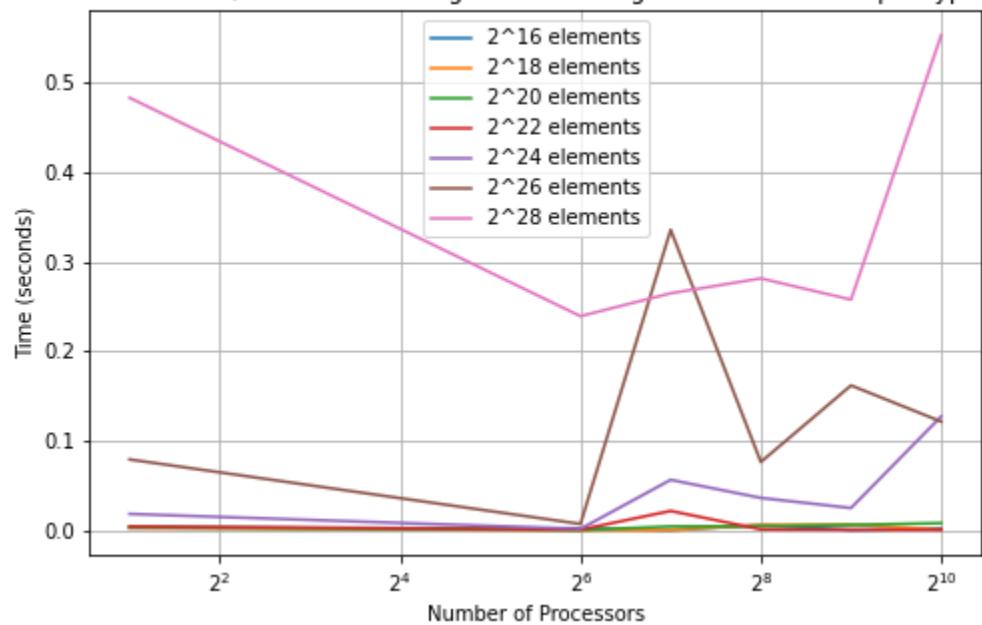
Bitonic Sort, MPI: Weak scaling of "comm" region with "random" input type



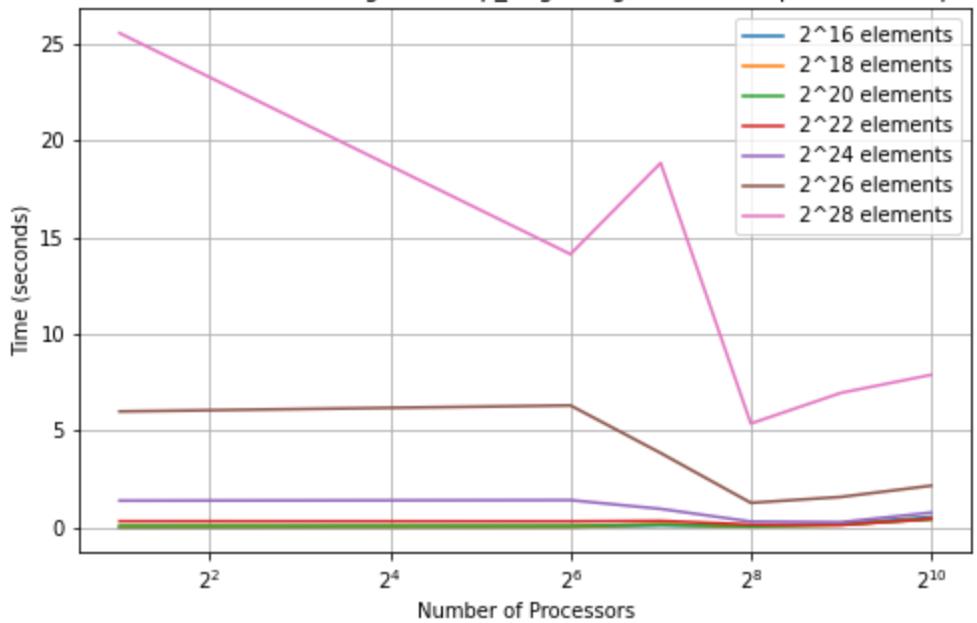
Bitonic Sort, MPI: Weak scaling of "comm" region with "reverse_sorted" input type



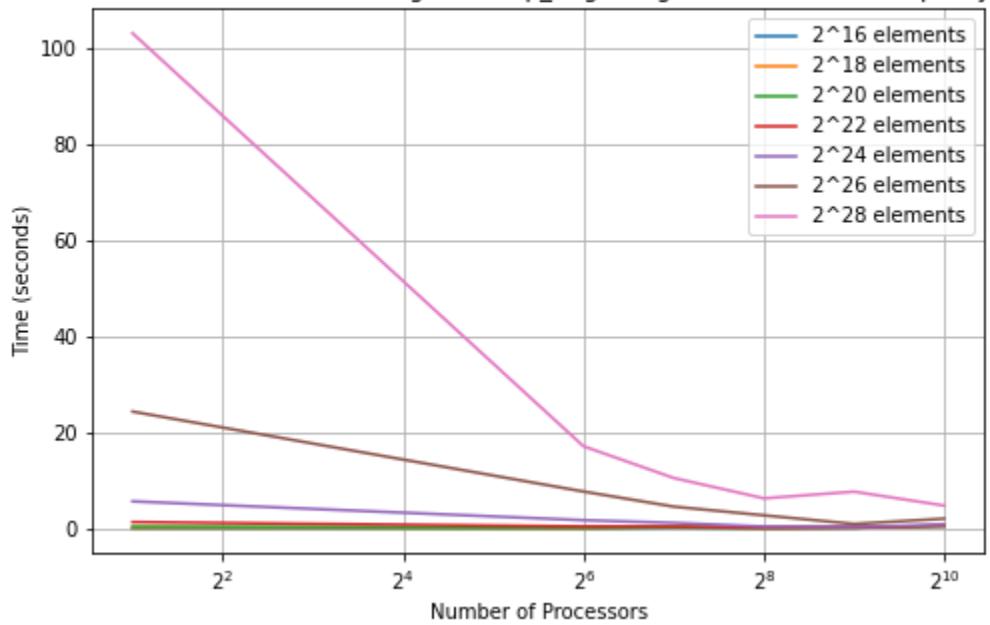
Bitonic Sort, MPI: Weak scaling of "comm" region with "sorted" input type



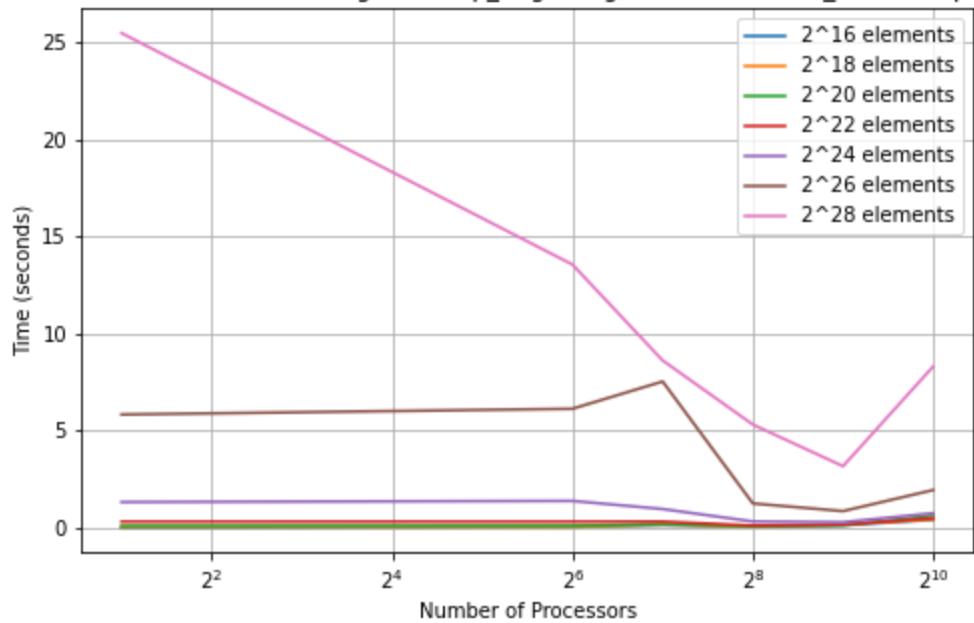
Bitonic Sort, MPI: Weak scaling of "comp_large" region with "1%perturbed" input type



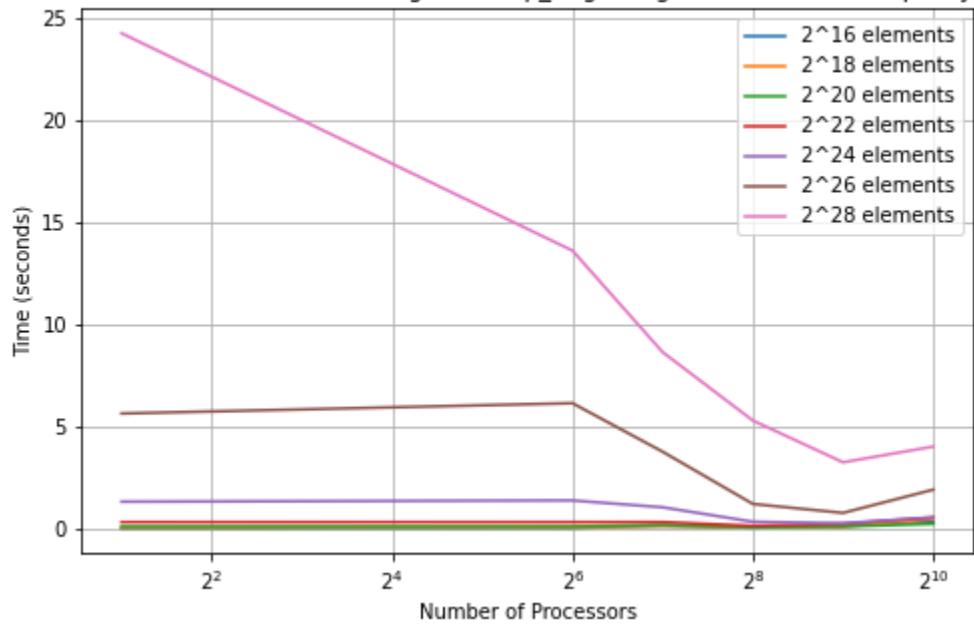
Bitonic Sort, MPI: Weak scaling of "comp_large" region with "random" input type



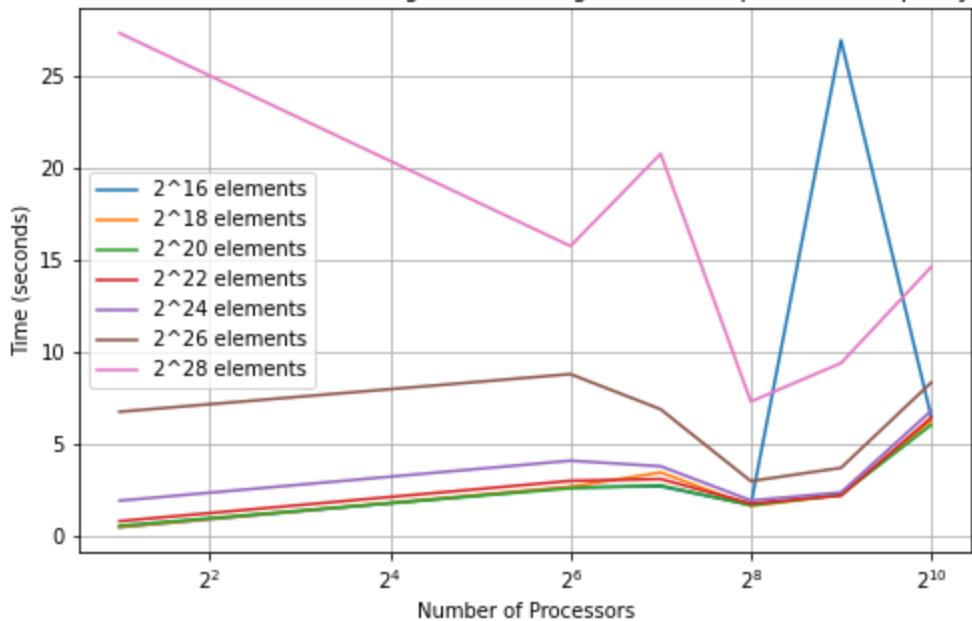
Bitonic Sort, MPI: Weak scaling of "comp_large" region with "reverse_sorted" input type



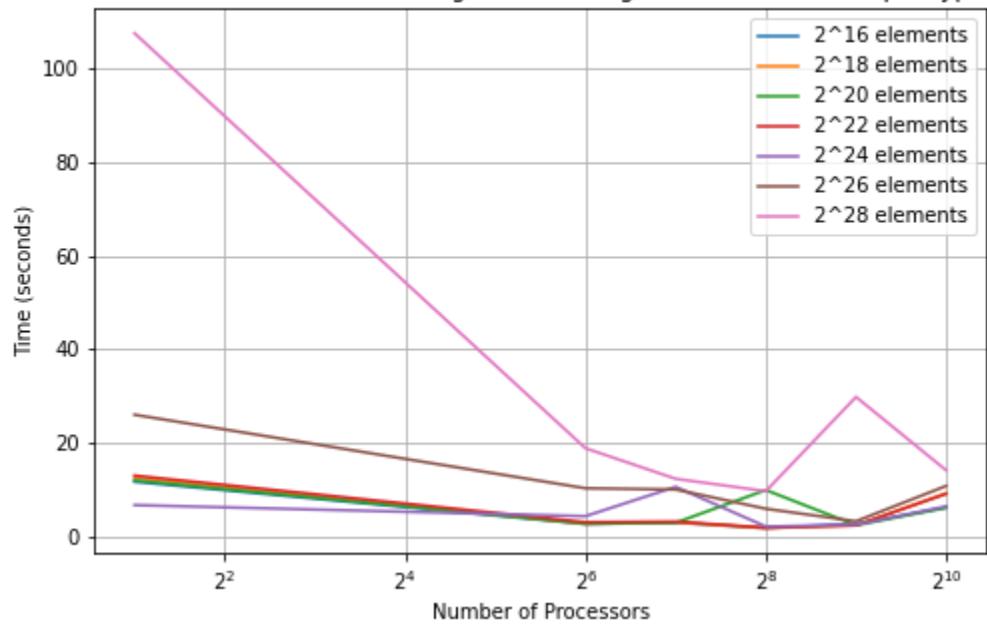
Bitonic Sort, MPI: Weak scaling of "comp_large" region with "sorted" input type



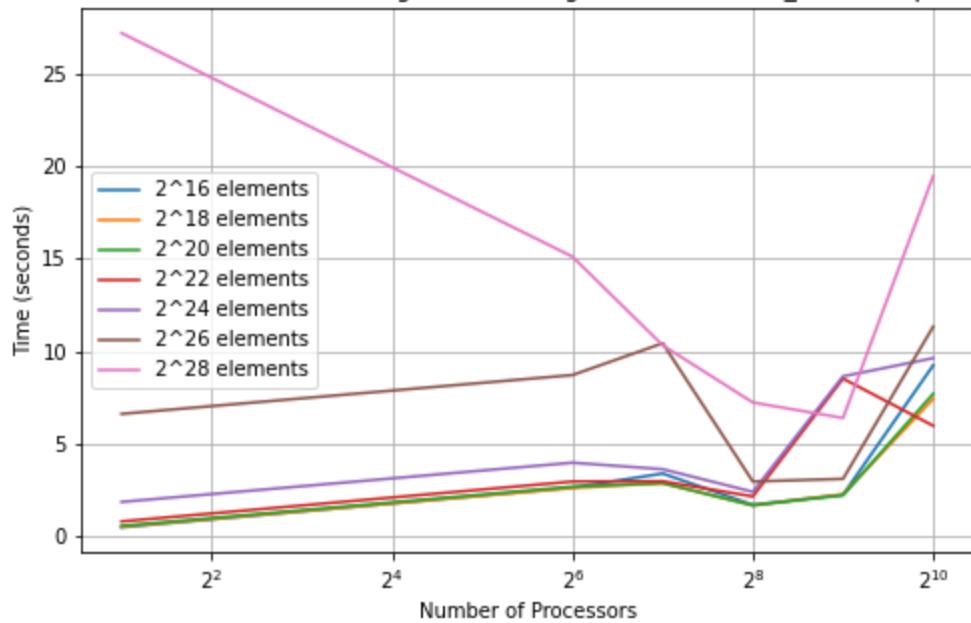
Bitonic Sort, MPI: Weak scaling of "main" region with "1%perturbed" input type



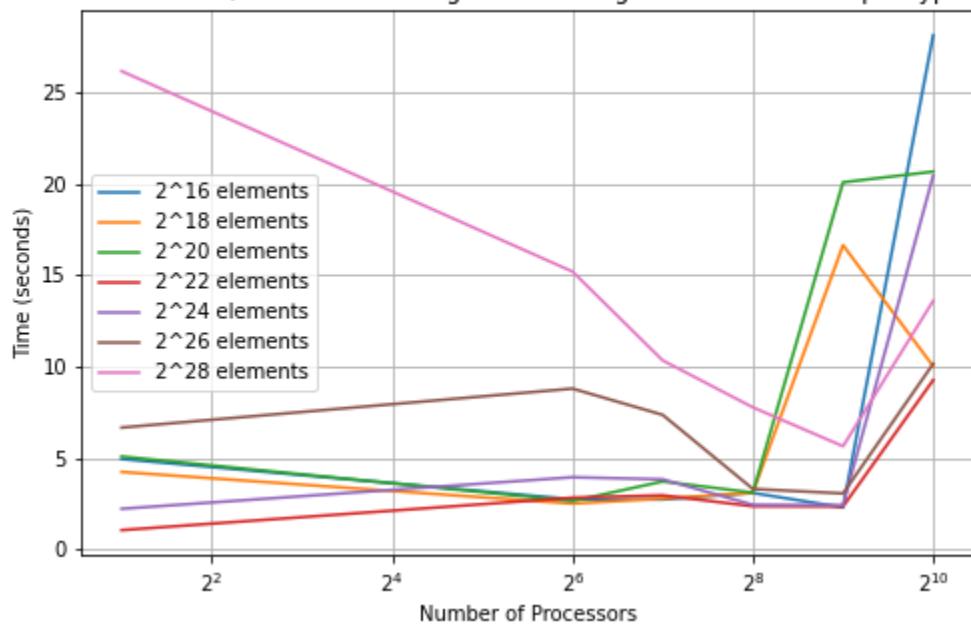
Bitonic Sort, MPI: Weak scaling of "main" region with "random" input type



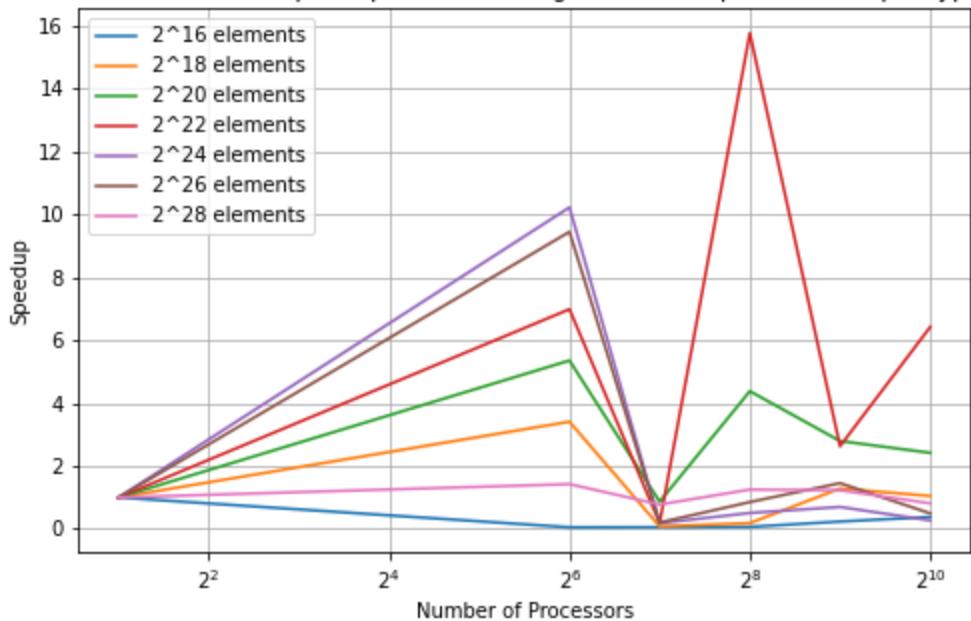
Bitonic Sort, MPI: Weak scaling of "main" region with "reverse_sorted" input type



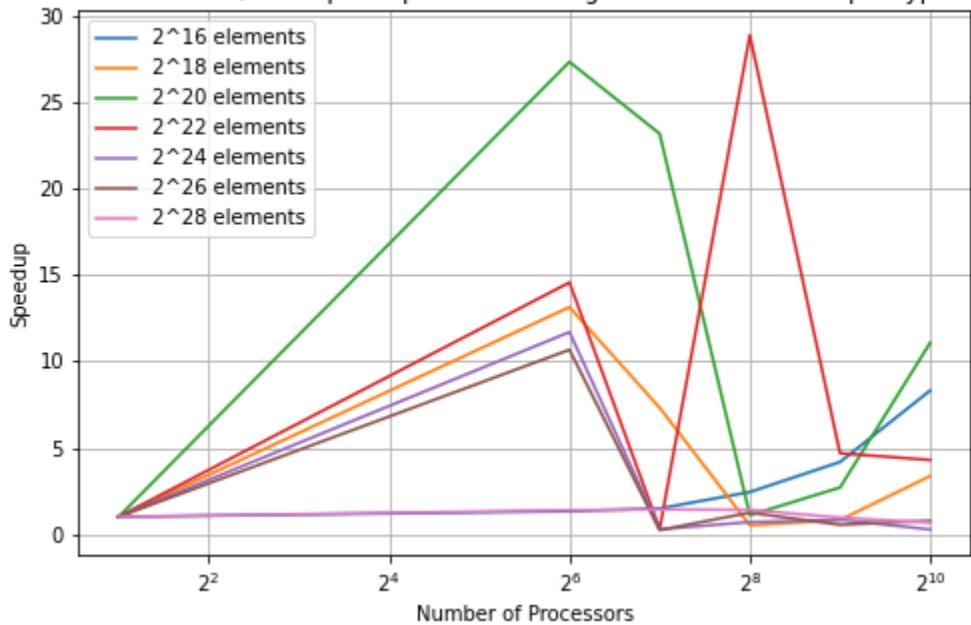
Bitonic Sort, MPI: Weak scaling of "main" region with "sorted" input type



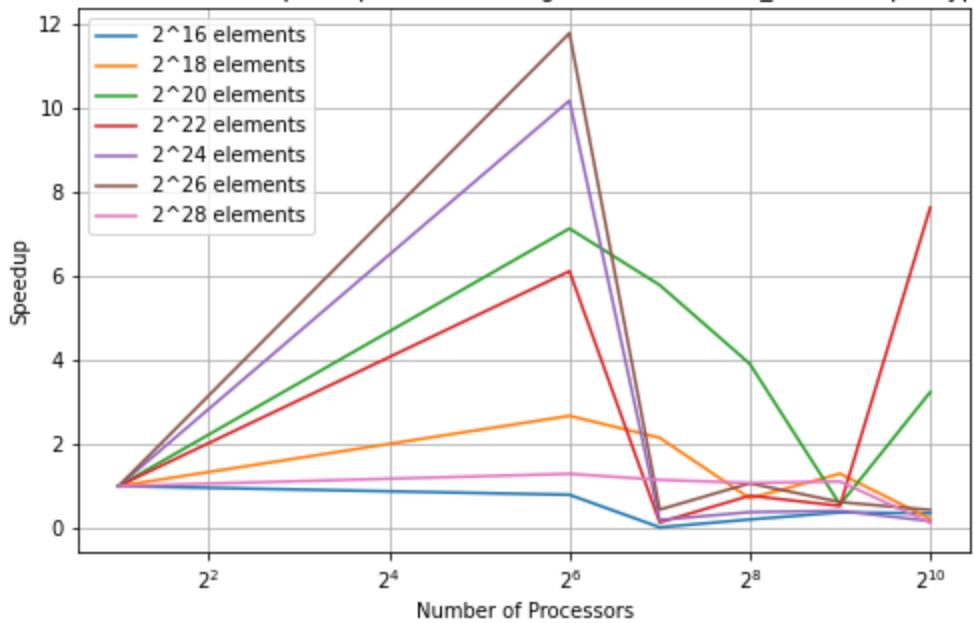
Bitonic Sort, MPI: Speedup of "comm" region with "1%perturbed" input type



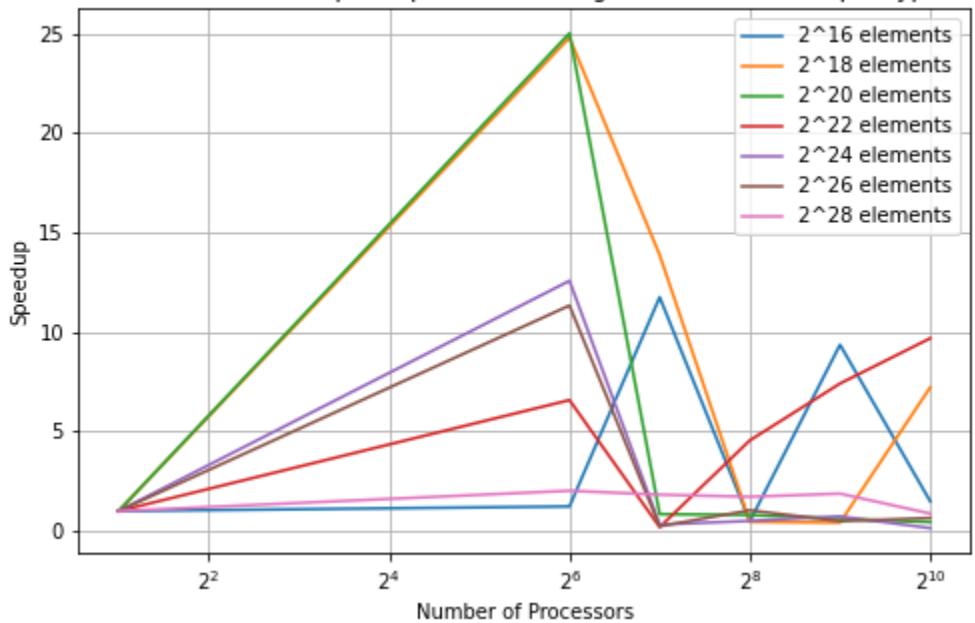
Bitonic Sort, MPI: Speedup of "comm" region with "random" input type



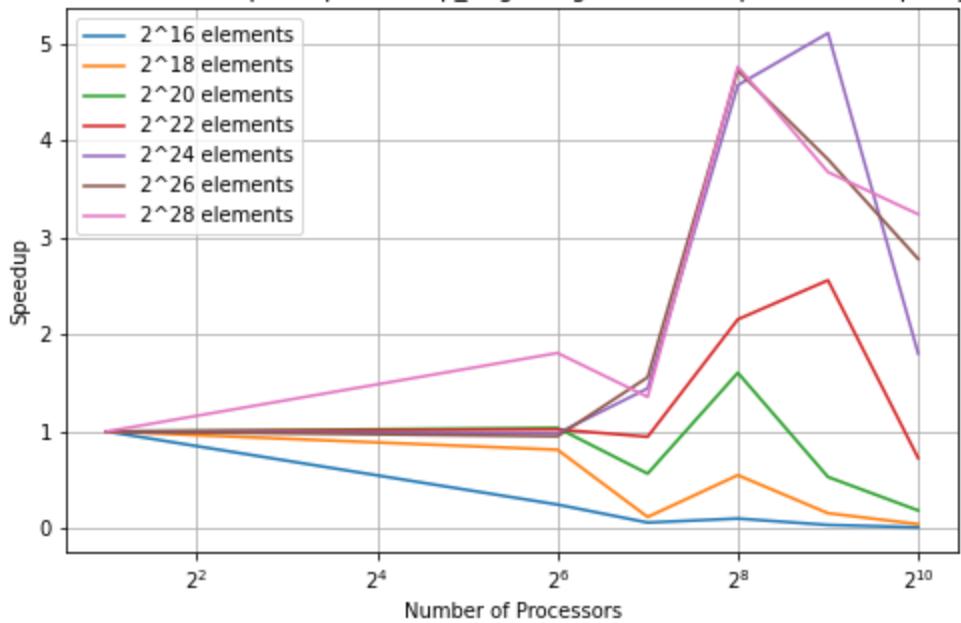
Bitonic Sort, MPI: Speedup of "comm" region with "reverse_sorted" input type



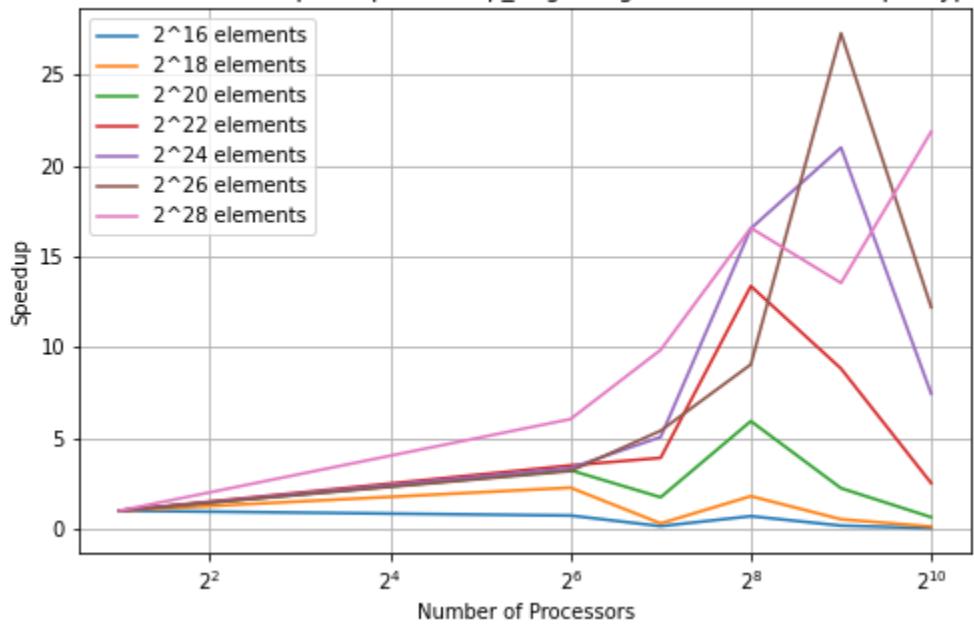
Bitonic Sort, MPI: Speedup of "comm" region with "sorted" input type



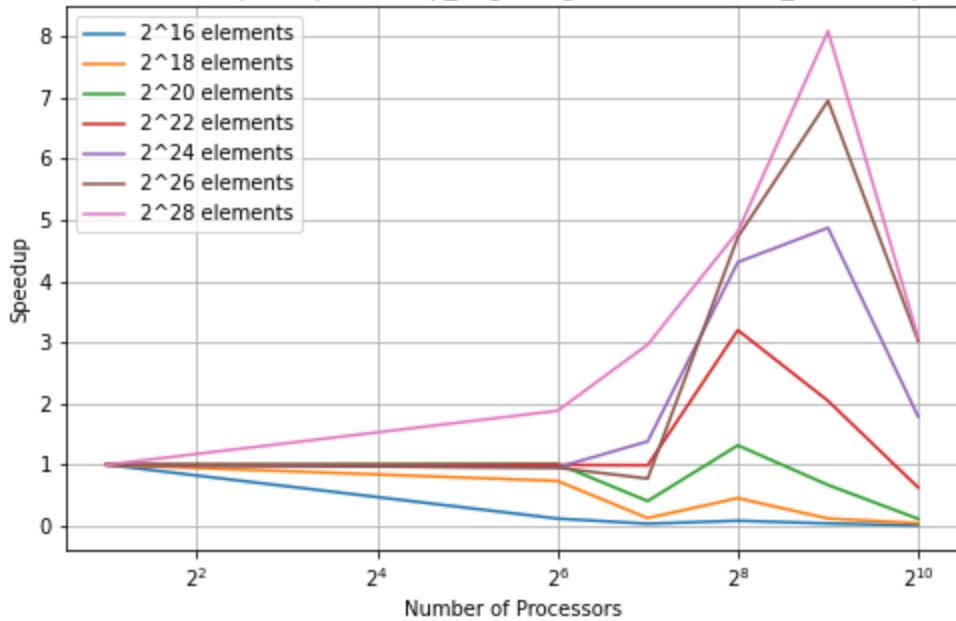
Bitonic Sort, MPI: Speedup of "comp_large" region with "1%perturbed" input type



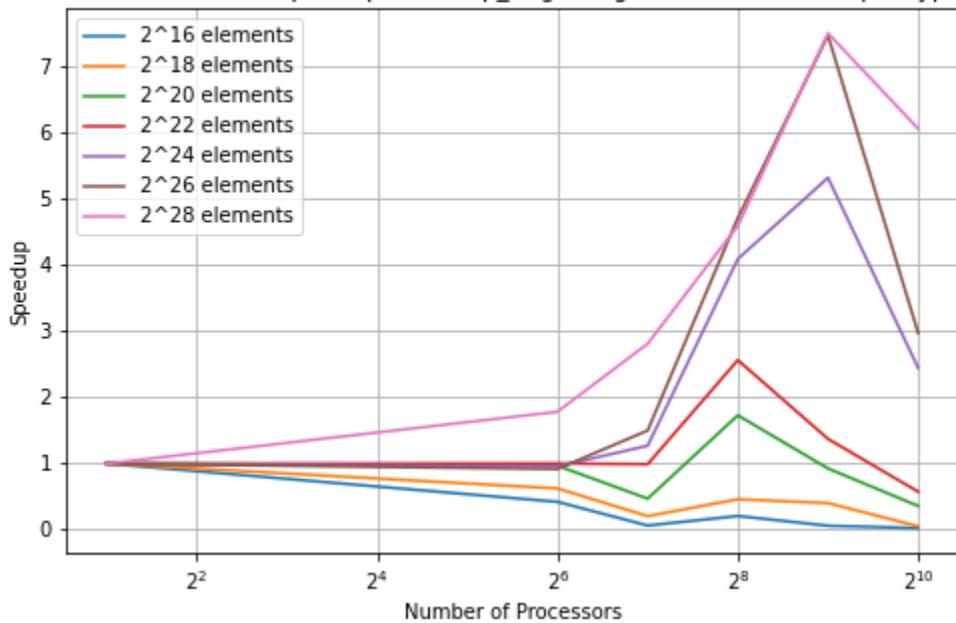
Bitonic Sort, MPI: Speedup of "comp_large" region with "random" input type



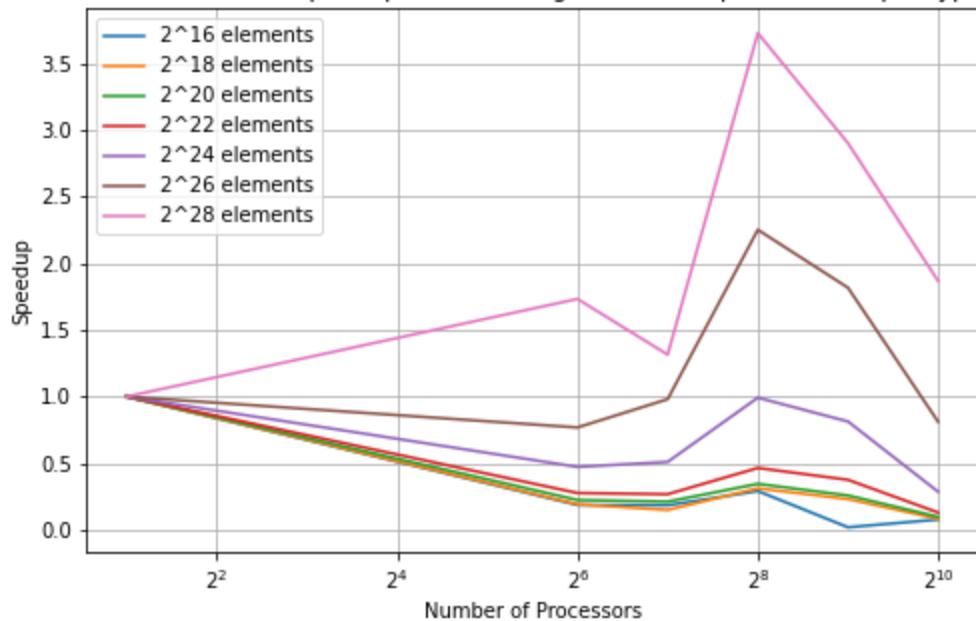
Bitonic Sort, MPI: Speedup of "comp_large" region with "reverse_sorted" input type



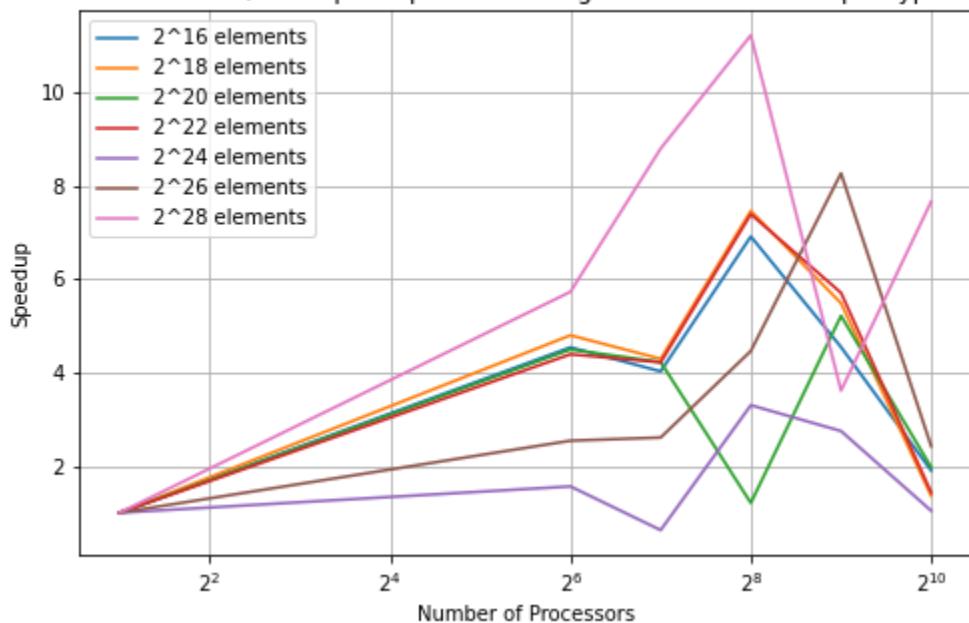
Bitonic Sort, MPI: Speedup of "comp_large" region with "sorted" input type

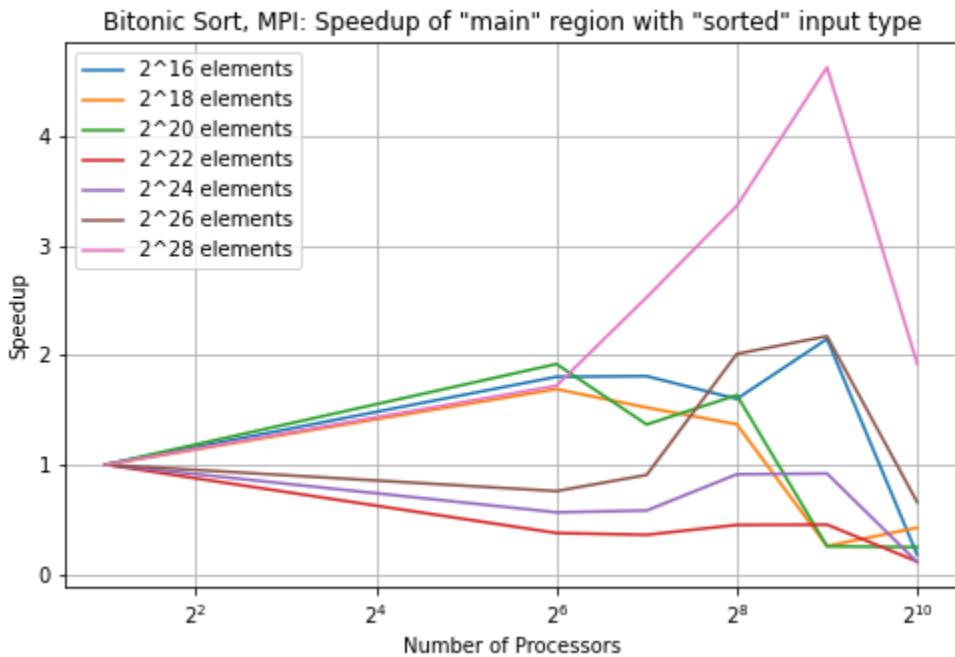
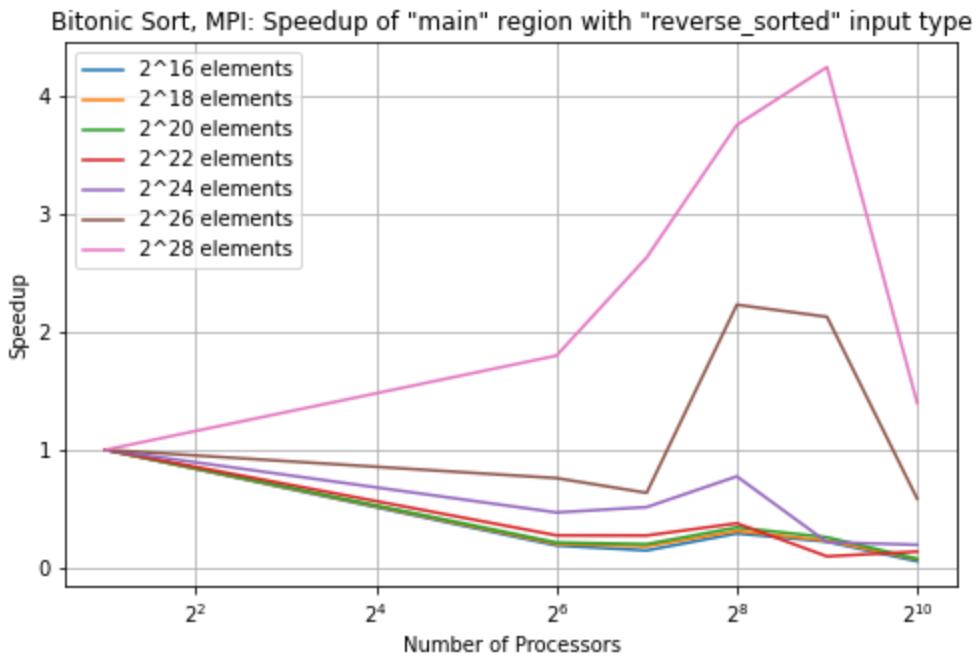


Bitonic Sort, MPI: Speedup of "main" region with "1%perturbed" input type



Bitonic Sort, MPI: Speedup of "main" region with "random" input type

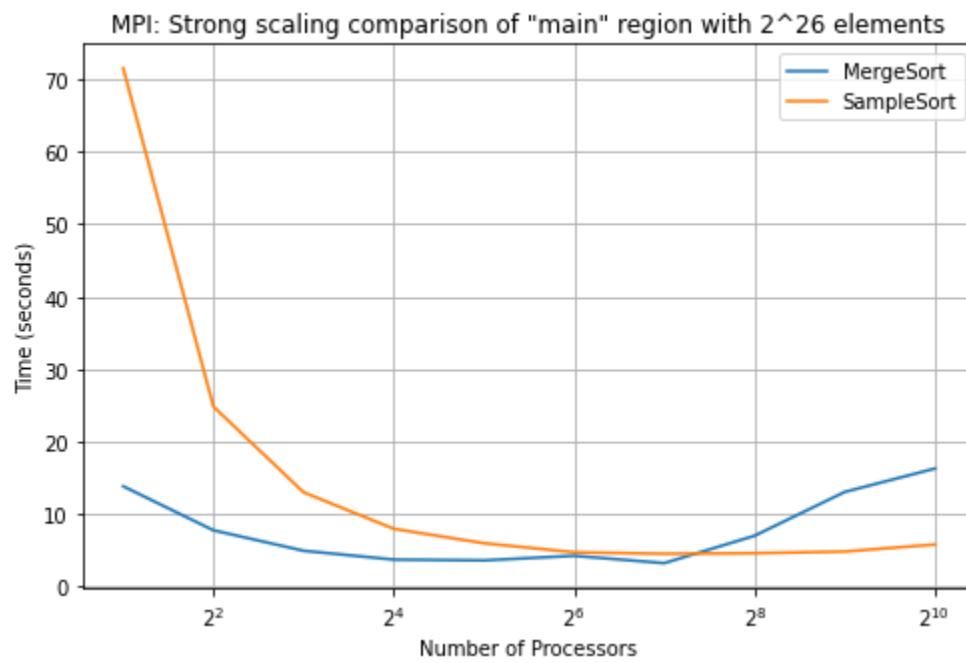
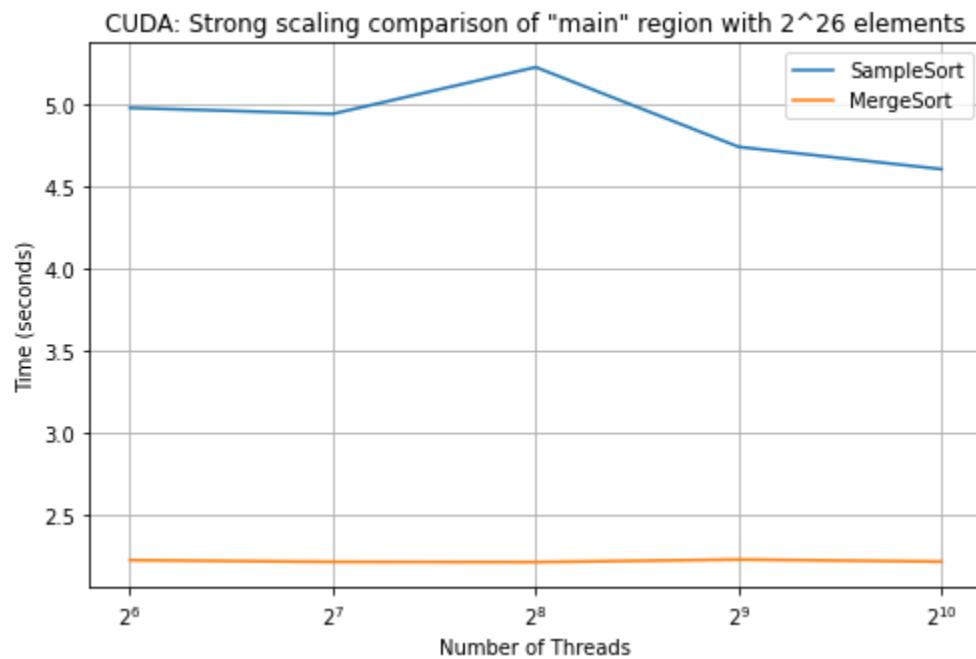




Throughout all of the implementations, we can expect an increase in the computational and communication time as the size of the array increases. We can also expect bitonic sort parallelism to reach a point of diminishing returns as we increase the number of threads. Additionally, as we change the different array populations such as random, sorted, reverse sorted, and 1% perturbed, we can expect slight differences in our

runtimes. For instance, a “sorted” array should have significantly less runtime than an unsorted array.

Comparison



When comparing Sample Sort and Merge Sort, you can see that Merge Sort was consistently faster with CUDA and MPI. The only time that Sample Sort was faster was with MPI for 2^7 , 2^8 , 2^9 , and 2^{10} processors.