High Performance Computing

Counting Sort with MPI

A.Y. 2021/22

Version 1.0.0

Alessio Pepe	0622701463	a.pepe108@studenti.unisa.it
Teresa Tortorella	0622701507	t.tortorella3@studenti.unisa.it
Paolo Mansi	0622701542	p.mansi5@studenti.unisa.it

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1 Index

2 Problem Description

2.1 Assignment

Parallelize and Evaluate Performances of "COUNTING SORT" Algorithm, by using MPI.

2.2 Algorithm description

Counting sort is a sorting algorithm that sorts the elements of an array of integer value by counting the number of occurrences of each unique element in the array. The count is stored in an auxiliary array and the sorting is done by mapping the count as an index of the auxiliary array.

The meta-algorithm of the counting sort is the following (from https://it.wikipedia.org/wiki/Counting_sort):

```
countingSort(fileName)
   A[] = read(fileName)

   //Calcolo degli elementi max e min
   max ← A[0]
   min ← A[0]
   for i ← 1 to length[A] do
      if (A[i] > max) then
         max ← A[i]
      else if(A[i] < min) then
         min ← A[i]
   end for

   * creates a C array of size max - min + 1
   for i ← 0 to length[C] do
      C[i] ← 0
   end for
   for i ← 0 to length[A] do
      C[A[i] - min] = C[A[i] - min] + 1
   end for
   k ← 0
   for i ← 0 to length[C] do
      while C[i] > 0 do
         A[k] ← i + min
         k ← k + 1
         C[i] ← C[i] - 1
   end for

   write(fileName, A[])
```

The algorithm consists of 5 parts:

- Read the array from file
- find the maximum and minimum value within the array to calculate the width of the range of values.
- use an auxiliary array to counter the frequency of each single value.
- repopulate the array in an orderly manner based on the newly calculated frequency array.
- Write the array from file

The computational complexity of the algorithm is given by the sum of the computational complexities of the three parts and is in total

$$\big(O(n)\big) + \big(O(n)\big) + \big(O(n) + O(k)\big) + \big(O(n)\big) + \big(O(n)\big)$$

where:

- n is the number of elements of the array.
- k is the size of the range of elements in the array (we are considering on average that the allocation of a block of memory costs $O(1)$.

We are aware that it is not possible to parallelize the PMF vector allocation operation and therefore the theoretical speedup curve should not be that of the images that will be shown in the solutions section later, but since this algorithm makes sense to be used when $k \ll n$ then we consider that represented a good approximation.

2.3 Solution idea

2.3.1 Read data

As first thing, the array of data must be read from the file. We implemented a fully parallel I/O, opening the File on the Communicator and let each thread read a different part of the file so that each threads owns a local copy of just a single part of the array. The size of the local array is the same for the first N-1 threads and equal to the length of the total size that must be read divided by the number of threads available. The last thread though is appointee of also reading the last part of the vector, for an additional number of elements up to N-1. The computational cost is ideally

$$O\left(\frac{n}{p}\right)$$

where n is the size of the array and p is the number of threads available

2.3.2 Min-max

In implementing this part of the algorithm with MPI we manually implemented the reduction mechanism parallelized between threads.

Each thread performs the sequential search of minimum and maximum on the local array it owns, then an AllReduce is called for both of them so that each thread associated to the communicator has access to the global minimum and maximum among all threads.

The computational complexity of this solution is ideally

$$O\left(\frac{n}{p}\right)$$

however, it is necessary to consider the transfer time of the min and max reductions, which is carried out in this phase of the algorithm.

2.3.3 Counting occurrence

To calculate the vector of PMF we allocate a local vector for each thread that will use it to calculate the element frequencies of its part of the array. In addition to this, the master thread also allocates an array of the same length which will contains the global PMF. Then each thread computes the local PMF on its part of the array and finally the reduction is implemented to sum together all the local PMF onto the master thread. The computational cost of this part is ideally

$$O\left(\frac{n}{p}\right)$$

The transfer time of the various local PMF carriers in the reduction must also be considered.

2.3.4 Populating array – CDF division

To repopulate the array, two different solutions are provided.

In the first one, only the main thread calculates sequentially the vector of the CDF starting from that of the PMF to allow the threads to know the position of the array in which the elements must be inserted. In fact, subsequently, MPI is used send, through the ScatterV, to each thread the portion of the CDF it must use to repopulate.

In the second solution instead, each thread calculates the CDF from the beginning to the element of its interest, leaving only the Nth thread to calculate the whole CDF.

Then, each thread must compute its part of the array to write based on the portion of the CDF. The computational complexity of first solution is ideally

$$O(k) + O\left(\frac{k}{p}\right)$$

However, in both cases, the times for transferring part of the CDF array and transferring the entire vector of the PMF must also be considered.

2.3.5 Write Data

Lastly, each thread must write its portion of data onto the file. We implemented the full parallel I\O where each thread autonomously calculates the offset at which write without collision with memory area related to other threads and then it writes the contiguous array of data starting at that location. The computational complexity is ideally

$$O\left(\frac{n}{p}\right)$$

2.3.6 Total solution

We have provided two files counting_sort.c and counting_sort_mpi.c (along the related header files) which contain the implementations of the sequential and parallel version of the algorithm. In the .c file some macros are set that allow you to enable the print of the time measures of each part of the algorithm.

An additional parameter is required when launching the main_measures.c and main_measures_sequential.c to select which of the two-algorithm version to use.

3 How to Run

3.1 Organization of the repository

The files containing the source code, the scripts, the license, and everything associated with it can be viewed at the link https://github.com/pepealessio/countingsort_mpi.git and have been provided together with the following document.

The following directories are present:

- **/include**: it contains all the header files necessary to solve the requested problem.
- **/script:** contains the scripts needed to evaluate the performance of the provided code.
- **/source:** it contains all the source code and the main file needed to solve the problem.
- **/test:** contains the test files, usable with the ctest framework (or make test), which demonstrate the actual functionality of the provided application.

The following folder is provided for the purpose of demonstrating the authors' measurements, but it can be deleted, or the content will be moved in the folder at the time of application launch:

The following folders are not present in the directory provided, but will be generated when launching the application:

- **/.oldmeasures:** If new measures will be generate when /measures folder already exists, the contents of the measurement folder will be moved to this folder in a subfolder named with the date and time of the move.
- **/.venv:** (will be generated at the script time if the user decide to use a virtual environment as described in the section below). It's just the virtual environment for python3.
- **/build:** Contains all built file with the make command.
- **/docs:** Contains the Doxygen documentation of the source code. The documentation was provided in 2 formats:
    - **/docs/html**
    - **/docs/latex**
- **/measures:** It contains all the raw and processed data of the evaluations (if provided contains measure made on one of the authors' machines). There are the following subfolders:
    - **/measures/plots:** Contains the plots, for each optimization level and size, of the speedup and efficiency as the number of threads used varies, extrapolated from the processed data. Each image has a name structure like *T_ALGO_opt_Y_size_Z.png*, where *Y* indicates the compiler optimization level used for those measurements, and *Z* indicates the size of the problem used for those measurements.
    - **/measures/processed:** Contains data extracted from raw measurements, where speedup and efficiency are compared based on the use of a different number of threads. Each file has a name structure like *T_ALGO_opt_Y_size_Z_range_R.csv*, where *Y* indicates the compiler optimization level used for those measurements, *Z* indicates the size of the problem used for those measurements, and *R* indicates the dimension of the element range of the array.
    - **/measures/raw:** Contains all raw measurements performed by the script. Each file has a name structure like *opt_Y_size_Z_range_R_threads_X.csv*, where *Y* indicates the compiler optimization level used for those measurements, *Z* indicates the size of the problem used for those measurements, and *R* indicates the dimension of the element range of the array and *X* indicate the number of OpenMP's threads used in the application run (0 means sequential application without OpenMP, so with 1 thread).

There are also the following files:

- **.gitignore:** Specifies intentionally untracked files to ignore. We used that for share our work on GitHub.
- **CmakeLists.txt:** contains all the directives needed to compile and test all the necessary executables.
- **Doxyfile:** it contains all the directives to automatically generate the documentation associated with the code.
- **licence.txt:** contains a transcript of the license to which the code is subject.
- **ReadMe.md:** probably the file that brought you here.
- **report.pdf:** it is this same file and contains an account of what has been done.
- **report_benchmark.zip**: it contains the raw data, the processed data and the plots present in the report.
- **requirements.txt:** contains all the packages required by python to run the measurement script.

3.2 Build phase

First, you need to compile all the files. For this you need a GCC compiler, MPI, Doxygen and Graphviz installed on your machine. If not, the process may fail.

1. Head to the main directory.
2. Create a new build folder inside:
```
user@PC:~/dir$ mkdir build
```
3. Go to the build folder:
```
user@PC:~/dir$ cd build
```
4. Launch the build command with the parameter the root directory containing the CMakeLists.txt file:
```
user@PC:~/dir/build$ cmake ..
```
5. Then run the make command
```
user@PC:~/dir/build$ make all
```

At this point, in the build folder there will be all the compiled executables and in the main directory there will be a new folder named docs containing the source code documentation in html and latex.

3.3 Test phase

You can verify the correct functioning of the application provided by using the cmake package.

6. Go to the build folder (if you are following all the directions then you will already be in this one).
7. Launch the ctest command (or also make test):
```
user@PC:~/dir/build$ ctest
```

The desired output is as follows

```
100% tests passed, 0 tests failed out of 2
```

3.4 Install Python3 requirements phase

You need to check whether the python requirements described in the requirements.txt file are satisfied or not. If they are not, the authors recommend using a virtual environment to not occupy memory and override the python installation. This can be done as:

8. Head to the main directory.
9. Create the virtual environment from the current python3 installation (we used python3.8). If you don't have it, you need the package python3.8-venv (installable with the command `sudo apt install python3.8-venv`).
```
user@PC:~/dir$ python3 -m venv .venv
```
10. Now you need to install all the required packages:
```
user@PC:~/dir$ source ./.venv/bin/activate
(.venv) user@PC:~/dir$ pip3 install -r requirements.txt
(.venv) user@PC:~/dir$ deactivate
```

3.5 Measure phase

At this point, everything is ready to run the measurement script. To run the script, you need to run the following commands

```
user@PC:~/dir$ bash script/measures.bash
```

You can see the output in the measurements folder as indicated in 3.1, but just the raw data. To extract the processed data and the plots you can use the command:

```
(.venv) user@PC:~/dir$ python3 ./script/extract_measures.py
```

Note: Once you have performed steps 3.2 to 3.4, you no longer need to perform them. Step 3.5 can be performed directly to obtain new measurements. In case of changes to the source code it will be necessary to repeat steps 3.2 and 3.3.

3.5.1 Change measurement parameters

it is likely that, depending on the machine on which you want to perform the measurement, you want to edit the measurement parameters such as the dimensions tested, the number of threads used, the optimizations used and the number of measurements for each configuration.

To do this, you need to go to the */script/measures.bash* and in the */script/extract_measures.py* file. Immediately after the license and the fields intended for the file's metadata, some constants are defined as shown here

```
ARRAY_RC=[1000, 10000, 100000, 1000000, 10000000]
ARRAY_RANGE=[1000, 10000, 100000]
ARRAY_THS=[0, 1, 2, 4, 8]
ARRAY_OPT=[0, 1, 2, 3]
```

and here

```
NMEASURES=100

ARRAY_RC=(1000 10000 100000 1000000 10000000)
ARRAY_RANGE=(1000 10000 100000)
ARRAY_THS=(0 1 2 4 8)
ARRAY_OPT=(0 1 2 3)
```

The configurations are given by all the possible combination *(ARRAY_RC, ARRAY_RANGE, ARRAY_THS, ARRAY_OPT)*. Therefore, if these have lengths 5, 3, 5 and 3 respectively, $3 \cdot 3 \cdot 5 \cdot 3 = 135$ configurations will be obtained.

- To change the configurations simply add or remove values from this fields (in the *ARRAY_RC* 0 **must** be present to correctly evaluate the speedup and the efficiency).
- To change the number of repetitions for each configuration just change the NMEASURES value.

Warning: the two configurations in the two files **must** match.

4 Experimental Setup

4.1 OS setup

All measurements were made using the Windows 10 operating system. Ubuntu 20.04 LTS virtualized using WSL 2 was used. We were using gcc9 as compiler

4.2 Hardware setup

The hardware configuration is reported at the time of measuring the performance of the solutions provided. The configuration is extracted from `/proc/cpuinfo` and `/proc/meminfo`.

```
user@PC:~/dir$ cat /proc/cpuinfo
(…omitted firt 7 threads…)
processor       : 7
vendor_id       : AuthenticAMD
cpu family      : 23
model           : 17
model name      : AMD Ryzen 5 2500U with Radeon Vega Mobile Gfx
stepping        : 0
microcode       : 0xffffffff
cpu MHz         : 1996.185
cache size      : 512 KB
physical id     : 0
siblings        : 8
core id         : 3
cpu cores       : 4
apicid          : 7
initial apicid  : 7
fpu             : yes
fpu_exception   : yes
cpuid level     : 13
wp              : yes
flags            : fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36
clflush mmx fxsr sse sse2 ht syscall nx mmxext fxsr_opt pdpe1gb rdtscp lm constant_tsc rep_good
nopl tsc_reliable nonstop_tsc cpuid extd_apicid pni pclmulqdq ssse3 fma cx16 sse4_1 sse4_2
movbe popcnt aes xsave avx f16c rdrand hypervisor lahf_lm cmp_legacy cr8_legacy abm sse4a
misalignsse 3dnowprefetch osvw topoext ssbd ibpb vmmcall fsgsbase bmi1 avx2 smep bmi2 rdseed
adx smap clflushopt sha_ni xsaveopt xsavec xgetbv1 xsaves clzero xsaveerptr virt_ssbd arat
bugs            : sysret_ss_attrs null_seg spectre_v1 spectre_v2 spec_store_bypass
bogomips        : 3992.37
TLB size        : 2560 4K pages
clflush size    : 64
cache_alignment : 64
address sizes   : 48 bits physical, 48 bits virtual
power management:
```

```
user@PC:~/dir$ cat /proc/meminfo
MemTotal:        5587448 kB
MemFree:         4452404 kB
MemAvailable:    4709292 kB
Buffers:           27012 kB
Cached:           417884 kB
SwapCached:            0 kB
Active:           207008 kB
Inactive:         783184 kB
Active(anon):        144 kB
Inactive(anon):   545212 kB
Active(file):     206864 kB
Inactive(file):   237972 kB
Unevictable:           0 kB
Mlocked:               0 kB
SwapTotal:       2097152 kB
SwapFree:        2097152 kB
```

```
Dirty:                12 kB
Writeback:             0 kB
AnonPages:        545324 kB
Mapped:            91368 kB
Shmem:                72 kB
KReclaimable:      37652 kB
Slab:              66744 kB
SReclaimable:      37652 kB
SUnreclaim:        29092 kB
KernelStack:        3472 kB
PageTables:        15132 kB
NFS_Unstable:          0 kB
Bounce:                0 kB
WritebackTmp:          0 kB
CommitLimit:     4890876 kB
Committed_AS:     756860 kB
VmallocTotal:   34359738367 kB
VmallocUsed:       23056 kB
VmallocChunk:          0 kB
Percpu:             2432 kB
AnonHugePages:    182272 kB
ShmemHugePages:        0 kB
ShmemPmdMapped:        0 kB
FileHugePages:         0 kB
FilePmdMapped:         0 kB
HugePages_Total:       0
HugePages_Free:        0
HugePages_Rsvd:        0
HugePages_Surp:        0
Hugepagesize:       2048 kB
Hugetlb:               0 kB
DirectMap4k:      101376 kB
DirectMap2M:     3590144 kB
DirectMap1G:     2097152 kB
```

Additionally, the available physical memory is reported:

```
user@PC:~/dir$ echo $(($(getconf _PHYS_PAGES) * $(getconf PAGE_SIZE) / (1024 * 1024))) "MB"
5456 MB
```

Finally, the available virtual memory is reported:

```
user@PC:~/dir$ echo $(($(getconf _AVPHYS_PAGES) * $(getconf PAGE_SIZE) / (1024 * 1024))) "MB"
4348 MB
```

5 Results

5.1 Results organization

Measurements are made on different dimensions of the problem, using a sequential version of the algorithms and a parallelized version by trying with different numbers of active threads. All the possibilities analyzed are divided by size of the problem.

Each box consists of 5 images, meaning the time of reading, minmax, pmf calculation, writing and complete timing of the whole algorithm. There is a box for each of the two version of the Algorithm. Each image contains the ideal speedup (blue line), experimental speedup (continuous broken lines), and efficiency (broken lines).

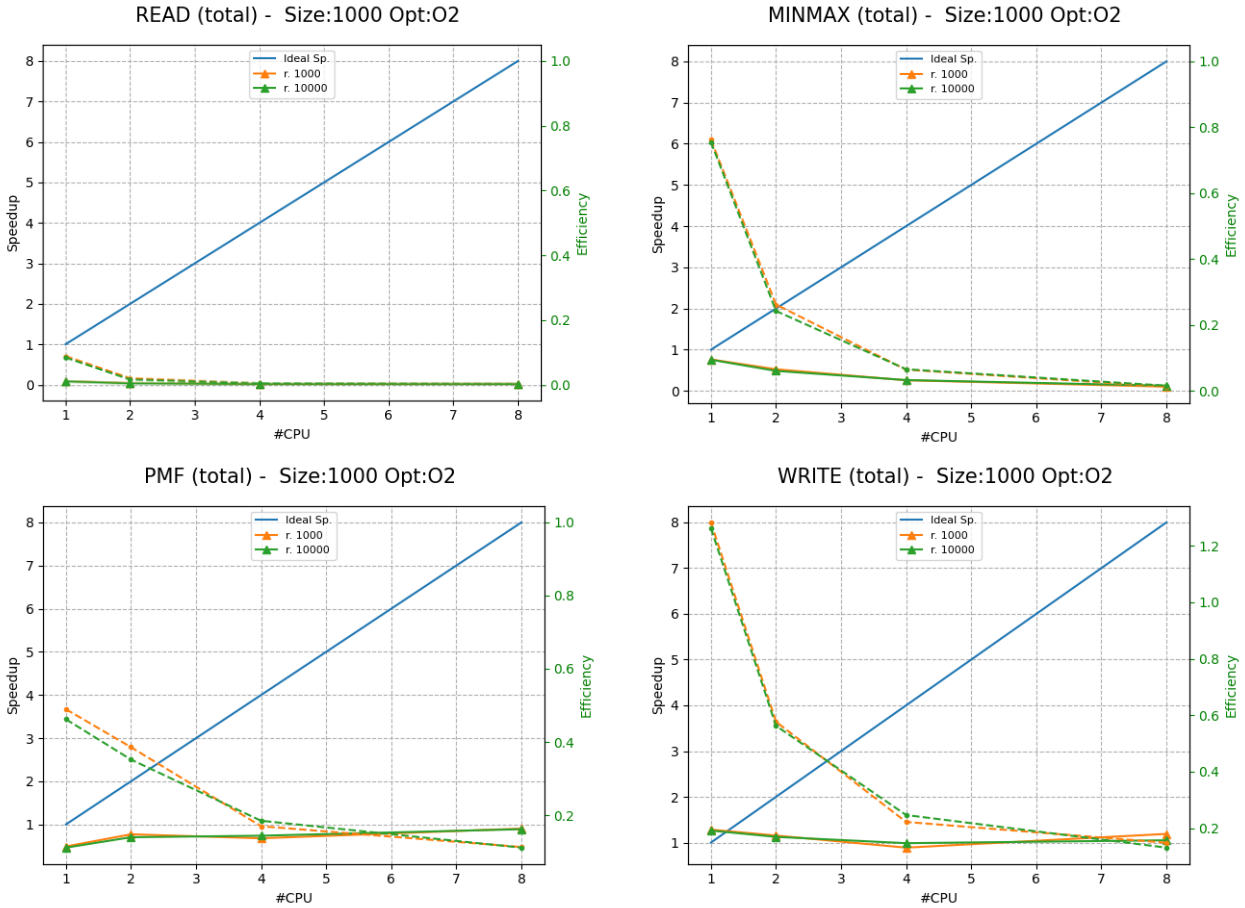
In this section of the report, only O2-optimized results are commented on because they give the best results. The others have however been provided and have the same meaning (except for performance improvements due to optimization).

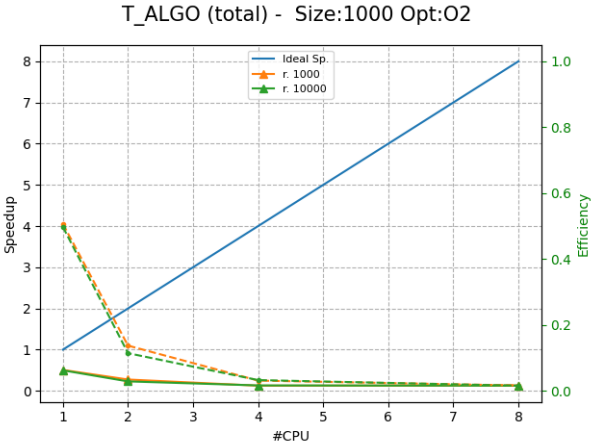
The raw, processed data and the plots of these were provided for the purpose of consultation. The tabulated data from which the plots are generated are not reported for the sake of brevity but are found within the project.

5.2 Optimization O2

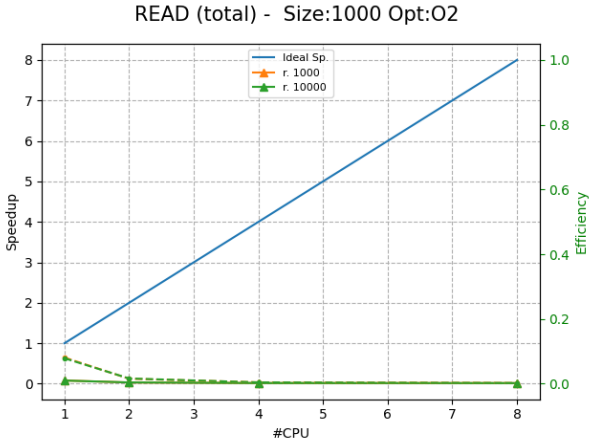
5.2.1 Size 10^3

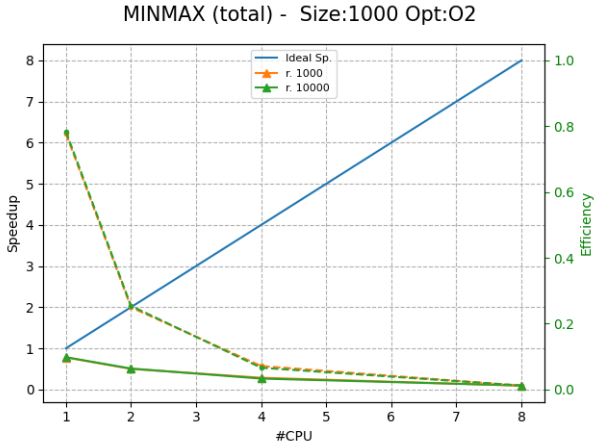
Version 1

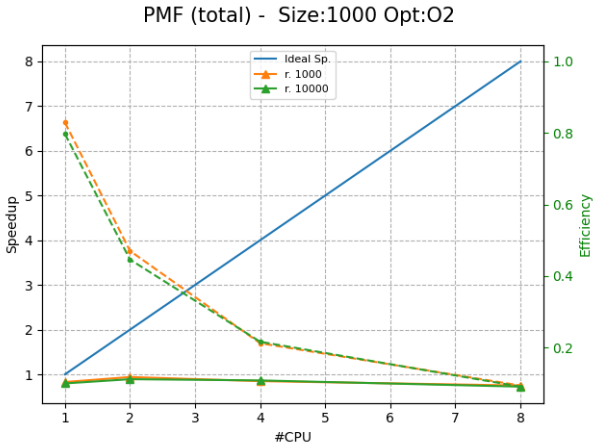


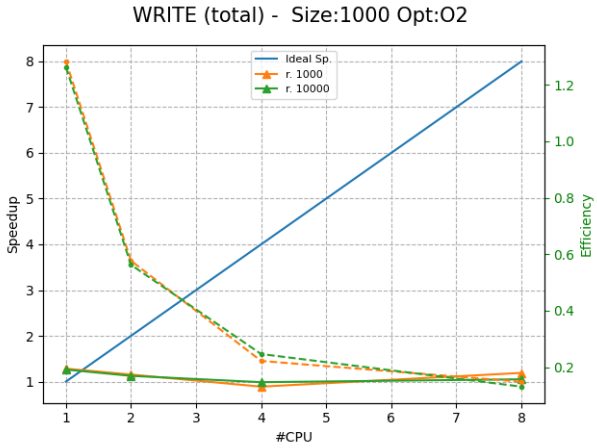

T_ALGO (total) - Size:1000 Opt:O2
Ideal Sp.
r. 1000
r. 10000
Speedup
Efficiency
#CPU

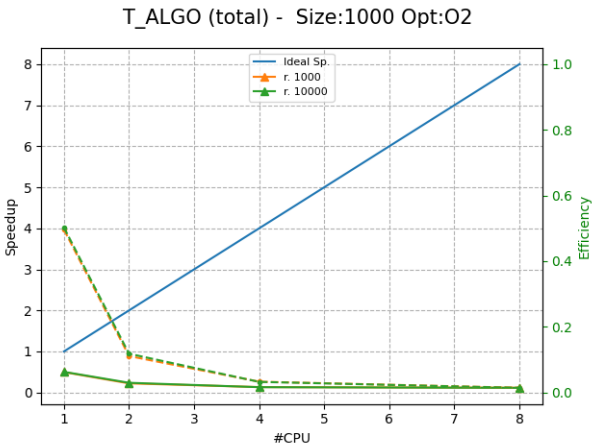
Version 2


READ (total) - Size:1000 Opt:O2
Ideal Sp.
r. 1000
r. 10000
Speedup
Efficiency
#CPU


MINMAX (total) - Size:1000 Opt:O2
Ideal Sp.
r. 1000
r. 10000
Speedup
Efficiency
#CPU


PMF (total) - Size:1000 Opt:O2
Ideal Sp.
r. 1000
r. 10000
Speedup
Efficiency
#CPU


WRITE (total) - Size:1000 Opt:O2
Ideal Sp.
r. 1000
r. 10000
Speedup
Efficiency
#CPU

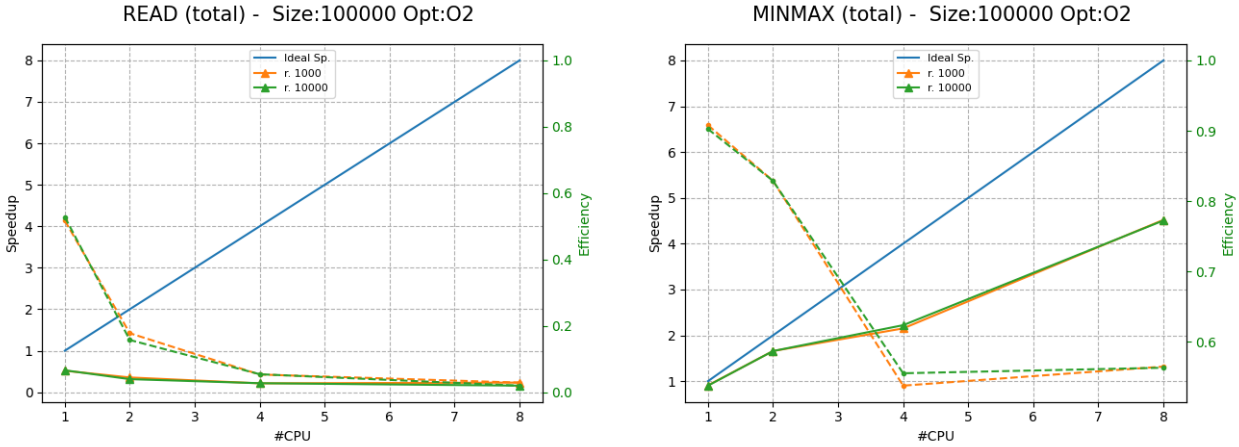


As we can see from the results, with this size of the array to be sorted it is not at all convenient to parallelize the algorithm in any case. This is because in none of the steps the size of the problem is such as to amortize the overhead introduced by MPI. Both versions of the algorithm show similar results, which means that for this size of the problem, no significant change is visible.

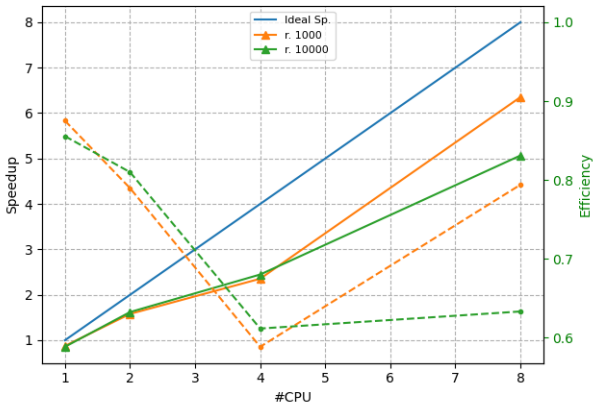
- **Read and Write:** these results are obtained launching the program on a single machine but, to achieve real increase of performances, it should be launched on a cluster where the accesses are not handled sequentially.
- **Minmax:** the overhead introduced for the reduction is too much for arrays of this size. That's why no significantly improvements are shown.
- **PMF:** the overhead is very high as extremely large arrays are allocated compared to the size of the problem (in the phase of counting the occurrences), one for each thread, bringing the speedup almost to 0.

5.2.2 Size 10^5

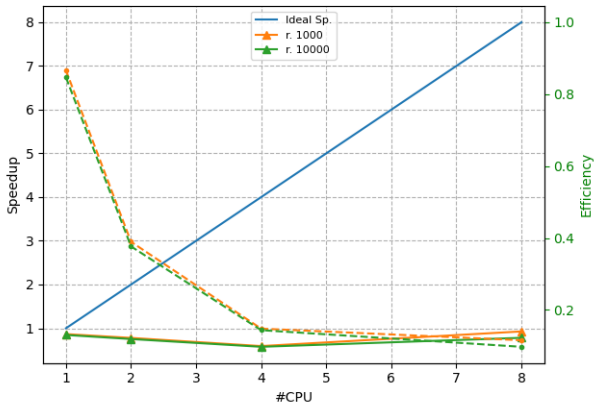
Version 1



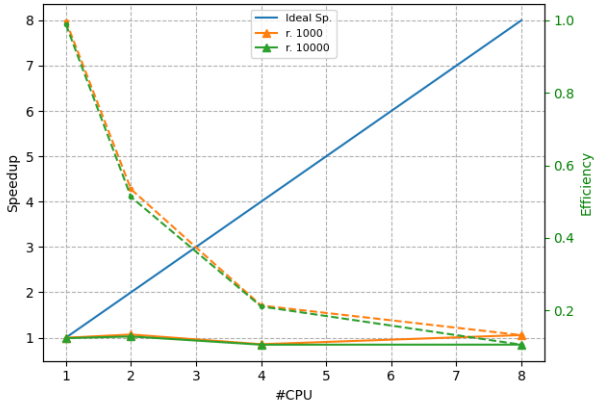
PMF (total) - Size:100000 Opt:O2



WRITE (total) - Size:100000 Opt:O2

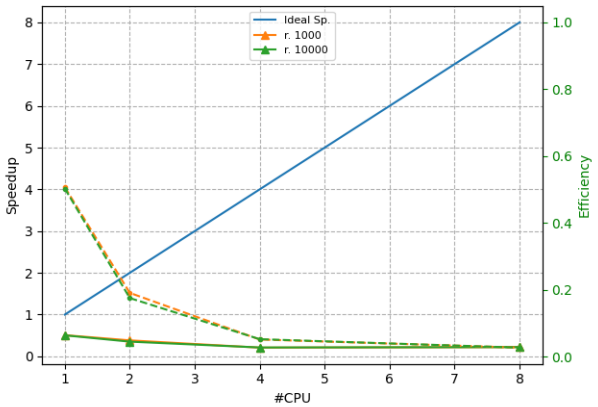


T_ALGO (total) - Size:100000 Opt:O2

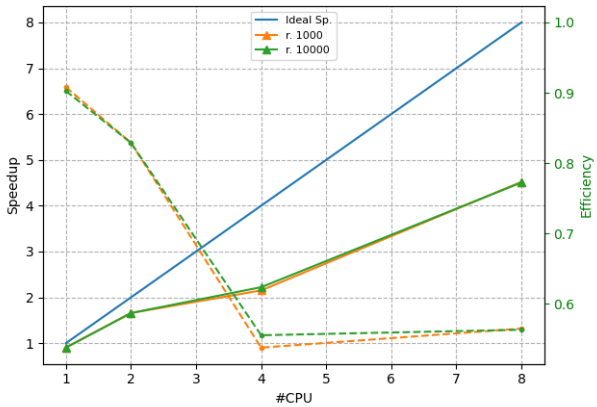


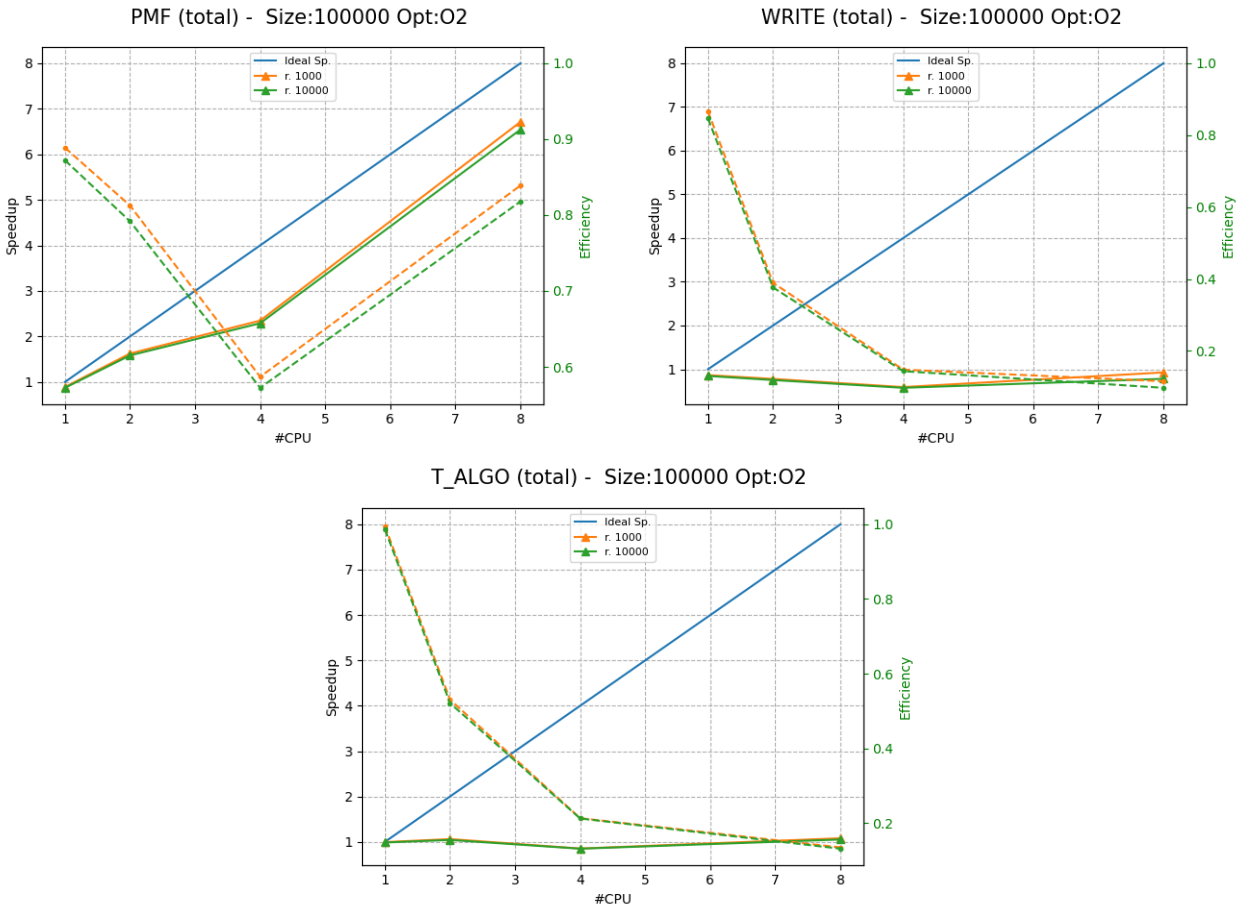
Version 2

READ (total) - Size:100000 Opt:O2



MINMAX (total) - Size:100000 Opt:O2





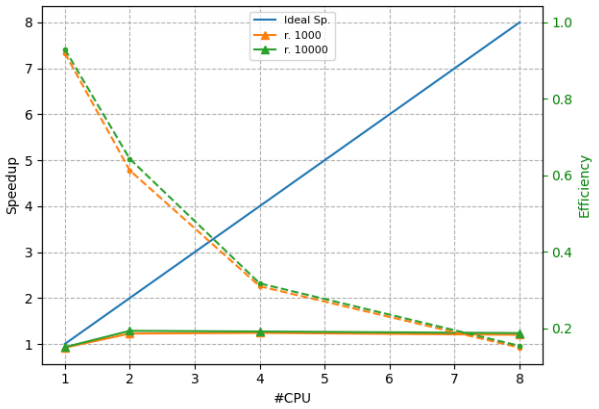
As the results show, increasing the size of the problem in both the algorithms is reducing the overhead because, as the size of the array increases, it justifies the overhead introduced by the generation of more threads. The read and write parts still don't show significant improvements because it's executed on a single machine, but the Minmax and PMF calculation achieve better performances, leading the speedup closer to the ideal.

You may notice problems in the plotted curves, especially when using 8 threads. This is because you are using a single machine and therefore, with 8 threads, these are running at least two for each core. Then they will share the cache and transfers, which are already faster than a cluster, will be even faster.

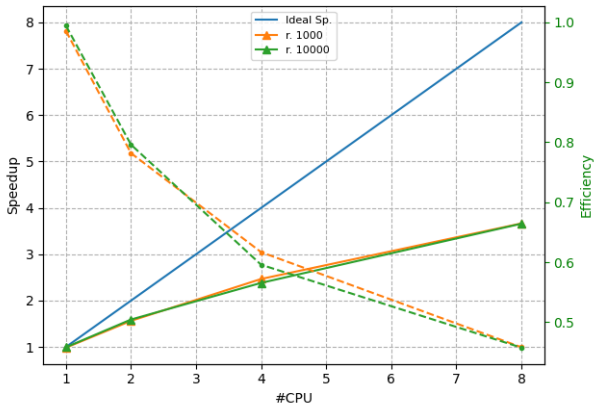
5.2.3 Size 10^7

Version 1

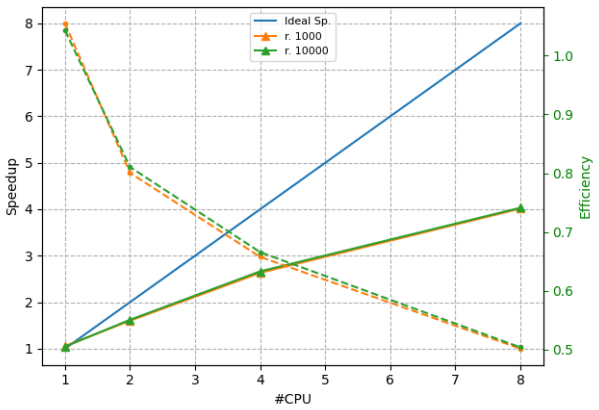
READ (total) - Size:10000000 Opt:O2



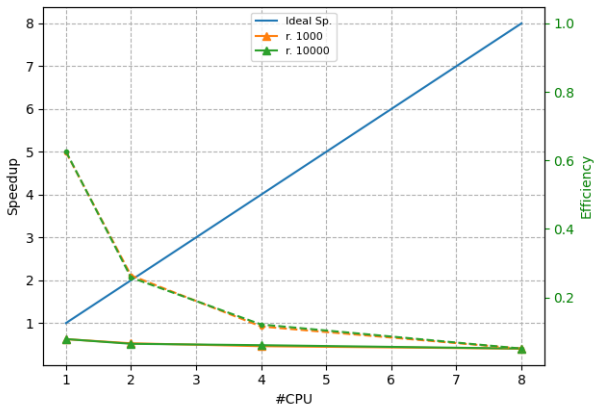
MINMAX (total) - Size:10000000 Opt:O2



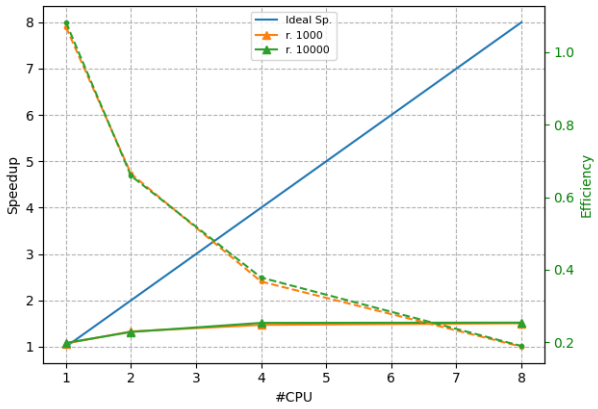
PMF (total) - Size:10000000 Opt:O2



WRITE (total) - Size:10000000 Opt:O2

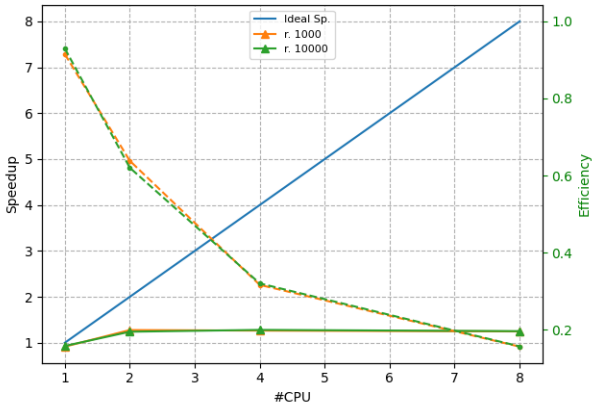


T_ALGO (total) - Size:10000000 Opt:O2

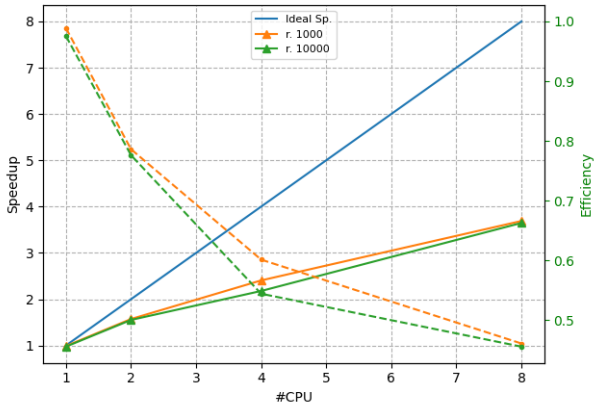


Version 2

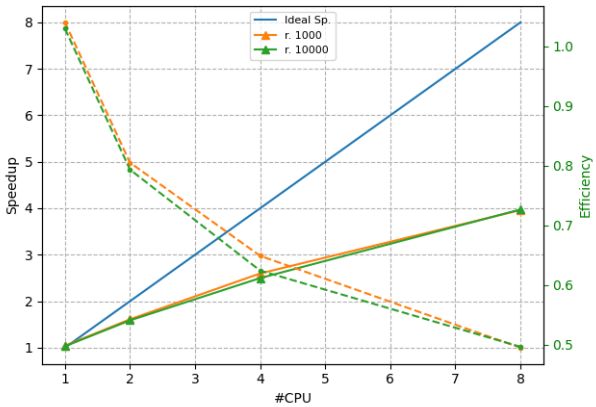
READ (total) - Size:10000000 Opt:O2



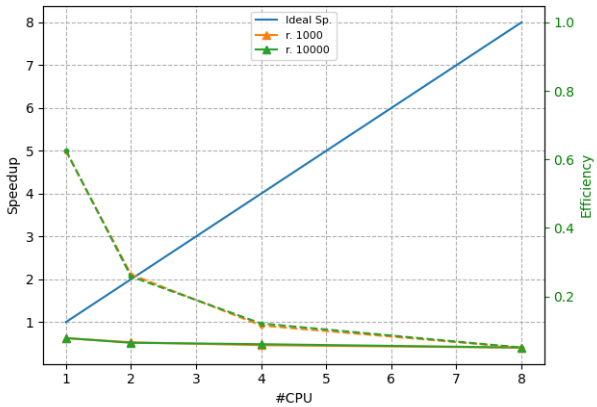
MINMAX (total) - Size:10000000 Opt:O2



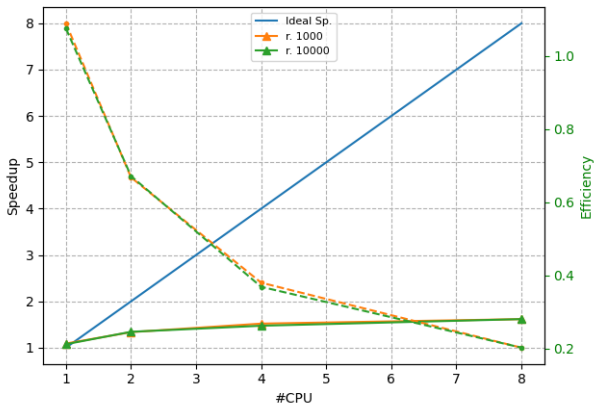
PMF (total) - Size:10000000 Opt:O2



WRITE (total) - Size:10000000 Opt:O2



T_ALGO (total) - Size:10000000 Opt:O2



With this dimension of the problem the improvement can be seen enough to justify the use of the algorithm.

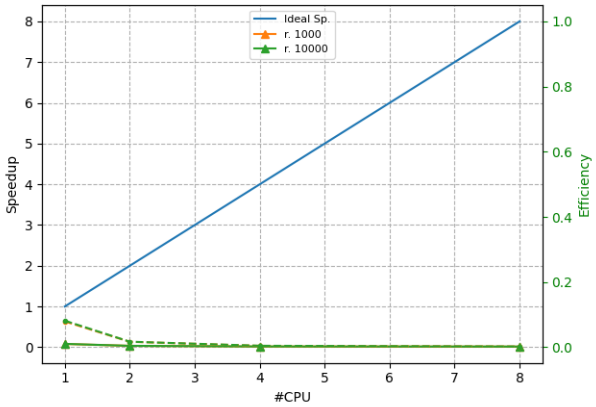
When n is greater than k by four orders of magnitude, improvements begin to appear, since the allocation of a vector of size k for each thread to count the occurrences and the relative reduction phase begin to be amortized. Moreover, they probably increase the performances of the maximum and minimum and PMF calculation.

5.3 Optimization O0

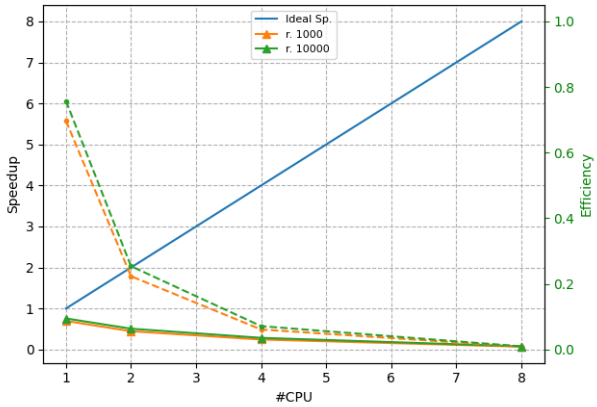
5.3.1 Size 10^3

Version 1

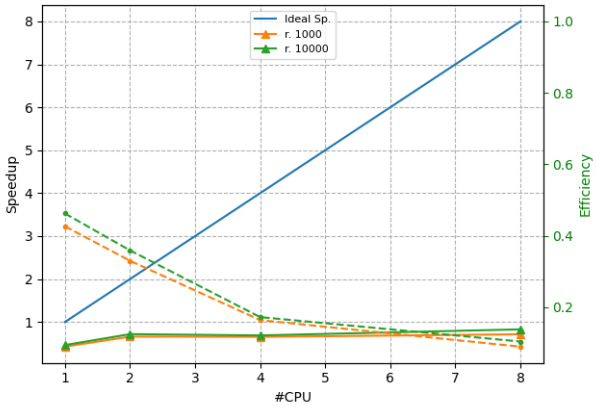
READ (total) - Size:1000 Opt:O0



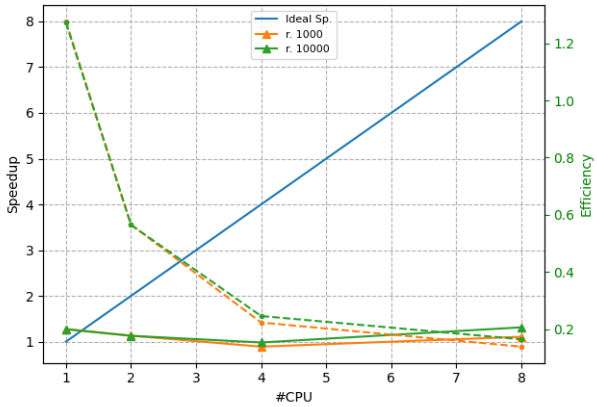
MINMAX (total) - Size:1000 Opt:O0



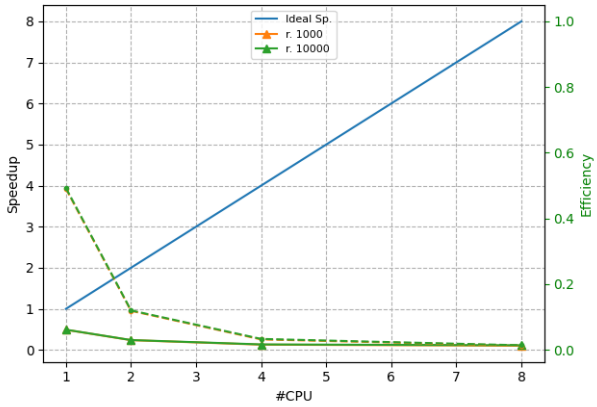
PMF (total) - Size:1000 Opt:O0



WRITE (total) - Size:1000 Opt:O0

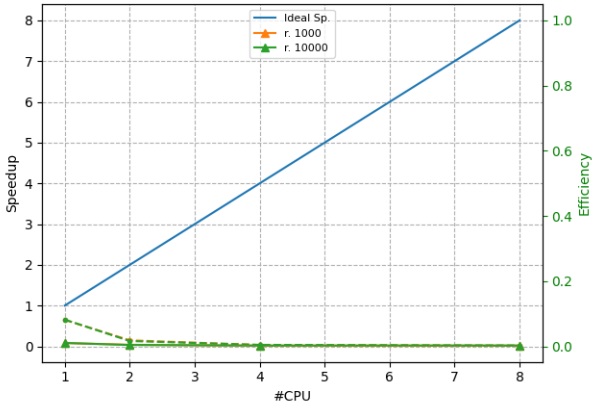


T_ALGO (total) - Size:1000 Opt:O0

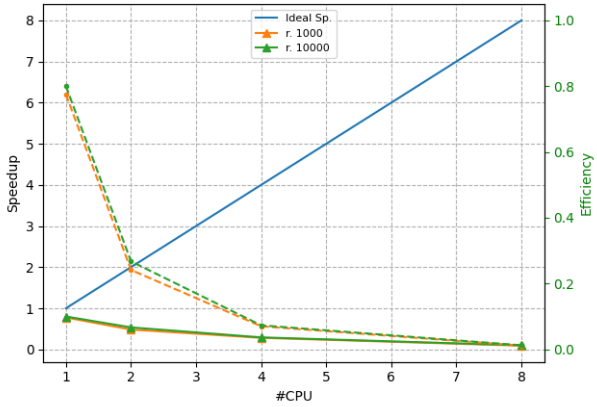


Version 2

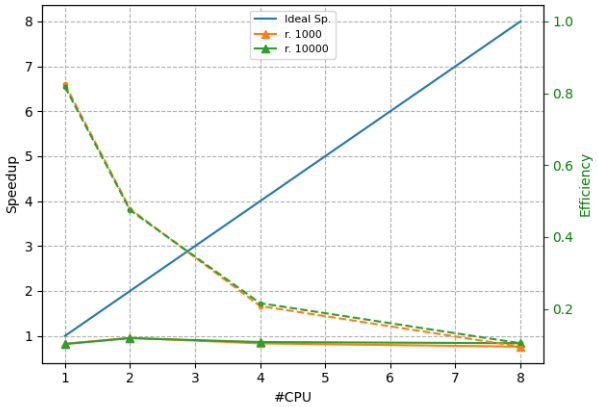
READ (total) - Size:1000 Opt:O0



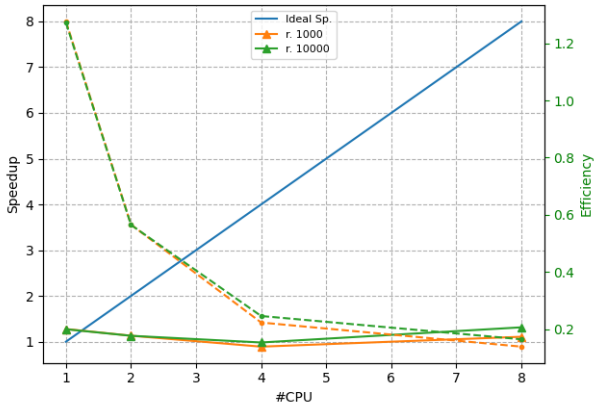
MINMAX (total) - Size:1000 Opt:O0



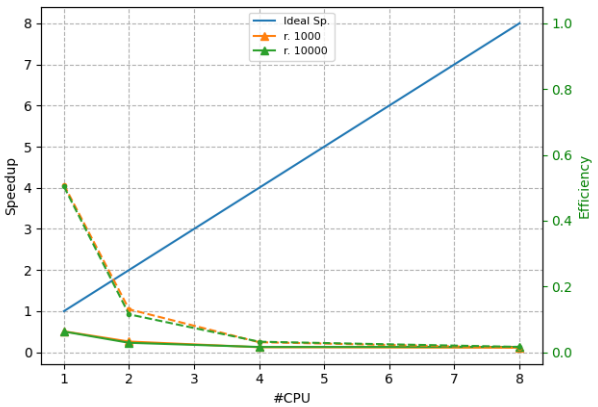
PMF (total) - Size:1000 Opt:O0



WRITE (total) - Size:1000 Opt:O0



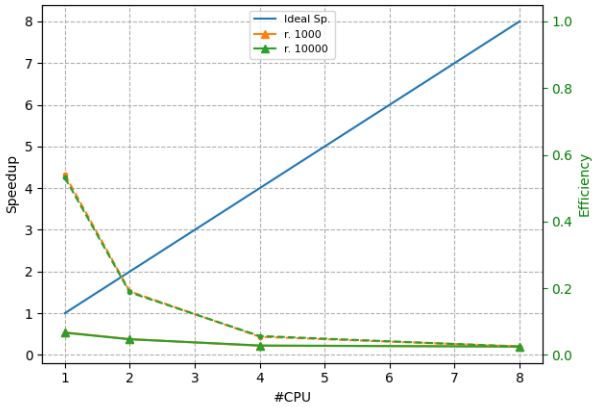
T_ALGO (total) - Size:1000 Opt:O0



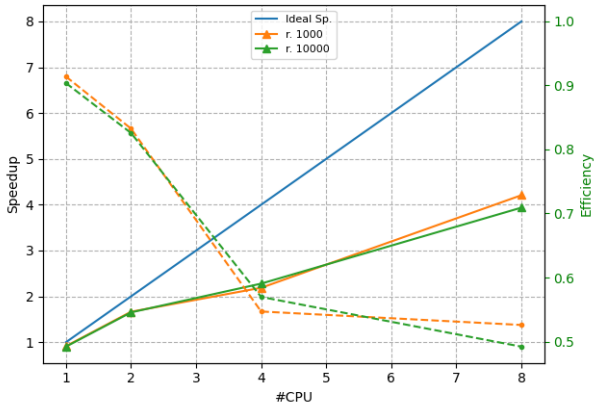
5.3.2 Size 10^5

Version 1

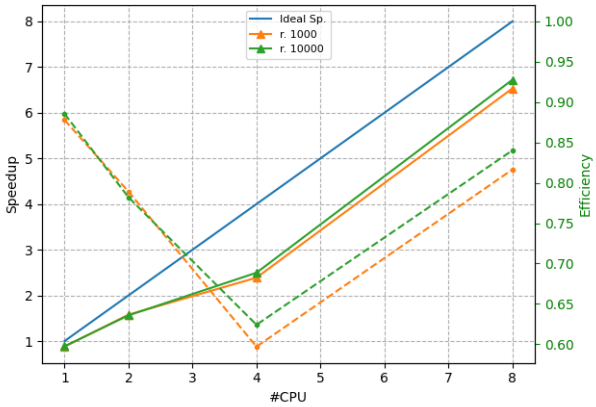
READ (total) - Size:100000 Opt:O0



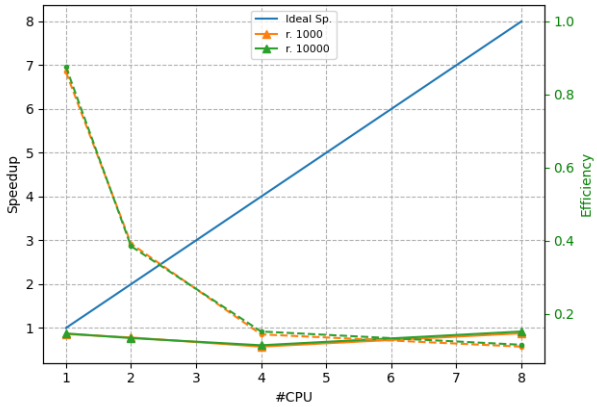
MINMAX (total) - Size:100000 Opt:O0



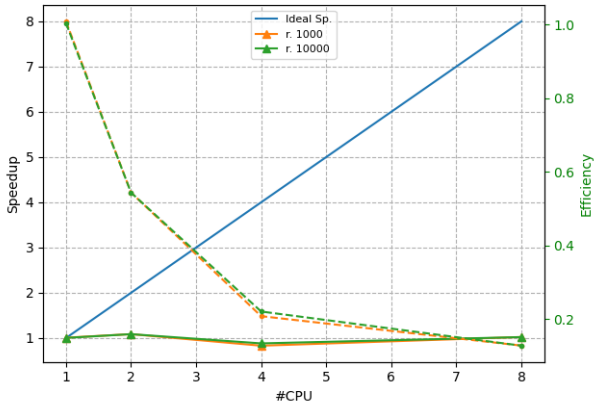
PMF (total) - Size:100000 Opt:O0



WRITE (total) - Size:100000 Opt:O0

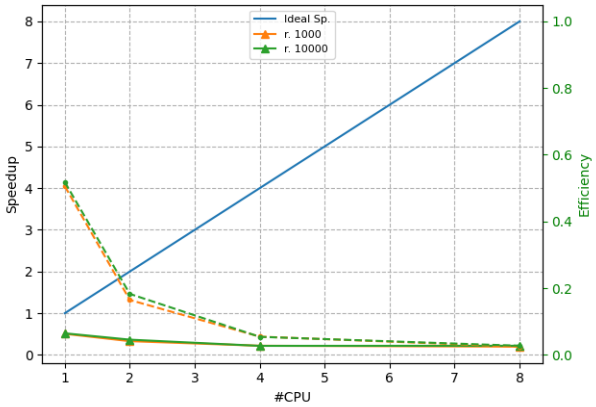


T_ALGO (total) - Size:100000 Opt:O0

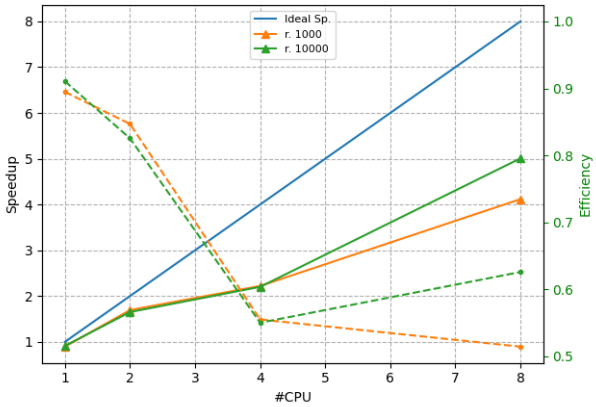


Version 2

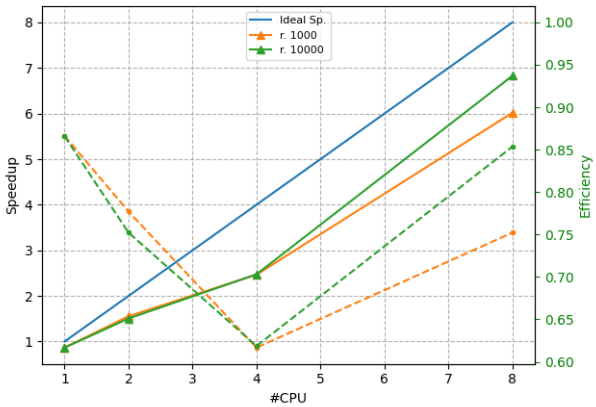
READ (total) - Size:100000 Opt:O0



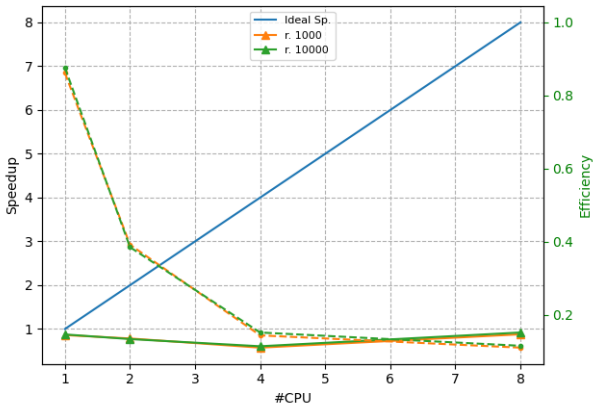
MINMAX (total) - Size:100000 Opt:O0



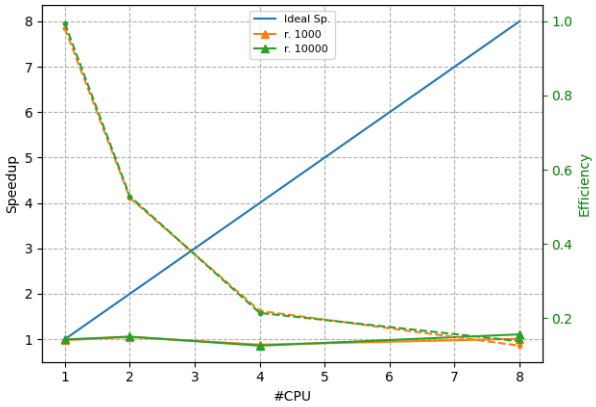
PMF (total) - Size:100000 Opt:O0



WRITE (total) - Size:100000 Opt:O0



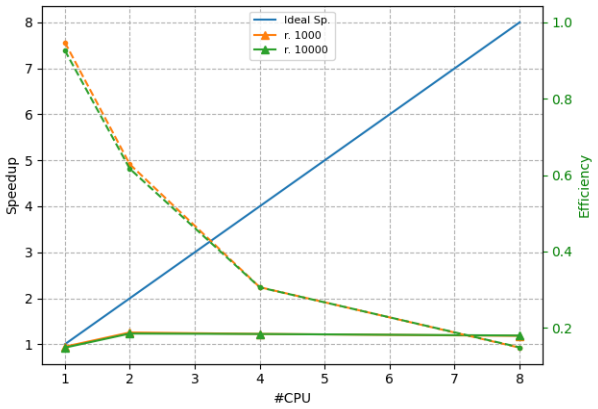
T_ALGO (total) - Size:100000 Opt:O0



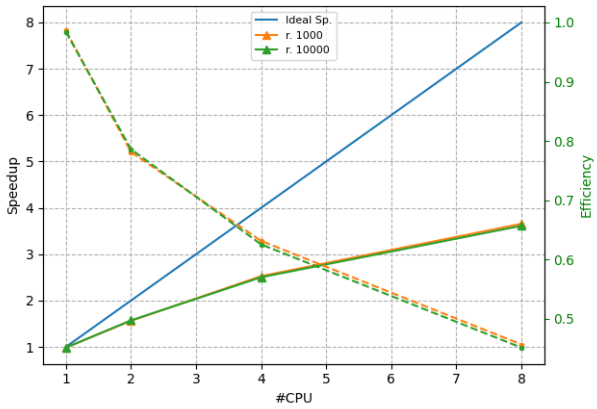
5.3.3 Size 10^7

Version 1

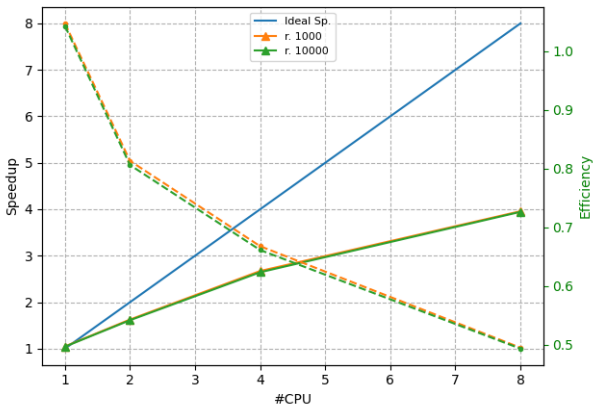
READ (total) - Size:10000000 Opt:O0



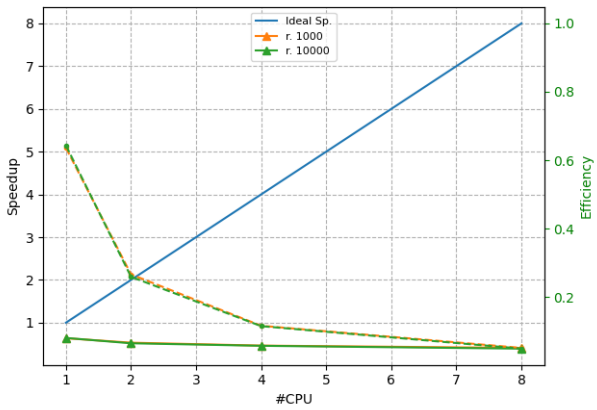
MINMAX (total) - Size:10000000 Opt:O0



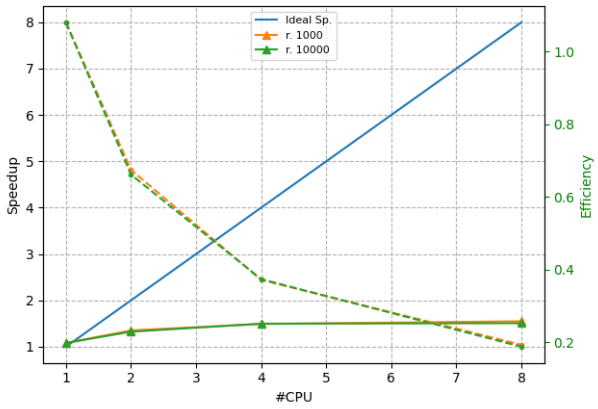
PMF (total) - Size:10000000 Opt:O0



WRITE (total) - Size:10000000 Opt:O0

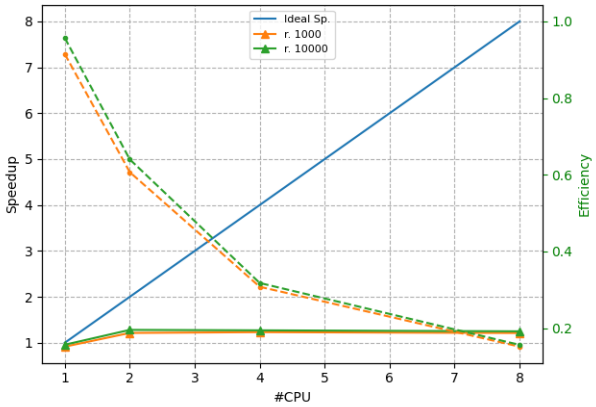


T_ALGO (total) - Size:10000000 Opt:O0

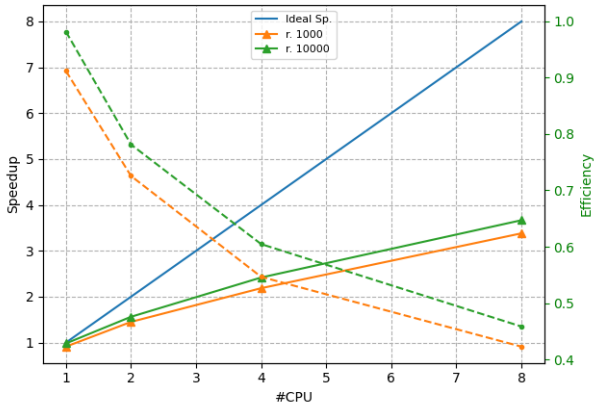


Version 2

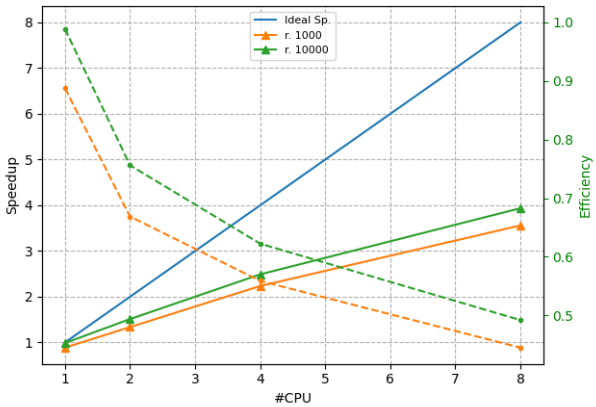
READ (total) - Size:10000000 Opt:O0



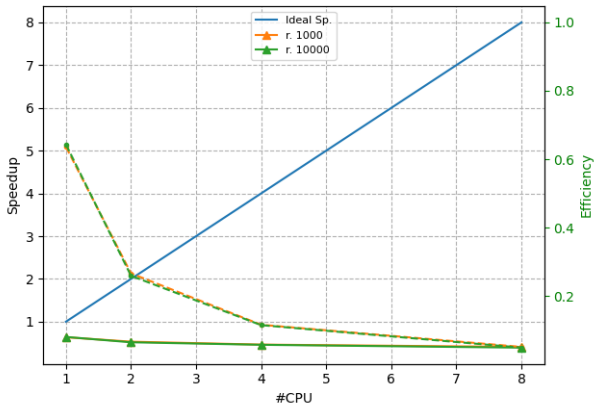
MINMAX (total) - Size:10000000 Opt:O0



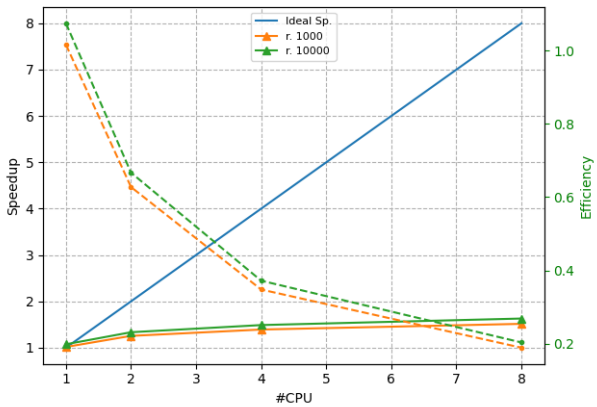
PMF (total) - Size:10000000 Opt:O0



WRITE (total) - Size:10000000 Opt:O0



T_ALGO (total) - Size:10000000 Opt:O0

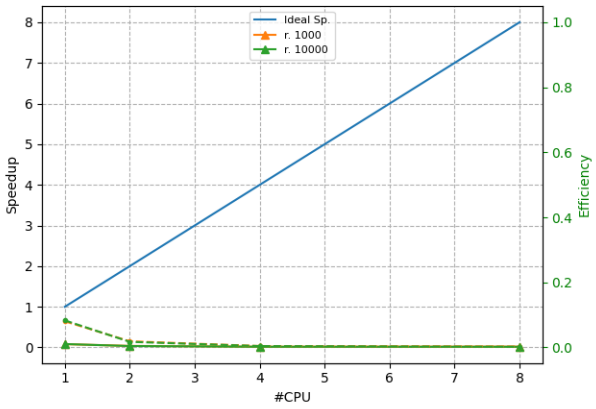


5.4 Optimization O1

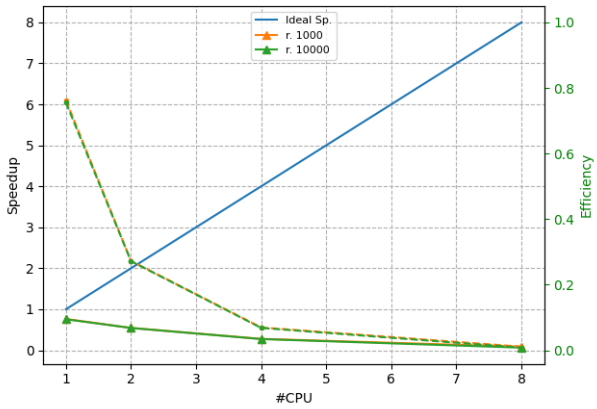
5.4.1 Size 10^3

Version 1

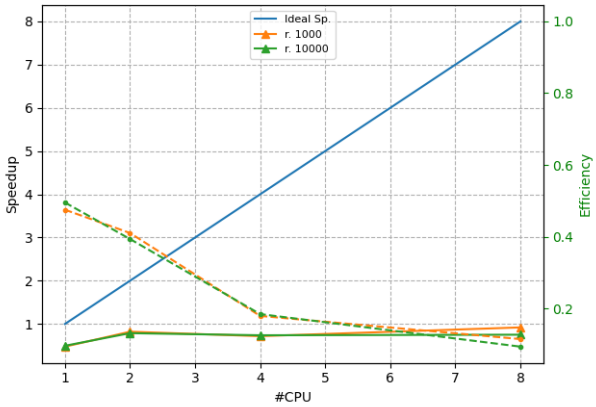
READ (total) - Size:1000 Opt:O1



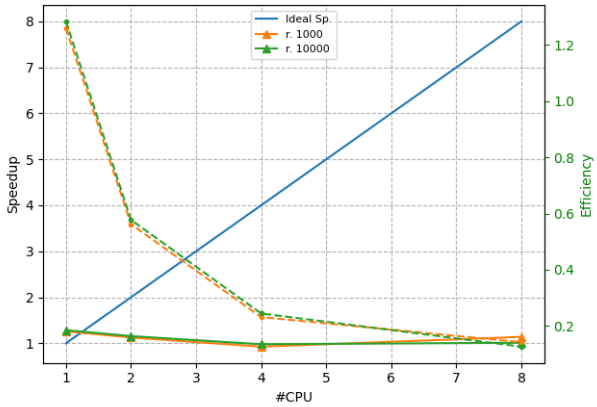
MINMAX (total) - Size:1000 Opt:O1



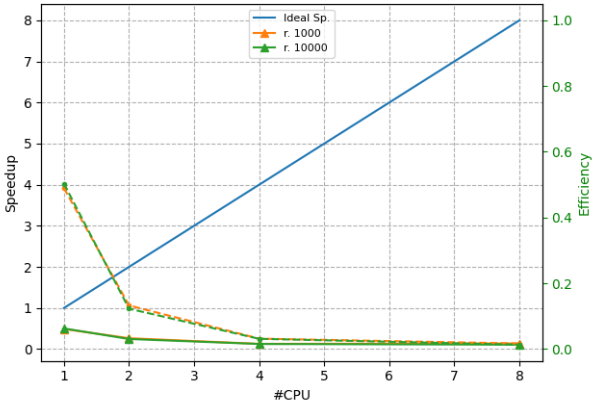
PMF (total) - Size:1000 Opt:O1



WRITE (total) - Size:1000 Opt:O1

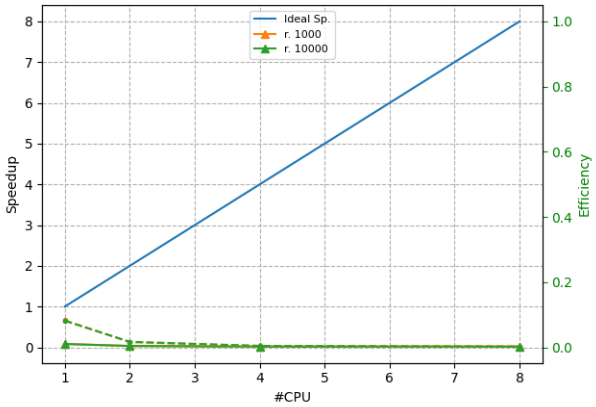


T_ALGO (total) - Size:1000 Opt:O1

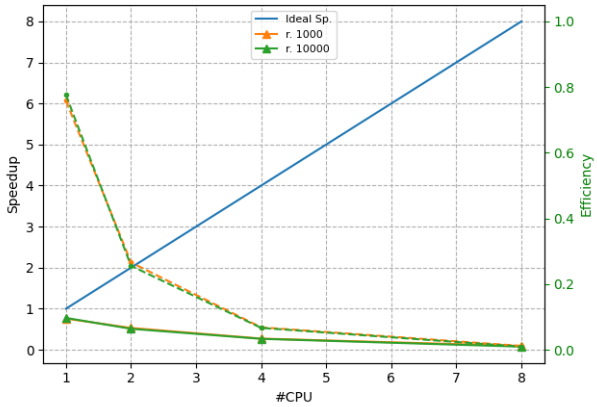


Version 2

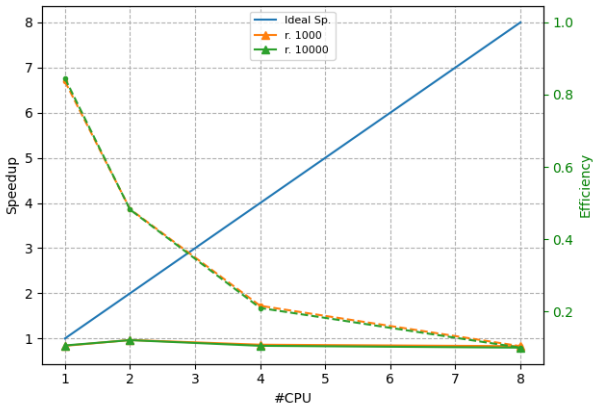
READ (total) - Size:1000 Opt:O1



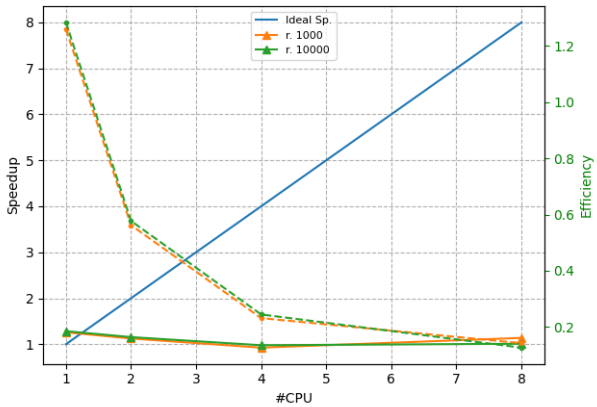
MINMAX (total) - Size:1000 Opt:O1



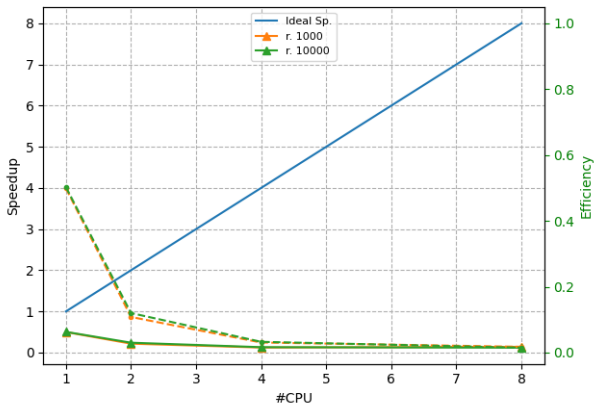
PMF (total) - Size:1000 Opt:O1



WRITE (total) - Size:1000 Opt:O1



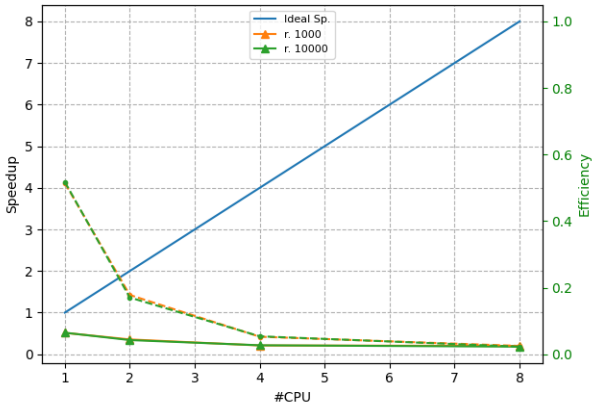
T_ALGO (total) - Size:1000 Opt:O1



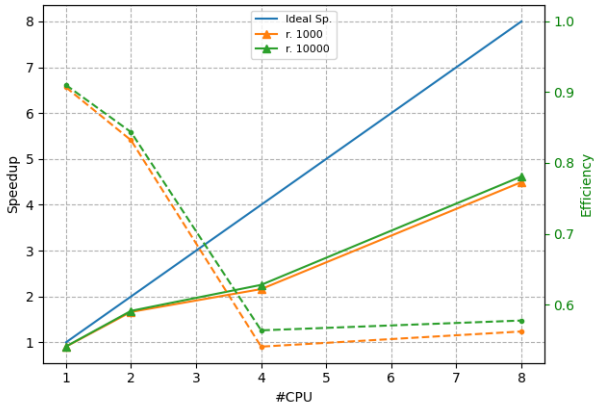
5.4.2 Size 10^5

Version 1

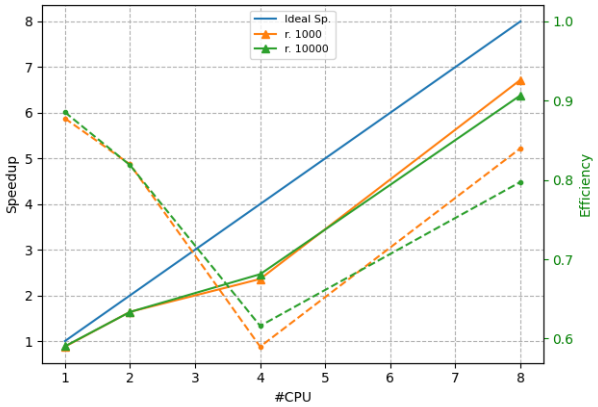
READ (total) - Size:100000 Opt:O1



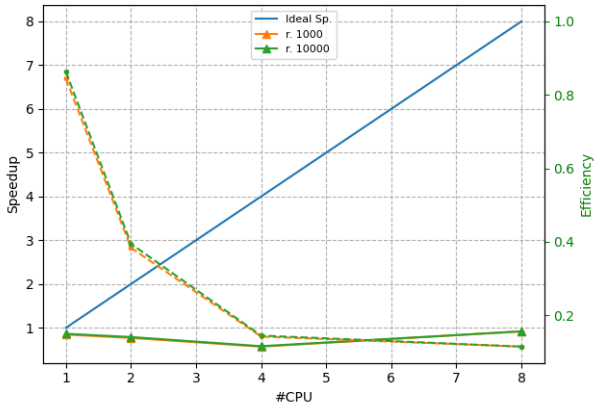
MINMAX (total) - Size:100000 Opt:O1



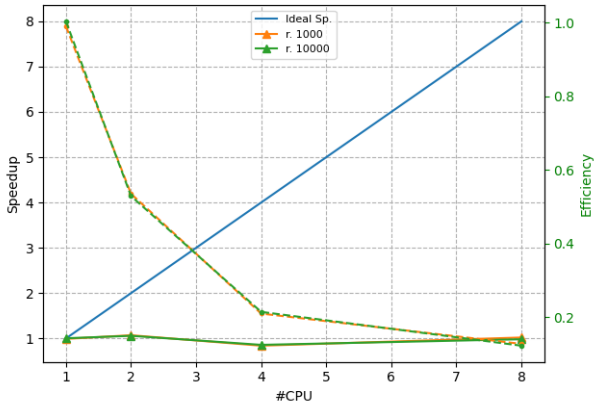
PMF (total) - Size:100000 Opt:O1



WRITE (total) - Size:100000 Opt:O1

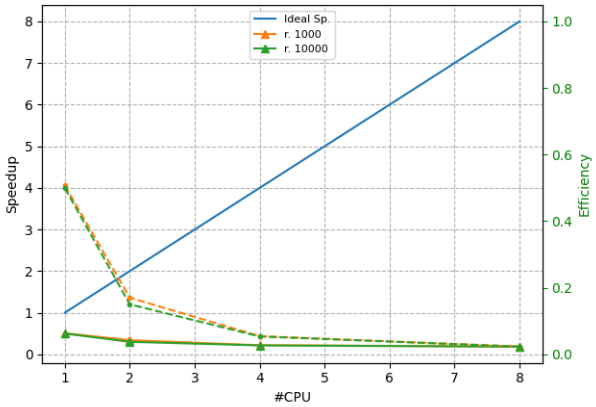


T_ALGO (total) - Size:100000 Opt:O1

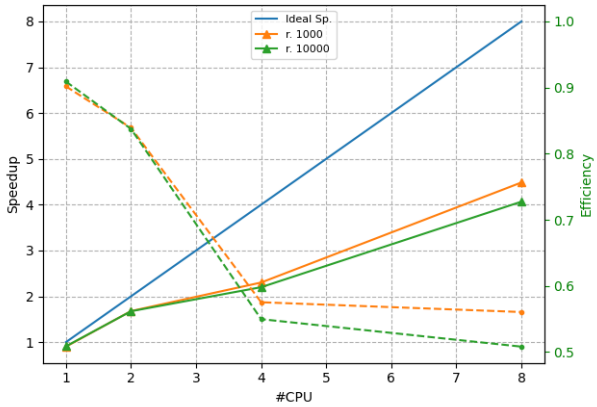


Version 2

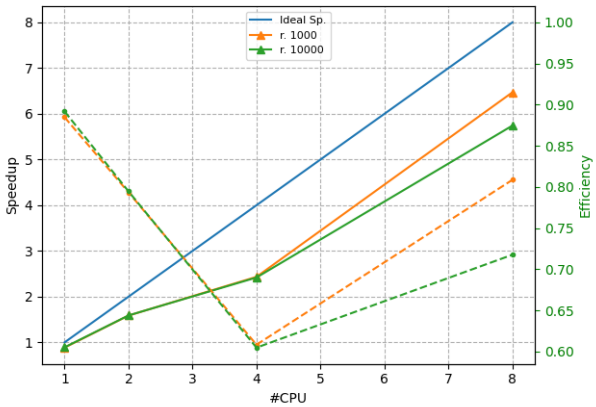
READ (total) - Size:100000 Opt:O1



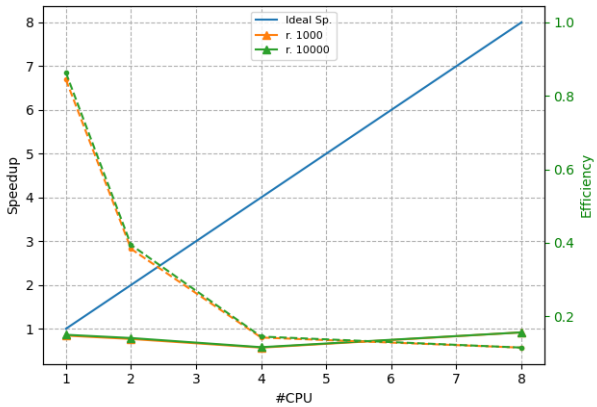
MINMAX (total) - Size:100000 Opt:O1



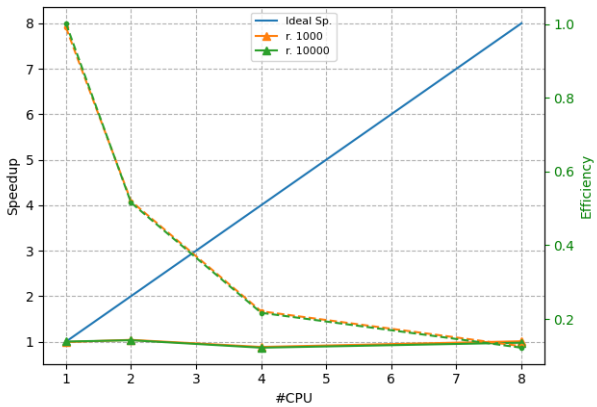
PMF (total) - Size:100000 Opt:O1



WRITE (total) - Size:100000 Opt:O1



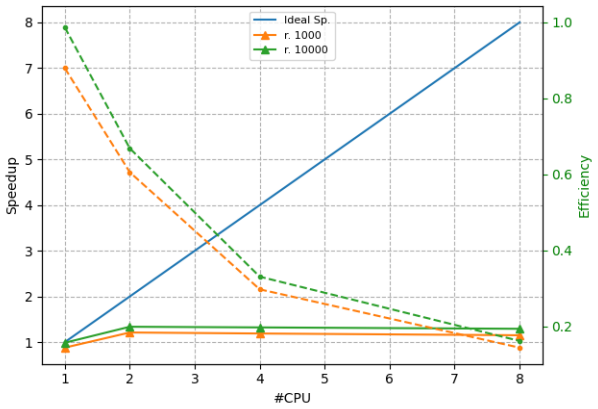
T_ALGO (total) - Size:100000 Opt:O1



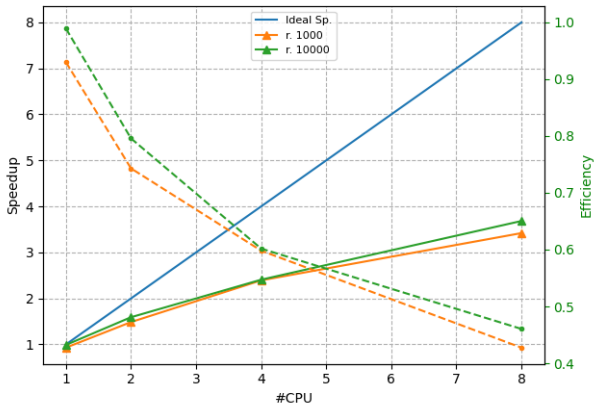
5.4.3 Size 10^7

Version 1

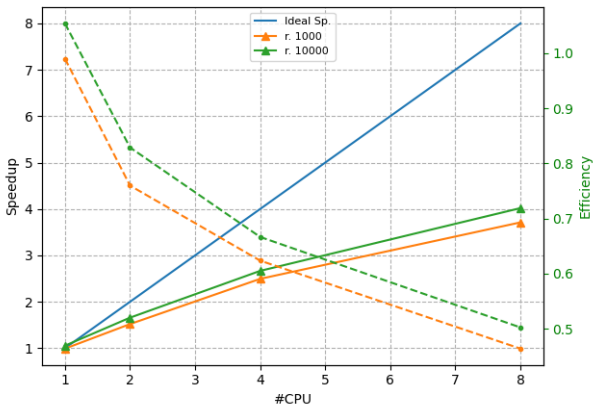
READ (total) - Size:10000000 Opt:O1



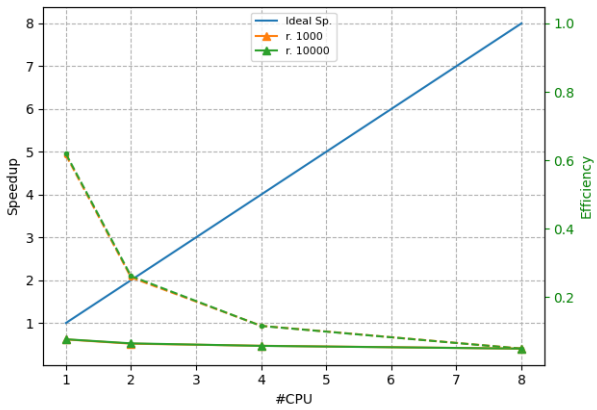
MINMAX (total) - Size:10000000 Opt:O1



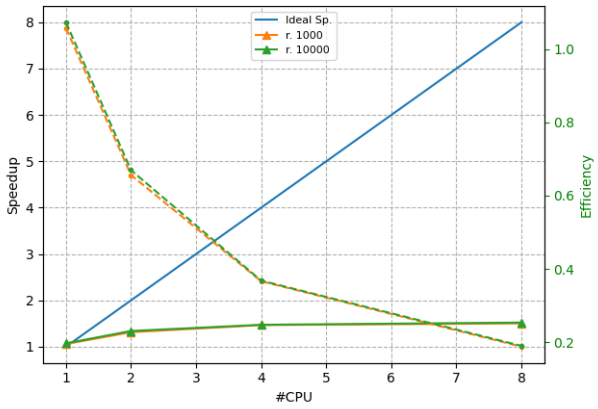
PMF (total) - Size:10000000 Opt:O1



WRITE (total) - Size:10000000 Opt:O1

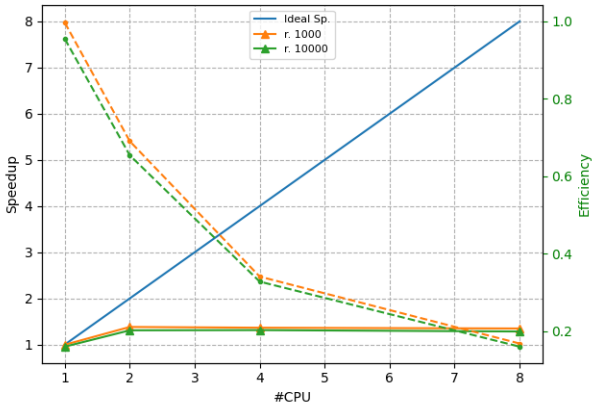


T_ALGO (total) - Size:10000000 Opt:O1

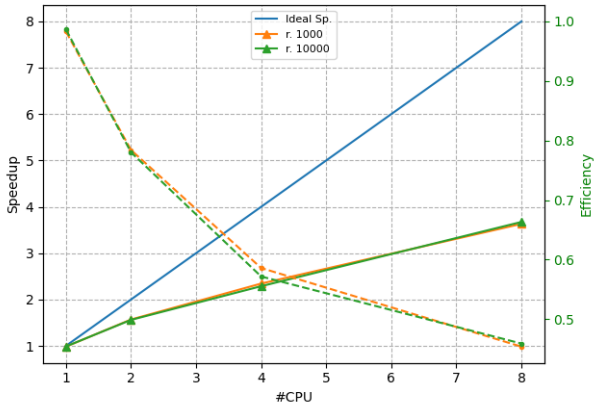


Version 2

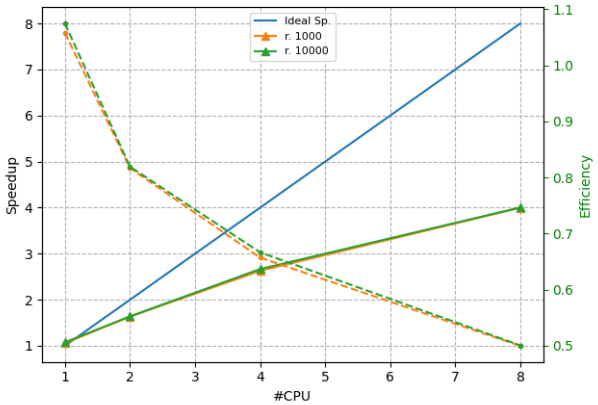
READ (total) - Size:10000000 Opt:O1



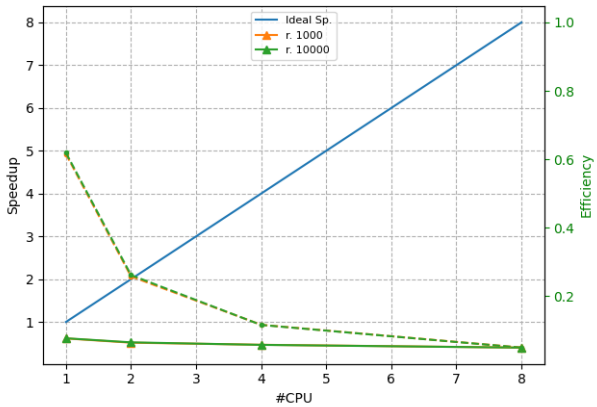
MINMAX (total) - Size:10000000 Opt:O1



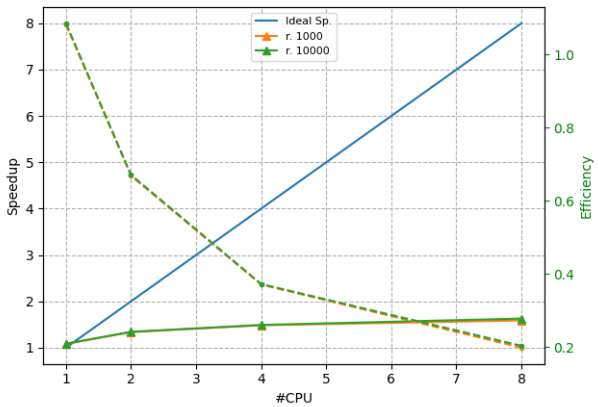
PMF (total) - Size:10000000 Opt:O1



WRITE (total) - Size:10000000 Opt:O1



T_ALGO (total) - Size:10000000 Opt:O1

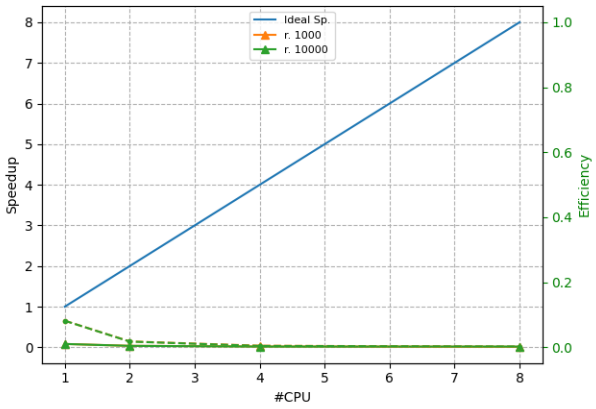


5.5 Optimization O3

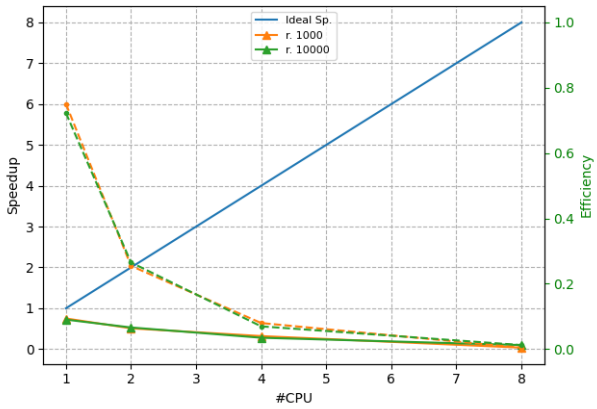
5.5.1 Size 10^3

Version 1

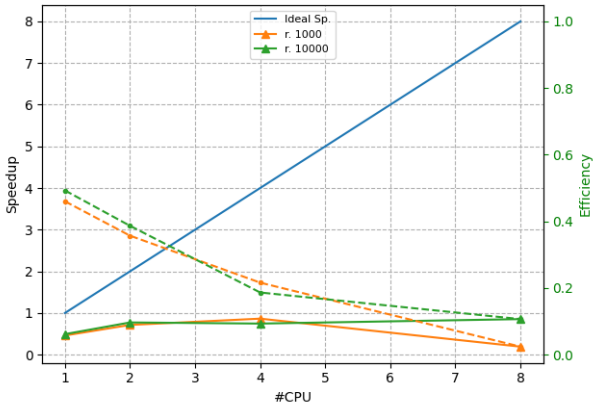
READ (total) - Size:1000 Opt:O3



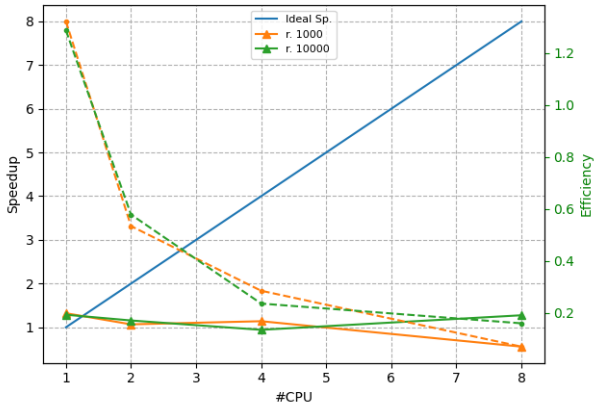
MINMAX (total) - Size:1000 Opt:O3



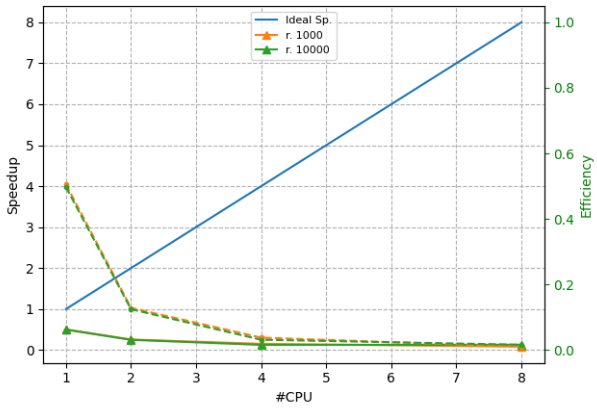
PMF (total) - Size:1000 Opt:O3



WRITE (total) - Size:1000 Opt:O3

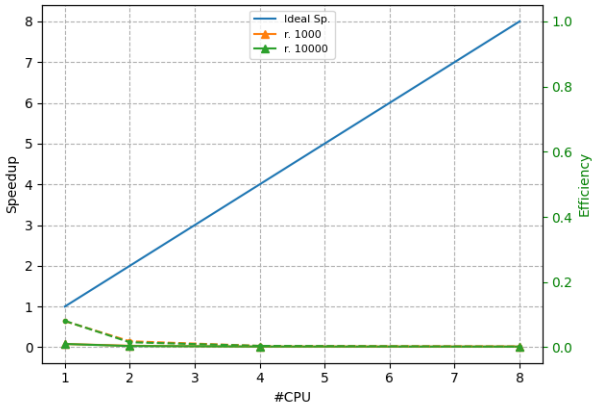


T_ALGO (total) - Size:1000 Opt:O3

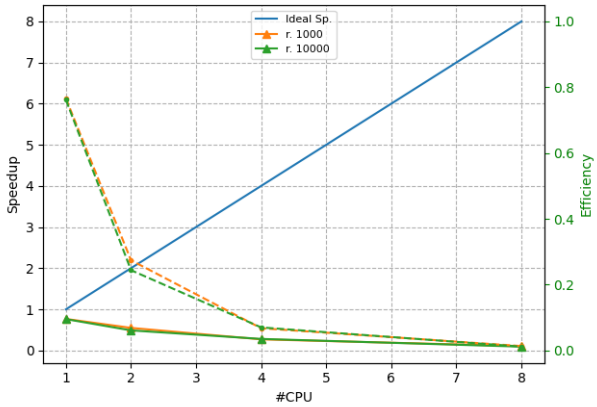


Version 2

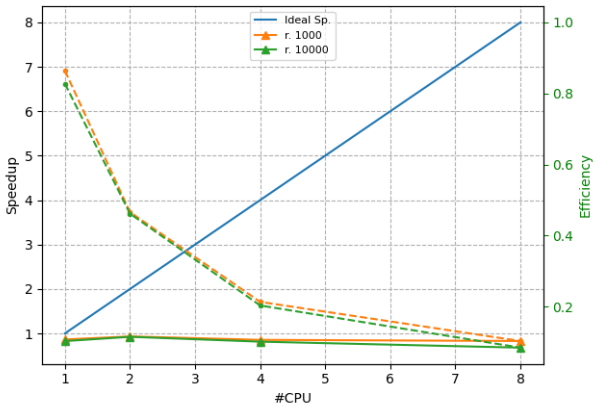
READ (total) - Size:1000 Opt:O3



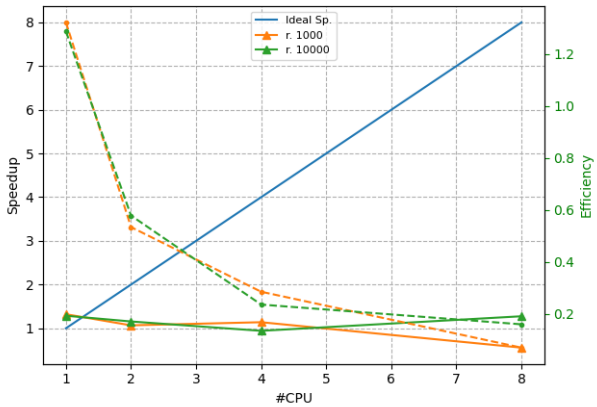
MINMAX (total) - Size:1000 Opt:O3



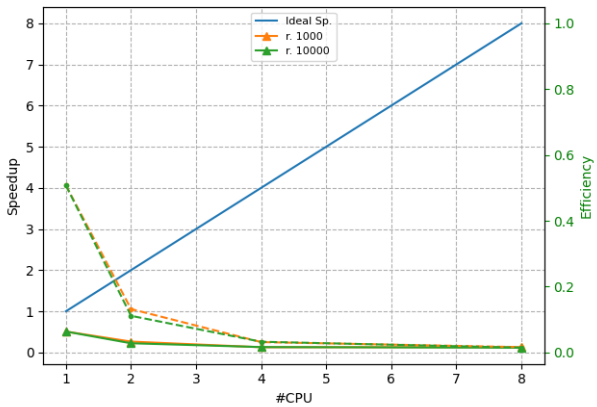
PMF (total) - Size:1000 Opt:O3



WRITE (total) - Size:1000 Opt:O3



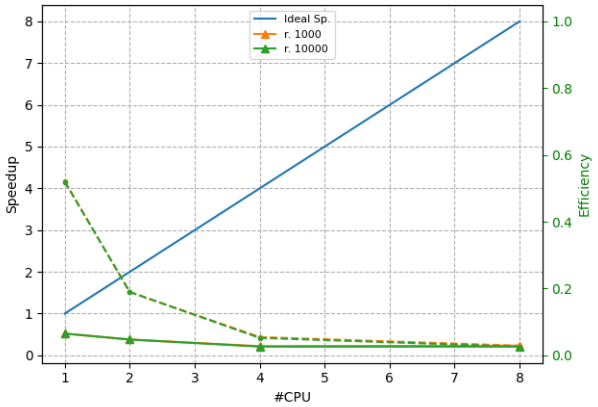
T_ALGO (total) - Size:1000 Opt:O3



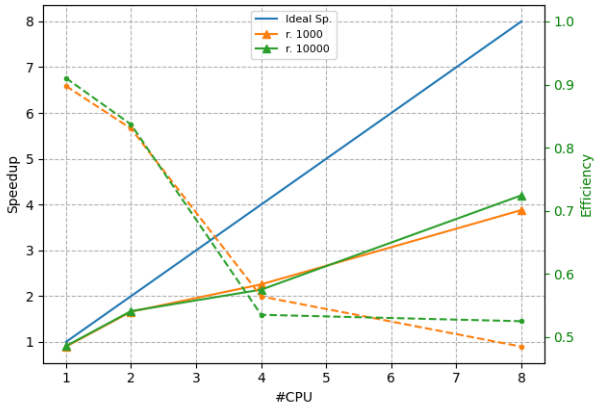
5.5.2 Size 10^5

Version 1

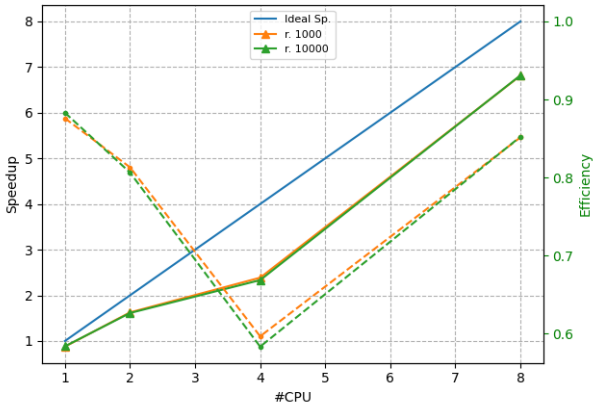
READ (total) - Size:100000 Opt:O3



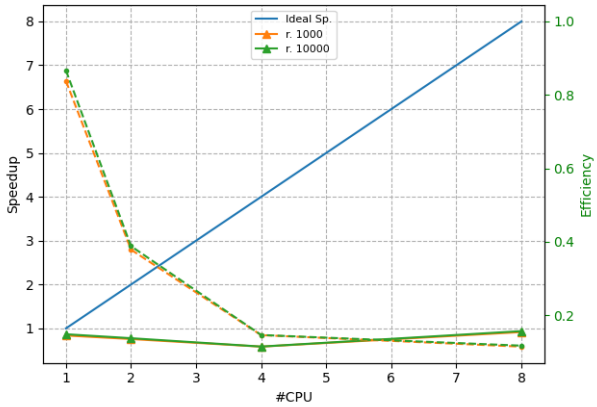
MINMAX (total) - Size:100000 Opt:O3



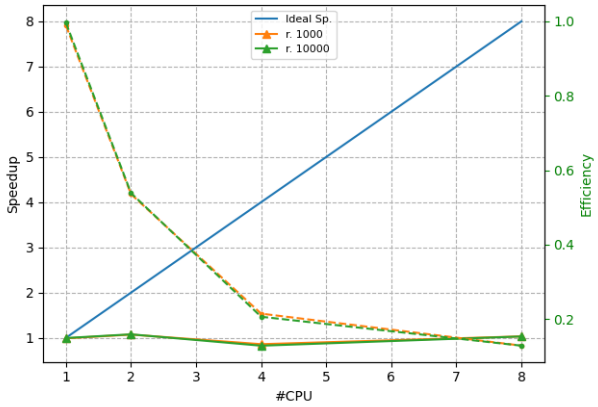
PMF (total) - Size:100000 Opt:O3



WRITE (total) - Size:100000 Opt:O3

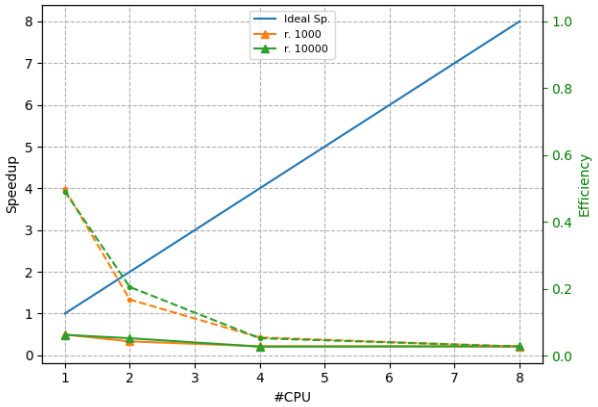


T_ALGO (total) - Size:100000 Opt:O3

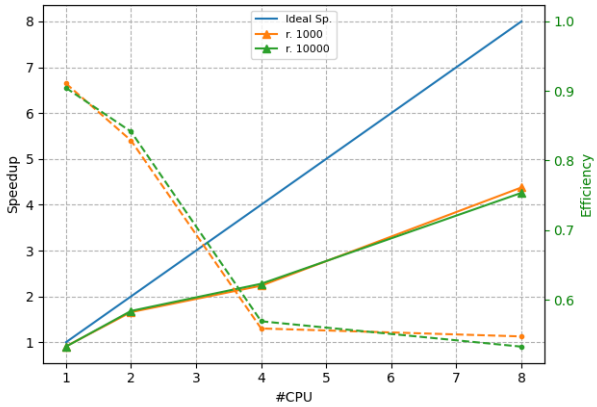


Version 2

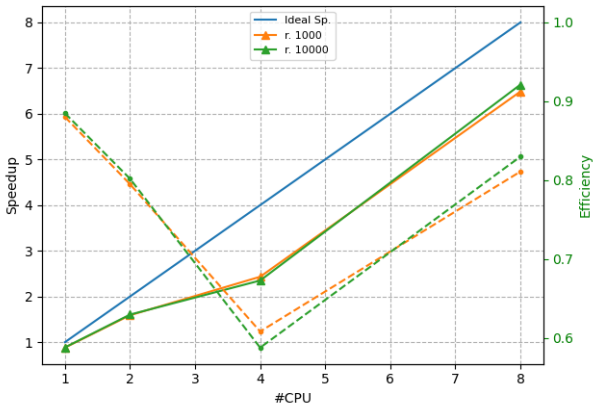
READ (total) - Size:100000 Opt:O3



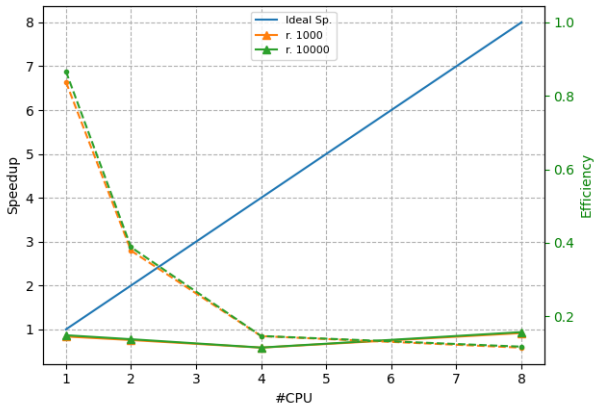
MINMAX (total) - Size:100000 Opt:O3



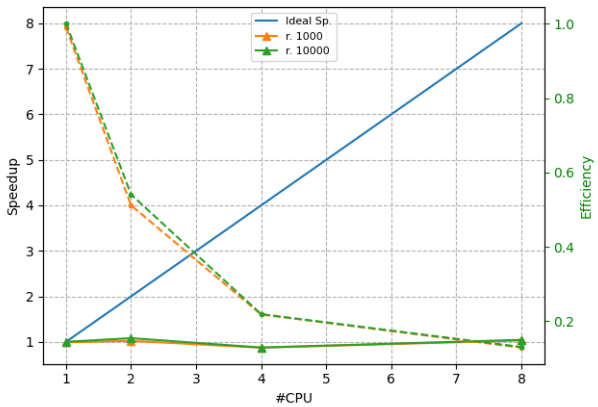
PMF (total) - Size:100000 Opt:O3



WRITE (total) - Size:100000 Opt:O3



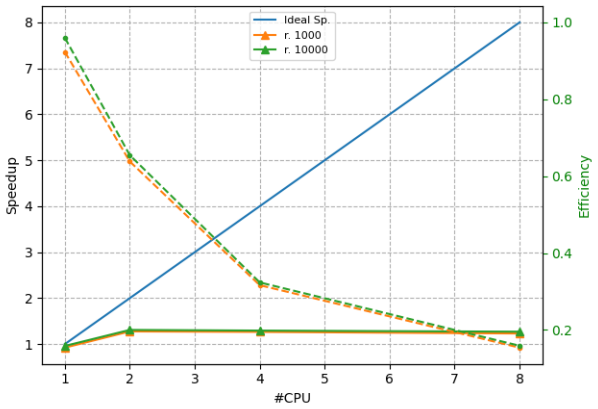
T_ALGO (total) - Size:100000 Opt:O3



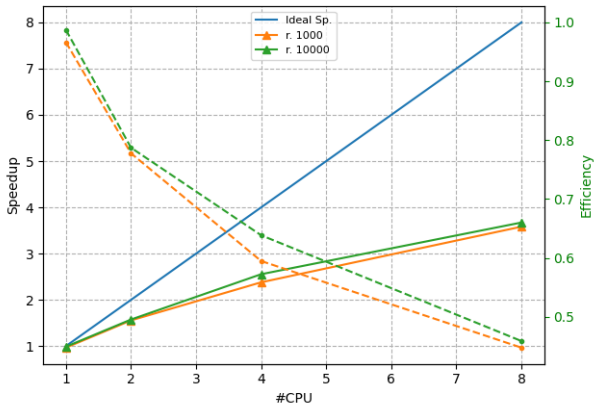
5.5.3 Size 10^7

Version 1

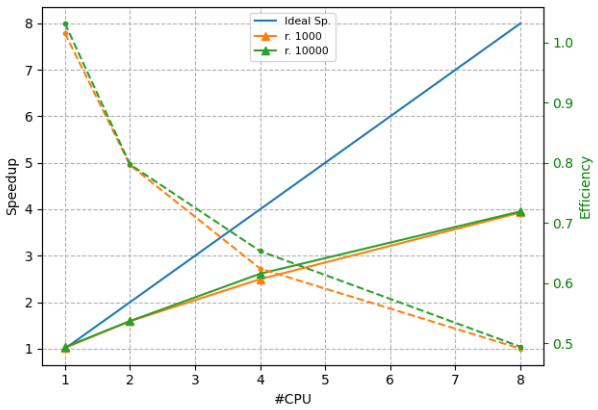
READ (total) - Size:10000000 Opt:O3



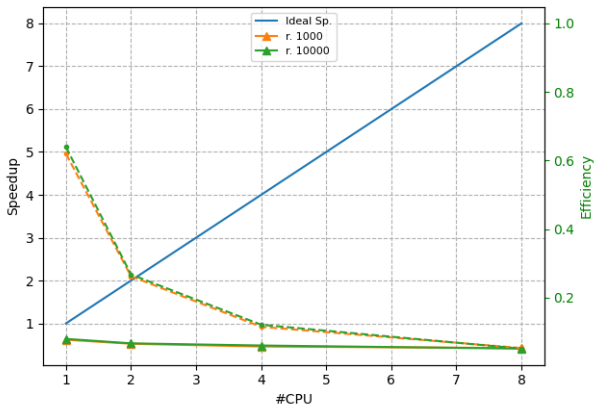
MINMAX (total) - Size:10000000 Opt:O3



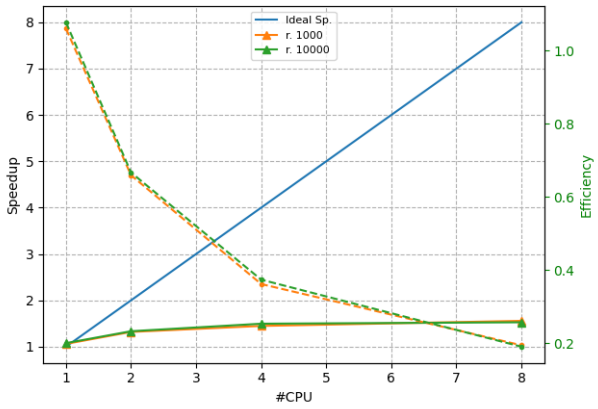
PMF (total) - Size:10000000 Opt:O3



WRITE (total) - Size:10000000 Opt:O3

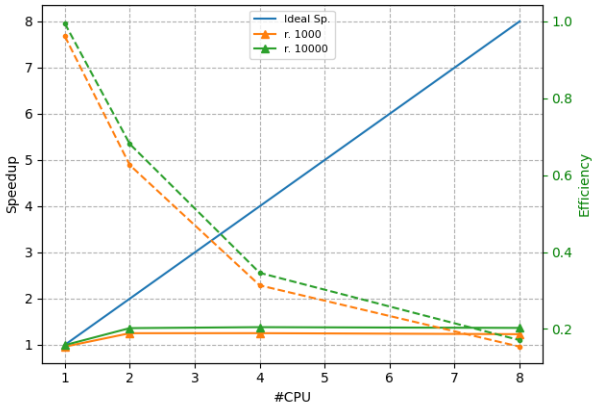


T_ALGO (total) - Size:10000000 Opt:O3

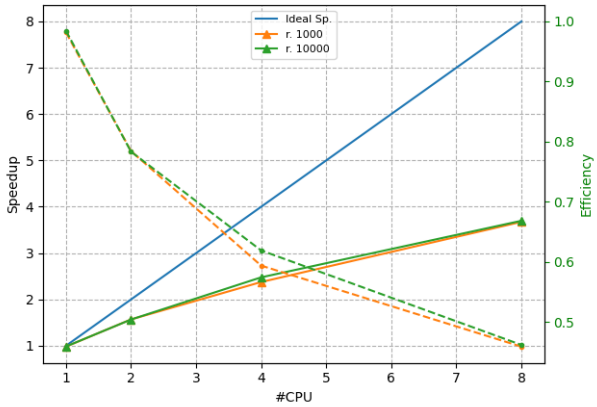


Version 2

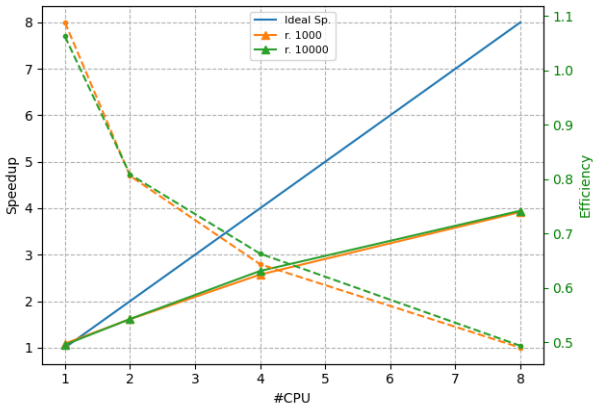
READ (total) - Size:10000000 Opt:O3



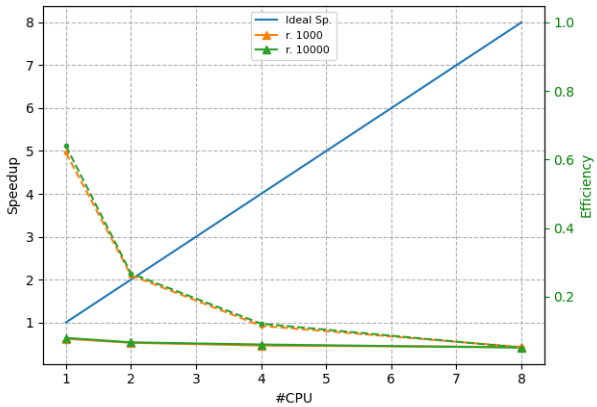
MINMAX (total) - Size:10000000 Opt:O3



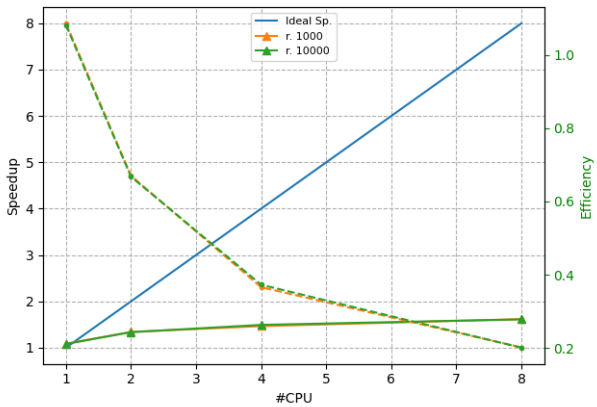
PMF (total) - Size:10000000 Opt:O3



WRITE (total) - Size:10000000 Opt:O3



T_ALGO (total) - Size:10000000 Opt:O3



5.6 Final considerations

Considering the two versions analyzed, we can see that there are not many differences in performance between the two. For this, considering that the transfers in version 2 are larger, we decide that version 1 is better and therefore the one to use.

We see that the number of threads suitable to use, on this setup (not considering more than 4 threads as there would be no coherence) for an array of size 10^5 it would be correct to use 2 threads, while on a size of 10^7 it would be correct to also use 4. It is difficult to estimate what can happen on a real cluster as we cannot estimate the time of transfers. However, we can certainly say that as the order of magnitude of the vector size increases, it is better to use more threads.

6 API

The APIs provided in the files are generated at compile time. Here is a copy.

6.1 util.h

This library provides the functions to generate random arrays and a macro which, if defined, always generates the same arrays by fixing the seed. The STARTTIME and ENDTIME macros are also defined.

◆ generate_rand_vector()

```
void generate_rand_vector ( const char *  file_name,
                            size_t        len,
                            long          min_value,
                            long          max_value
                          )
```

Initialize an array of a certain lenght with random integers in a certain range.

Parameters

file_name	The file name into write the array.
len	The desidered lenght of the array.
min_value	The minimum value inside the array.
max_value	The maximum value inside the array.

6.2 counting_sort.h

This library provides the functions to use the implemented counting sort algorithm, sequential and parallel, respectively.

◆ counting_sort()

```
void counting_sort ( const char *  file_name,
                     size_t        len
                   )
```

Reorder the array using a counter-sort alghorithm.

Parameters

file_name	The file name into write the array.
len	The lenght to the array to reorder

◆ counting_sort_mpi()

```
void counting_sort_mpi ( const char *  fileName,
                         size_t        arrayLen
                       )
```

Implementation of counting sort using MPI. Order an array writed in a byte file.

Parameters

fileName	The path to the file containing an array of integert to order (in
arrayLen	The lenght of the array in the file.

6.3 main_measures.c

This is the file used to launch the benchmarks you've seen in the previous sections.

6.4 generate.c

This file contains a main to generate a random vector in a file.

7 Test Case

A total of 8 test cases were provided, 4 of which in the sequential version and 4 in the parallel version. Both run both compiled with OpenMP and not.

```
unsigned short empty_test();
unsigned short pempty_test();
unsigned short test1();
unsigned short test2();
unsigned short test3();
unsigned short ptest1();
unsigned short ptest2();
unsigned short ptest3();
```

- `empty_test` and `pempty_test` sort on an empty array. The expected result is that no errors occur.
- `test1` and `ptest1` perform a test on a random array and verify its ordering
- `test2` and `ptest2` perform a test where the ordered vector is known in advance and check the actual operation.
- `test` and `ptest3` perform the same test case as the previous ones, but on larger arrays.

Tests pass by all algorithm combinations at 100%.

License

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.