

PART 1. Performance timing and analysis using MPI.

Objective: To find the minimum data size for which it is useful to parallelize this program on a four-core system such as USS.

Methodology: First, I attempted get a better understanding of the data size by using a small data size of 4, medium data size of 40, large data size of 400 on different processors. I kept the parameters of bucket, min, and max measurements constant through my testing because I wanted to select a data size first. Additionally, I used MPI_Wtime() from textbook section 3.6.1 to measure the time taken to compute.

Table 1: Small data size of 4 on different cores.

	1 processor 4.321638e+00 sec	2 processors 1.656019e+01 sec	4 processors 1.125786e+01 sec
1.000-1.900:			
1.900-2.800:			
2.800-3.700:			
3.700-4.600:	X	X	X
4.600-5.500:			
5.500-6.400:			
6.400-7.300:			
7.300-8.200:	XX	XX	XX
8.200-9.100:	X	X	X
9.100-10.000:			

Observation: The data is same for all the processors. The elapsed time for 1 processor was 4.321638e+00 seconds. The elapsed time for 2 processors was 1.656019e+01 seconds. The elapsed time for 4 processors was 1.125786e+01 seconds. It is clear, the more processors the time is more efficient.

Table 2: Medium data size of 40 on different cores.

	1 processor 2.161180e+02 sec	2 processors 3.433713e+00 sec	4 processors 5.260822e+00 sec
1.000-1.900:	X	X	X
1.900-2.800:	XXXXXX	XXXXXX	XXXXXX
2.800-3.700:	XXXXX	XXXXX	XXXXX
3.700-4.600:	XXX	XXX	XXX
4.600-5.500:	XXX	XXX	XXX
5.500-6.400:	XXXX	XXXX	XXXX
6.400-7.300:	XXXXX	XXXXX	XXXXX
7.300-8.200:	XXXXX	XXXXX	XXXXX
8.200-9.100:	XXX	XXX	XXX
9.100-10.000:	XXXXX	XXXXX	XXXXX

Observation: Again, the data is same for all the processors. The elapsed time for 1 processor was 2.161180e+02 seconds. The elapsed time for 2 processors was 3.433713e+00 seconds. The elapsed time for 4 processors was 5.260822e+00 seconds. Unlike observation for table 1, this time it is the exact opposite the less processors the faster to compute.

Table 3: Large data size of 400 on different cores.

	1 processor 4.454129e+00 sec	2 processors 4.454129e+00 sec	4 processors 6.894232e+00 sec
1.0 00- 1.9 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX
1.9 00- 2.8 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
2.8 00- 3.7 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
3.7 00- 4.6 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
4.6 00- 5.5 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
5.5 00- 6.4 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX
6.4 00- 7.3 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX XXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX XXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX XXXXX
7.3 00- 8.2 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
8.2 00- 9.1 00:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
9.1 00- 10. 00 0:	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX X	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX X	XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX X

Observation: The data is same for all the processors. The elapsed time for 1 processor was 4.454129e+00 seconds. The elapsed time for 2 processors was 7.705539e+00 seconds. The elapsed time for 3 processors was 6.894232e+00 seconds. Unlike observation for table 1 and table 2, it was the 2 processors time which was the fastest to compute.

Conclusion:

1. For what combinations of p (number of processors) and N (dataset size) is it worthwhile to run the program in parallel? Define worthwhile as $E > 0.7$ (where E is efficiency).
I tried multiple combinations and could not find any.
2. If no such combination exists, what is closest?
The closet I was able to find is 4 processors and 4000 data size.
3. Under what circumstances would you, based in your results, run this as a parallel program on USS?
For larger data sizes I would run this parallel program on the USS.