

Plasma scaling study

Severin Staudinger, Simon Haendeler

January 30, 2017

Abstract

We are examining the scaling of the PLASMA library in case of the Cholesky Decomposition. We compare the runtime when using a different amount of cores and compare this to the runtime of LAPACK.

1 Introduction

How to cite: [?]

2 Implementation

3 Evaluation

Because the measurement was done on a shared system we could not guarantee that our program was the only one running. Measurement errors through context switches and alike are possible. Each datapoint represents 5 measurements that were averaged so that small measurement errors cancel out. Systematic error could still be present i.e. long running processes which affect all measurements are still possible. Although we tried to control for those manually via htop and similar tools, we cannot exclude those with 100% certainty.

In figure we see the runtime of the LAPACK version and of the PLASMA version with different amount of cores/threads used. As expected the runtime mostly increases like x^3 . The single threaded PLASMA version is slightly slower than the sequential LAPACK version. The other PLASMA versions are faster the more cores there are, at least for n large enough. The versions for 8 and 16 threads are very close together. 32 and 48 threads are also close, although we expect that this gap should be relatively narrow. There are bumps in the graph of the LAPACK version for $n = 4100$ and $n = 5100$.

In figure we see the relative speedup for each version. For comparison there are black lines at 2, 4, 8, 16, 32, 48 which represent the theoretical possible speedups. We can see that the parallel versions start slower than the sequential version. Then for relatively low n each version achieves a local maxima. For a higher thread number the n is higher when the maxima is reached.

Only 2 and 4 threads nearly reaches its theoretical performance, the higher the amount of threads, the worse the performance relatively to the theoretical performance.

The residuum is shown in figure . All PLASMA versions have the same residuum behavior, but LAPACK is always a bit better. There are certain

thresholds after which the residual gets worse (200, 500, 1000, 2000, 4000). After these sudden increases the residuum gets better with higher n .

4 Conclusions

We can see clearly that PLASMA scales with the cores as expected. Sadly for a high number of threads it does not reach its theoretical peak performance.