

# Intro HPC: Blatt 7

24.11.1014

## **7.1 Reading: LogP: Towards a Realistic Model of Parallel Computation**

The author is introducing a model, that describe the bottlenecks of a distributed memory multiprocessor in which processors communicate by point-to-point messages. Therefore it is based on a few of parameters like latency (L), overhead (o), bandwidth (g) of communication and the number of processes (P) and the assumption of a finite capacity. Furthermore it is generally for different types of communication protocols or applications.

The author is finding, that for some applications some parameters can get negligible, which makes a simplification of the model possible. He tested the model on different workloads like the FFT and the LU-Decomposition, to show how the use of the model can lead to efficient parallel applications in practice. In the paper he is also comparing this model to the widely used PRAM and the BSP model, which do not accurately reflect the performance characteristics of such systems, in his opinion.

This model is promising a couple of advantages to other often used models like including asynchronous algorithms. It is very general, so it might give a good overview of a system but not very detailed.

## **7.2 Reading: Roofline - An Insightful Visual Performance Model for Multicore Architectures**

In this paper the author introduces a visual computational model for multicore Architectures called 'roofline' model. It relates the operational intensity (mean operations per byte of DRAM traffic) with an upper bound for performance of a kernel (Attainable GFlops/s). For that 'roofline' uses the minimum of Peak Floating-Point Performance (constant) and the Peak memory Bandwidth multiplied with the operational intensity (line with positive slope), so whether the problem is compute-bound or memory-bound.

In that way it tells the user what optimizations should be implemented and in what order, by pointing out the limiting factor of performance. Using micro benchmarks he is testing 4 different kernels to apply this model.

The Roofline model seems to be a very clear and easy to apply way for characterizing multicore architectures for to find a first starting point for optimizations.

## **7.3 n-Body Problem – Partitioning/Communication Design**

In our n-Body code firstly the arrays of the acceleration and positions and the mass are allocated dynamically. Then we have the functions 'accelerations' and 'pos\_update' that

update first the acceleration and afterwards the positions of all the particles in every step.

Since the masses are fixed it will be the best to broadcast it once at the beginning of the calculation. For the communication to update the positions will have a ring communication. So everybody sends the new position to the neighbour till it is broadcasted completely.

To hide the latencies we use an unblocking send so the processes can go on computing while sending the results of the particle before.

To avoid collective calls the sum over all other particles in the force calculation starts at  $i+1$  so firstly we make sure no pair is counted twice and secondly the firstly needed particle position is different for every process. In this case you have to consider Newton's laws by counting force and anti-force.